**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**ĐỒ ÁN HỌC PHẦN**

TÊN HỌC PHẦN: **Kho dữ liệu (DWH)**

MÃ SỐ LỚP HP: **DAWH430784\_06**

Tên đề tài Nhóm: **Phát Triển Mô Hình Kho Dữ Liệu Game Sales Khai Thác Tại Công Ty Phân Phối Game**

Lớp: **181330A, 181330B, 181330C**

**NHÓM : 1**

THÀNH VIÊN:

1. Nguyễn Tuấn Vũ [55]\_Nhóm trưởng Phân hệ phân tích DW theo nhà phát hành Publisher
2. Trương Hùng Anh [01]: Phân hệ phân tích DW theo khu vực Region
3. Trần Thị Lệ Xuân [56]: Phân hệ phân tích DW theo thể loại Gerne
4. Nguyễn Xuân Hiệu [10]: Phân hệ phân tích DW theo nền tảng Platform

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác}

**Ngày nộp: 18/06/2021**

**TP.HCM, ngày 15 Tháng 06 năm 2021**

**NHẬN XÉT CỦA GIẢNG VIÊN ĐỐI VỚI NHÓM:**

**………………………………………………………………..**

**………………………………………………………………..**

**……………………………………………………………….**

**NHÓM : 1**

**ĐIỂM ĐÁNH GIÁ**

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác: NẾU DS DÀI = Trang riêng: 1 lần :BỎ DS BÌA & BÌA LÓT}

1. Nguyễn Tuấn Vũ [55] **:………….(…………………)………….**
2. Trương Hùng Anh [01] **:………….(…………………)………….**
3. Trần Thị Lệ Xuân [56] **:………….(…………………)………….**
4. Nguyễn Xuân Hiệu [10] **:………….(…………………)………….**

**Ngày ……../06/2021**

**Giảng viên Ký tên**

THÀNH VIÊN:

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác}

1. Nguyễn Tuấn Vũ [55]\_Nhóm trưởng Phân hệ phân tích DW theo nhà phát hành Publisher
2. Trương Hùng Anh [01]: Phân hệ phân tích DW theo khu vực Region
3. Trần Thị Lệ Xuân [56]: Phân hệ phân tích DW theo thể loại Gerne
4. Nguyễn Xuân Hiệu [10]: Phân hệ phân tích DW theo nền tảng Platform

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**ĐỀ TÀI ĐỒ ÁN HỌC PHẦN**

TÊN HỌC PHẦN: **Kho dữ liệu (DWH)**

MÃ SỐ LỚP HP: **DAWH430784\_06**

Tên đề tài Nhóm: **Phát Triển Mô Hình Kho Dữ Liệu Game Sales Khai Thác Tại Công Ty Phân Phối Game**

**Giảng viên giảng dạy: VÕ XUÂN THỂ \_ Giảng viên chính**

Lớp: **181330A, 181330B, 181330C (Sáng thứ 4)**

**Tên sản phẩm đề tài: 56.N1.04.TranThiLeXuan.GameSale.DWH.rar** (Microsoft SQL Server 2014)

**DB Nguồn: video\_game**

**DB DWH: video\_game\_DW**

**Tên máy chủ DB SQL:**

**Tên đăng nhập DB SQL: sa**

**Mật khẩu (password) đăng nhập SQL: 123**

**ProjectName BIDS: 56N1TranThiLeXuanVideogamesDW: 05N1VXTheQLKCBBHYT\_DWH**

**NHÓM : 1**

THÀNH VIÊN:

1. Nguyễn Tuấn Vũ [55]\_Nhóm trưởng Phân hệ phân tích DW theo nhà phát hành Publisher
2. Trương Hùng Anh [01]: Phân hệ phân tích DW theo khu vực Region
3. Trần Thị Lệ Xuân [56]: Phân hệ phân tích DW theo thể loại Gerne
4. Nguyễn Xuân Hiệu [10]: Phân hệ phân tích DW theo nền tảng Platform

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác

**TP.HCM, ngày 15 Tháng 06 năm 2021**

# LỜI CẢM ƠN

Chúng em xin chân thành cảm ơn sự hướng dẫn tận tình của thầy ***Võ Xuân Thể*** đã giúp chúng em hoàn thành bài báo cáo cuối kì bộ môn Kho dữ liệu. Tuy đã cố gắng học tập, tìm hiểu, phân tích thiết kế và xây dựng hệ thống nhưng ắt hẳn thiếu sót là điều không thể tránh khỏi. Chúng em rất mong nhận được sự thông cảm và góp ý chân thành của Thầy. Chúng em xin chân thành cảm ơn.

# DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |
| --- | --- |
| CSDL hoặc DB | Cơ sở dữ liệu: DataBase |
| LĐQH | Lược đồ quan hệ |
| PTH | Phụ thuộc hàm = Ràng buộc tham chiếu: Reference Constraint |
| PTTK | Phân tích và thiết kế |
| RBTV | Ràng buộc toàn vẹn [dữ liệu] |
|  |  |
| BIDS | Công cụ hỗ trợ tổ chức, quản lý và khai thác kho dữ liệu (Hệ thống tổ chức dữ liệu thông minh cho các doanh nghiệp): Business Intelligence Development Studio của Microsoft |
| DWH, DW | Kho dữ liệu (Nhà kho dữ liệu): DataWareHouse |
| ERD | Sơ đồ thực thể kết hợp: Entity Relationship Diagram |
| MS | Công ty Microsoft |
| NF | Dạng chuẩn của CSDL: Normal Form |
| SQL | Ngôn ngữ vấn tin có cấu trúc: Structured Query Language |

# DANH MỤC CÁC THUẬT NGỮ

|  |  |
| --- | --- |
| Giao diện người dùng  (User-Interface) | Là hệ thống các màn hình giao tiếp cho phép người sử dụng tương tác với các thành phần phần mềm, điều khiển phần mềm hoạt động theo yêu cầu của người dùng - tương ứng các chức năng hiện có của phần mềm. |
| Người dùng (User):  Tài khoản (Account) | Là một quyền làm việc trên hệ thống phần mềm được cấp phát cho một cá nhân thông qua tên tài khoản (username) và mật khẩu (password). |
|  |  |
| BigData | Dữ liệu lớn: là một tập hợp dữ liệu rất lớn và phức tạp, không thể xử lý dữ liệu bằng các phương pháp truyền thống. |

# MỤC LỤC

[LỜI CẢM ƠN 4](#_Toc73666776)

[DANH MỤC CÁC TỪ VIẾT TẮT 5](#_Toc73666777)

[DANH MỤC CÁC THUẬT NGỮ 6](#_Toc73666778)

[MỤC LỤC 7](#_Toc73666779)

[Chương 1. GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN 10](#_Toc73666780)

[1.1. Tổng quan về ĐỒ ÁN HỌC PHẦN 10](#_Toc73666781)

[1.2. Nội dung chuyên môn chính của ĐỒ ÁN HỌC PHẦN 10](#_Toc73666782)

[1.3. Mục tiêu của ĐỒ ÁN HỌC PHẦN 10](#_Toc73666783)

[1.4. Công cụ và nền tảng kỹ thuật thực hiện ĐỒ ÁN HỌC PHẦN 10](#_Toc73666784)

[1.5. Sản phẩm của ĐỒ ÁN HỌC PHẦN 10](#_Toc73666785)

[1.6. Bố cục của báo cáo 10](#_Toc73666786)

[Chương 2. CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN 11](#_Toc73666787)

[2.1. Tổng quan về kho dữ liệu (DWH) 11](#_Toc73666788)

[2.1.1. Các khái niệm cơ bản về Kho dữ liệu 11](#_Toc73666789)

[2.1.2. Kiến trúc cơ bản của kho dữ liệu: 11](#_Toc73666790)

[2.1.3. Mô hình CSDL đa chiều (Dimensional Modeling) 11](#_Toc73666791)

[2.2. Phân tích và thiết kế Kho dữ liệu 11](#_Toc73666792)

[2.2.1. Phân tích và thiết kế kho dữ liệu theo hướng mô hình hóa nhiều chiều (Dimensional Modeling) 11](#_Toc73666793)

[2.2.2. Các dạng (phương pháp) tích hợp dữ liệu 11](#_Toc73666794)

[2.2.3. Các bước cơ bản trong tiến trình tích hợp dữ liệu 11](#_Toc73666795)

[2.2.4. Các công cụ tích hợp dữ liệu: 11](#_Toc73666796)

[2.2.5. Lập các luồng công việc tích hợp dữ liệu (Data Integration Workflows) 11](#_Toc73666797)

[2.2.6. Hệ thống OLTP (On-Line Transaction Processing = Xử lý giao dịch trực tuyến) 11](#_Toc73666798)

[2.3. Vấn tin khai thác sử dụng Kho dữ liệu 11](#_Toc73666799)

[2.3.1. Vấn tin với các phép toán tổng hợp của SQL trên kho dữ liệu 11](#_Toc73666800)

[2.3.2. Vấn tin với các hàm phân tích của SQL trên kho dữ liệu 11](#_Toc73666801)

[2.3.3. Vấn tin dùng Materialized View để phân tích trên kho dữ liệu 11](#_Toc73666802)

[2.3.4. Hệ thống OLAP 11](#_Toc73666803)

[Chương 3. TỔNG QUAN VỀ KHO DỮ LIỆU …<theo tên đề tài>…… TẠI ……<theo tên đề tài>………… 12](#_Toc73666804)

[3.1. Ý tưởng hình thành DWH …<theo tên đề tài>…… tại ……<theo tên đề tài>………… 12](#_Toc73666805)

[3.2. Giới thiệu chung về Kho dữ liệu 12](#_Toc73666806)

[3.2.1. Mô tả tổng quan về DWH 12](#_Toc73666807)

[3.2.2. Các dữ liệu nguồn hình thành DWH 12](#_Toc73666808)

[3.3. Xác định nhu cần tổ chức và phân tích dữ liệu của toàn bộ DWH 12](#_Toc73666809)

[3.4. Các thành phần phân hệ Kho dữ liệu phân công các cá nhân phụ trách 12](#_Toc73666810)

[Chương 4. PHÂN TÍCH, THIẾT KẾ VÀ KHAI THÁC SỬ DỤNG CÁC PHÂN HỆ KHO DỮ LIỆU …<theo tên đề tài>…… TẠI ……<theo tên đề tài>………… 13](#_Toc73666811)

[4.1. <STT>\_ Ho Tên 1\_ <thành phần kho dữ liệu đươc phân công> 13](#_Toc73666812)

[4.1.1. Mô tả tổng quan về phân hệ DWH…<tên phân hệ DWH>… 13](#_Toc73666813)

[4.1.2. Xác định nguồn dữ liệu tích hợp phân hệ Kho dữ liệu 13](#_Toc73666814)

[4.1.3. Xác định ý tưởng hình thành phân hệ của Kho dữ liệu 13](#_Toc73666815)

[4.1.4. Mô tả hệ thống các DB gốc liên quan phân hệ DWH 13](#_Toc73666816)

[4.1.5. Phân phân tích các DB gốc xác định yêu cầu phân tích DWH của phân hệ 13](#_Toc73666817)

[4.1.6. Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ 14](#_Toc73666818)

[4.1.7. Thực hiện các thủ tục tích hợp các DB gốc vào phân hệ DWH 15](#_Toc73666819)

[4.1.8. Thiết lập các Views tính toán cần thiết để nạp dữ liệu từ DB gốc vào các Factors của phân hệ DWH 15](#_Toc73666820)

[4.1.9. Triển khai phân hệ DWH lên BIDS 15](#_Toc73666821)

[4.1.10. Khai thác sử dụng phân hệ DWH dùng các vân tin SQL 15](#_Toc73666822)

[4.2. <STT>\_ Ho Tên 2\_ <thành phần kho dữ liệu đươc phân công> 15](#_Toc73666823)

[Chương 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN 16](#_Toc73666824)

[5.1. Kết luận 16](#_Toc73666825)

[5.1.1. Những kết quả đạt được 16](#_Toc73666826)

[5.1.2. Hạn chế 16](#_Toc73666827)

[5.2. Hướng phát triển 16](#_Toc73666828)

[5.2.1. Hướng khắc phục các hạn chế 16](#_Toc73666829)

[5.2.2. Hướng mở rộng ĐỒ ÁN HỌC PHẦN 16](#_Toc73666830)

[DANH MỤC TÀI LIỆU THAM KHẢO 17](#_Toc73666831)

[CÁC PHỤ LỤC 18](#_Toc73666832)

[ĐÁNH GIÁ KẾT QUẢ HỌC TẬP HỌC PHẦN 19](#_Toc73666833)

1. GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN
   1. Tổng quan về ĐỒ ÁN HỌC PHẦN

Đồ án Học phần môn Kho dữ liệu: **Phát Triển Mô Hình Kho Dữ Liệu Game Sales Khai Thác Tại Công Ty Phân Phối Game**

* 1. Nội dung chuyên môn chính của ĐỒ ÁN HỌC PHẦN

Phân tích, thiết kế, chuyển đổi dữ liệu database gốc sang nhà Kho dữ liệu

* 1. Mục tiêu của ĐỒ ÁN HỌC PHẦN

Nắm được cách phân tích thiết kế Data Warehouse theo nhu cầu phân tích dữ liệu

Nắm được kỹ thuật tích hợp dữ liệu từ các nguồn về Data Warehouse

Nắm được kỹ thuật phân tích bằng BIDS, Subtotal,…

* 1. Công cụ và nền tảng kỹ thuật thực hiện ĐỒ ÁN HỌC PHẦN

Các công cụ chính sử dụng trong đồ án này là

+ Microsoft SQL Server

+ SQL Server Business Intelligence Development Studio (BIDS)

* 1. Sản phẩm của ĐỒ ÁN HỌC PHẦN

Kho dữ liệu “video\_games\_DW”

BIDS Project

Các view phân tích, truy vấn Database gốc, DW

* 1. Bố cục của báo cáo

Báo cáo gồm những nội dung như sau:

Chương 1: Giới thiệu đồ án của học phần DWH

Chương 2: Các cơ sở lý thuyết của HP phục việc thực hiện đề tài HP

Chương 3: Tổng quan về kho dữ liệu game sales khai thác tại công ty phân phối game

Chương 4: Phân tích, thiết kế và tổ chức khai thác sử dụng từng phân hệ của Kho dữ liệu

Chương 5: Tổng kết các kết quản đạt được và còn hạn chế của đồ án, đồng thời đề xuất hướng khắc phục hạn chế và phát triển Đồ án.

1. CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN
   1. Tổng quan về kho dữ liệu (DWH)
      1. Các khái niệm cơ bản về Kho dữ liệu
         1. Kho dữ liệu (DWH)?
         2. Nguyên lý hình thành DWH
         3. Các đặc điểm cơ bản của DWH
         4. Phân biệt [SS] DWH với CSDL truyền thống
      2. Kiến trúc cơ bản của kho dữ liệu:
         1. Tầng đáy: nạp dữ liệu
         2. Tầng giữa: OLAP (OLAP server): xử lý DWH
         3. Tầng trên cùng: ứng dụng = khai thác sử dụng DWH
      3. Mô hình CSDL đa chiều (Dimensional Modeling)
   2. Phân tích và thiết kế Kho dữ liệu
      1. Phân tích và thiết kế kho dữ liệu theo hướng mô hình hóa nhiều chiều (Dimensional Modeling)
      2. Các dạng (phương pháp) tích hợp dữ liệu
      3. Các bước cơ bản trong tiến trình tích hợp dữ liệu
      4. Các công cụ tích hợp dữ liệu:
      5. Lập các luồng công việc tích hợp dữ liệu (Data Integration Workflows)
      6. Hệ thống OLTP (On-Line Transaction Processing = Xử lý giao dịch trực tuyến)
   3. Vấn tin khai thác sử dụng Kho dữ liệu
      1. Vấn tin với các phép toán tổng hợp của SQL trên kho dữ liệu
      2. Vấn tin với các hàm phân tích của SQL trên kho dữ liệu
      3. Vấn tin dùng Materialized View để phân tích trên kho dữ liệu
      4. Hệ thống OLAP
2. TỔNG QUAN VỀ KHO DỮ LIỆU GAME SALES KHAI THÁC TẠI CÔNG TY PHÂN PHỐI GAME
   1. Ý tưởng hình thành DWH Game Sales khai thác tại công ty phân phối game

Ý tưởng: Xây dựng 1 Data Warehouse dưới dạng mô hình Star Schema có các Dim table (chứa các đối tượng cần phân tích) và Fact table (chứa các id của các đối tượng và 1 cột chứa giá trị được group by từ các id)

* 1. Giới thiệu chung về Kho dữ liệu
     1. Mô tả tổng quan về DWH

1 tập đoàn chuyên phân phối game cho các thị trường trên thế giới có 1 hệ thống database lưu trữ doanh số bán game qua hằng năm.

* + 1. Các dữ liệu nguồn hình thành DWH

Nguồn dữ liệu tích hợp cho phân hệ phân tích được lấy từ Database video\_games đặt tại trụ sở của công ty phân phối game.

* 1. Xác định nhu cần tổ chức và phân tích dữ liệu của toàn bộ DWH

Tập đoàn có nhu cầu phân tích các doanh số bán hàng đã thu thập qua các năm để hiểu thị trường, hiểu thể loại game nào được ưa thích, nhà phát hành, thể loại game nào nào đang làm mưa làm gió trên thị trường…

* 1. Các thành phần phân hệ Kho dữ liệu phân công các cá nhân phụ trách

<55>\_Nguyễn Tuấn Vũ\_ Phân hệ phân tích DW theo nhà phát hành Publisher

<01>\_Trương Hùng Anh\_ Phân hệ phân tích DW theo khu vực Region

<10>\_ Nguyễn Xuân Hiệu\_ Phân hệ phân tích kho dữ liệu theo nền tảng Platform

<56>\_Trần Thị Lệ Xuân\_ Phân hệ phân tích DW theo thể loại Gerne

1. PHÂN TÍCH, THIẾT KẾ VÀ KHAI THÁC SỬ DỤNG CÁC PHÂN HỆ KHO DỮ LIỆU GAME SALES KHAI THÁC TẠI CÔNG TY PHÂN PHỐI GAME
   1. Trần Thị Lệ Xuân [56]\_ Phân hệ phân tích DW theo Genre
      1. Mô tả tổng quan về Phân hệ phân tích DW theo Genre

Phân hệ gồm Dim Genre và Dim Year, Fact table chứa tổng doanh thu

* + 1. Xác định nguồn dữ liệu tích hợp phân hệ Kho dữ liệu

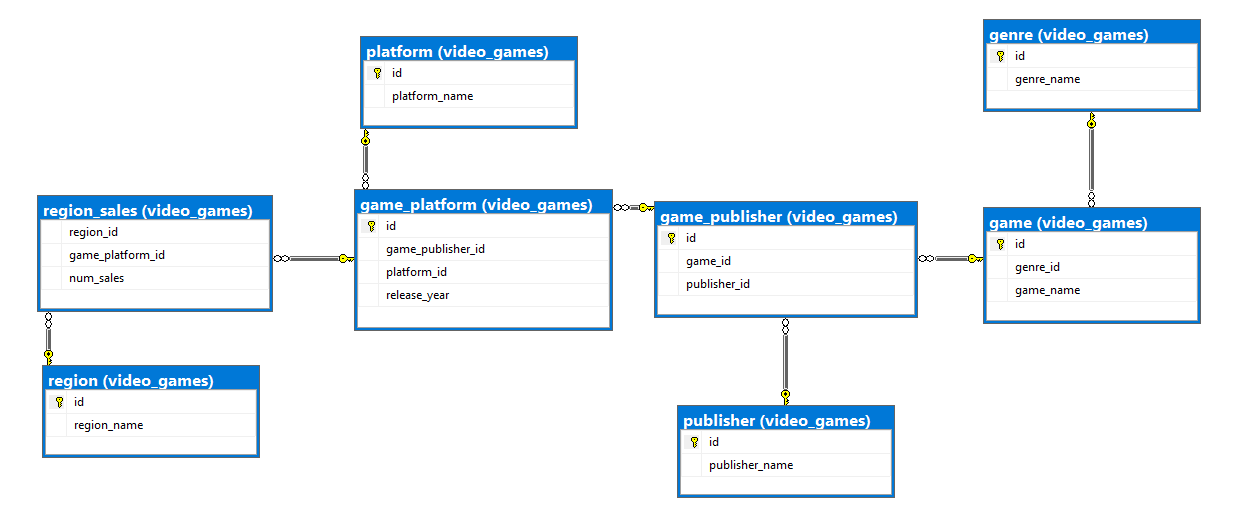
Nguồn dữ liệu tích hợp cho phân hệ phân tích được lấy từ Database video\_games đặt tại trụ sở của công ty phân phối game.

* + 1. Xác định ý tưởng hình thành phân hệ của Kho dữ liệu

Việc xác định doanh thu theo thể loại game giúp người quản lý, lập trình game biết được nhà phát hành để sản xuất theo nhu cầu thể loại mà khách hàng sử dụng, để đàm phán hợp tác tối ưu hoá doanh thu game

* + 1. Mô tả hệ thống các DB gốc liên quan phân hệ DWH

Hệ thống Database lưu trữ



Mô tả DB:

**region**(id, region\_name)

**genre**(id, genre\_name)

**publisher**(id, publisher\_name)

**platform**(id, platform\_name)

**game**(id, genre\_id, game\_name)

**game\_publisher**(id, game\_id, publisher\_id)

**game\_platform**(id, game\_publisher\_id, platform\_id, release\_year)

**region\_sales**(region\_id, game\_platform\_id, num\_sales)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **region** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | id | int | PRIMARY KEY | Khoá chính phân biệt vùng |
| 2 | region\_name | nvarchar(150) | NOT NULL | Tên vùng |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **genre** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | id | int | PRIMARY KEY | Khoá chính phân biệt thể loại |
| 2 | genre\_name | nvarchar(150) | NOT NULL | Tên thể loại |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **publisher** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | id | int | PRIMARY KEY | Khoá chính phân biệt nhà phát hành |
| 2 | publisher\_name | nvarchar(150) | NOT NULL | Tên nhà phát hành |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **platform** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | id | int | PRIMARY KEY | Khoá chính phân biệt nền tảng |
| 2 | platform\_name | nvarchar(150) | NOT NULL | Tên nền tảng |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **game** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | id | int | PRIMARY KEY | Khoá chính phân biệt game |
| 2 | game\_name | nvarchar(150) | NOT NULL | Tên game |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **game\_publisher** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | id | int | PRIMARY KEY | Khoá chính phân biệt game và nhà phát hành |
| 2 | game\_id | int | NOT NULL | ID game |
| 3 | publisher\_id | int | NOT NULL | ID nhà phát hành |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **game\_platform** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | id | int | PRIMARY KEY | Khoá chính phân biệt game và nền tảng |
| 2 | game\_publisher\_id | int | NOT NULL | ID game và nhà phát hành |
| 3 | platform\_id | int | NOT NULL | ID nền tảng |
| 4 | release\_year | int | NOT NULL | Năm phát hành |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **region\_sale** | | | | |
| **STT** | **Thuộc tính** | **Kiểu dữ liệu** | **Ràng buộc** | **Diễn giải** |
| 1 | region\_id | int | PRIMARY KEY | ID khu vực |
| 2 | game\_platform\_id | int | PRIMARY KEY | ID game và nền tảng |
| 3 | num\_sales | Decimal(5,2) | NOT NULL | Đơn vị bán (1.000.000 USD) |

* + 1. Phân phân tích các DB gốc xác định yêu cầu phân tích DWH của phân hệ

Dim:

DimRegion: Dim về khu vực/thị trường bán game

DimGerne: Dim về thể loại game

DimPublisher: Dim về nhà phát hành

DimPlatform: Dim về nền tảng phát triển

DimGame: Dim về game

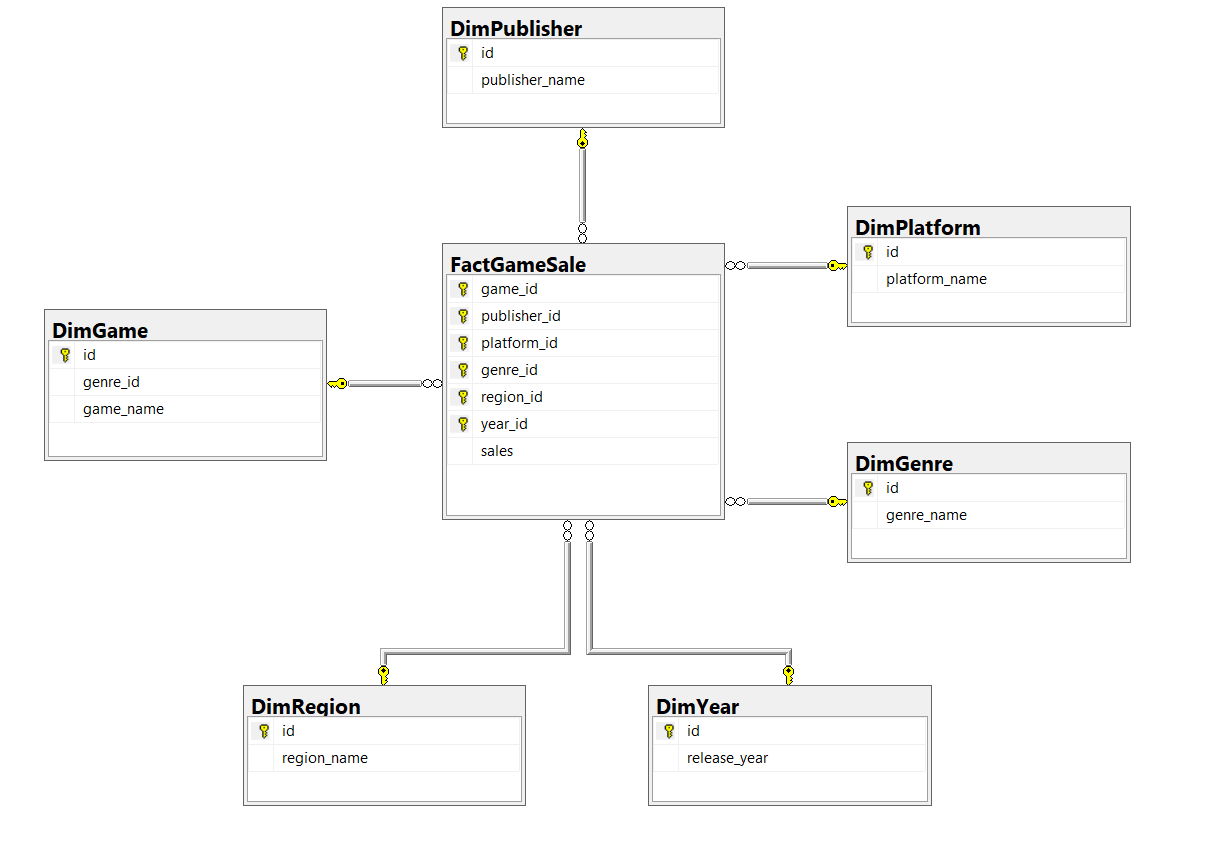
DimYear: Dim về năm phát hành

Có 1 fact:

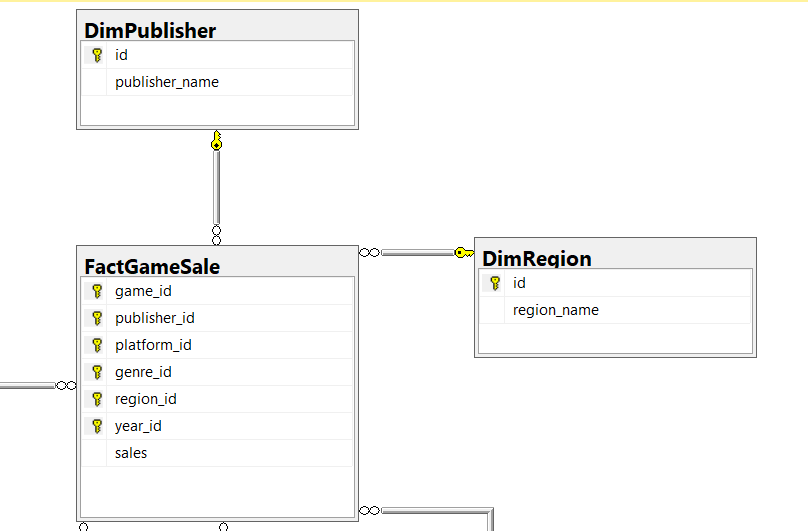
Fact doanh thu = sum(num\_sales) group by (id DimRegion, id DimGerne, id DimPublisher, id DimPlatform, id DimGame, id DimYear)

* + 1. Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ

DW chung :

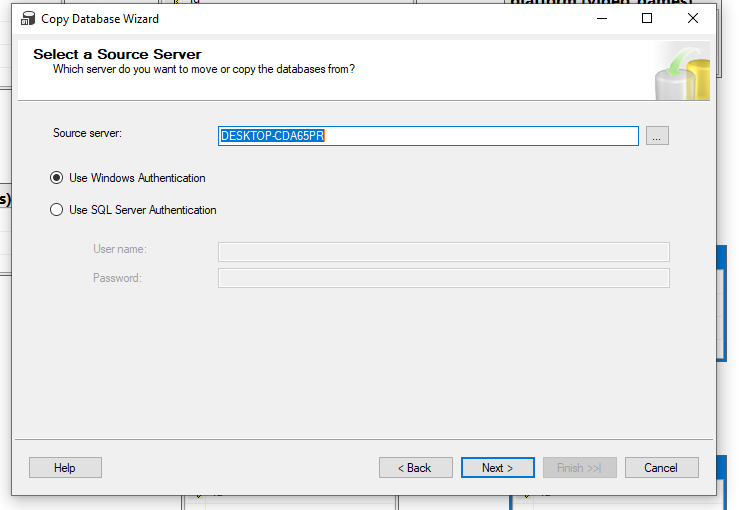


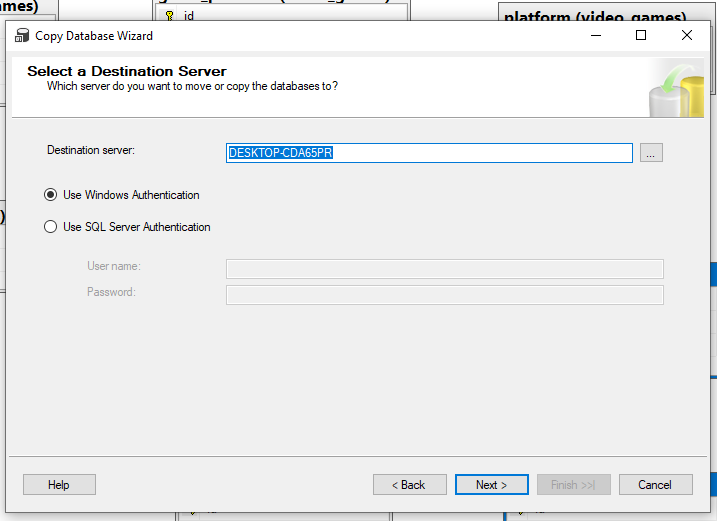
DW phân hệ phân tích **doanh thu theo nhà phát hành và khu vực**

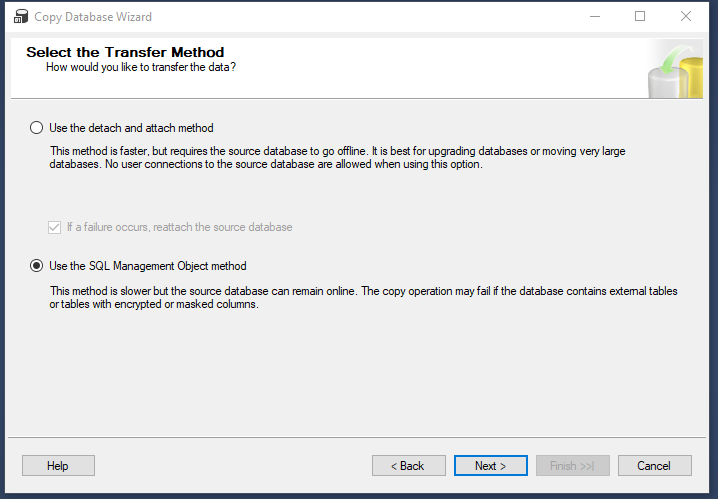


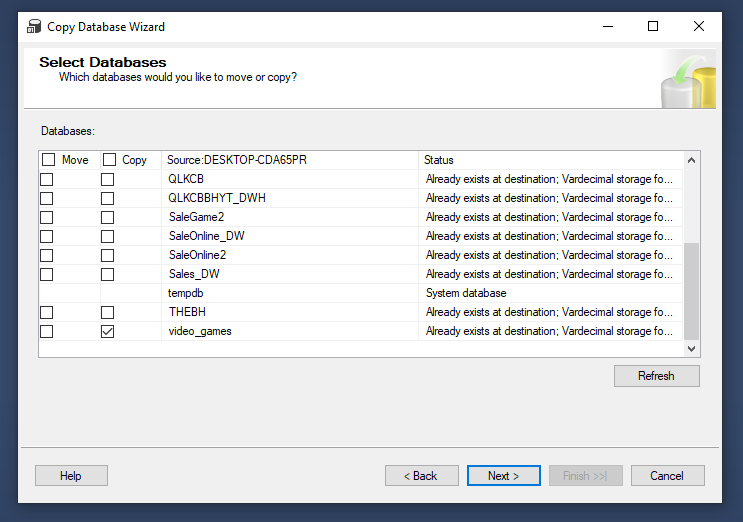
* + 1. Thực hiện các thủ tục tích hợp các DB gốc vào phân hệ DWH

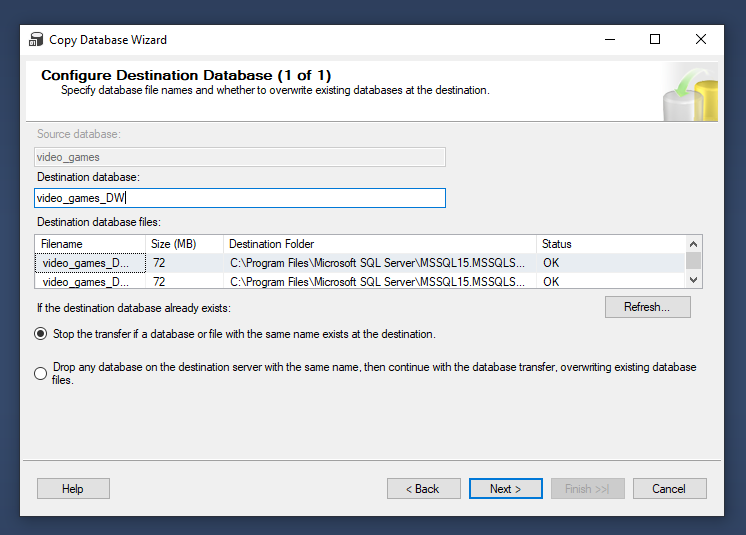
**1.** Thực hiện các thủ tục Copy DataBase cần thiết (để lập bản sao phục vụ phân tích) từ DB gốc sang DB của DHW phân tích.

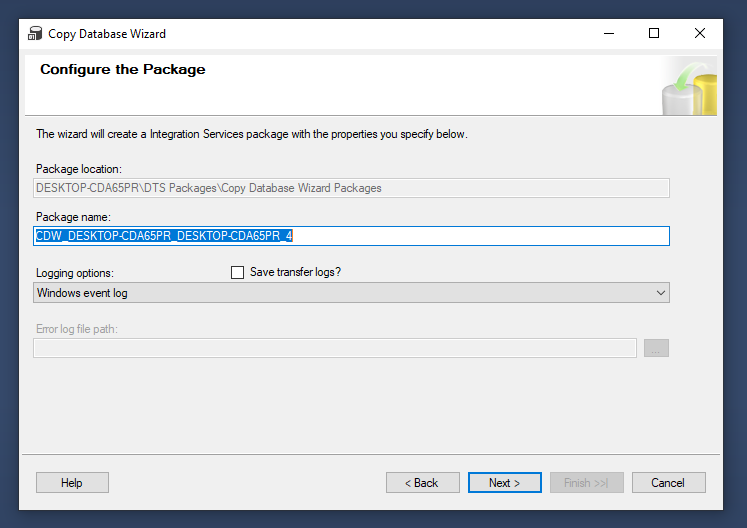


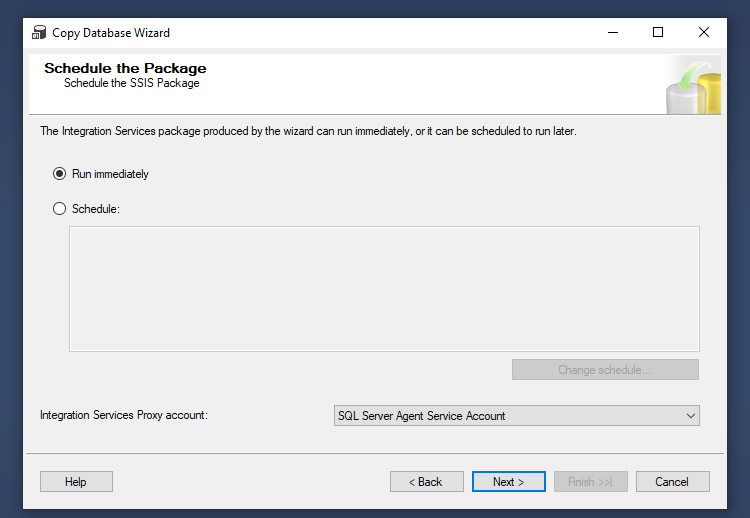


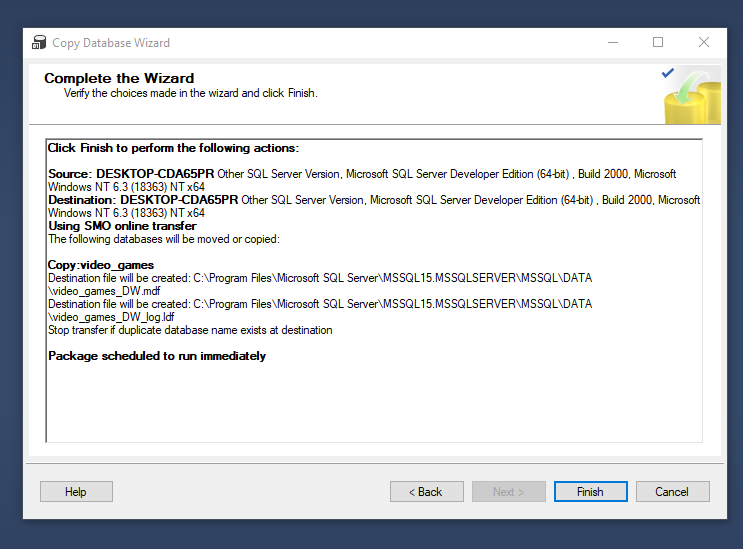


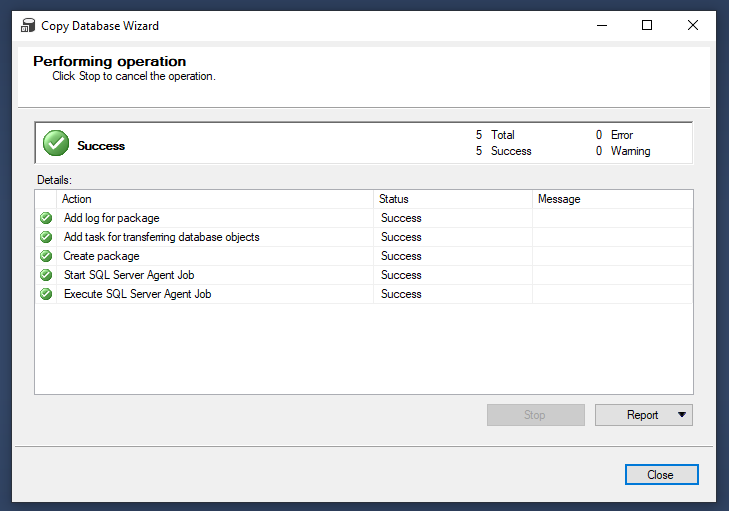








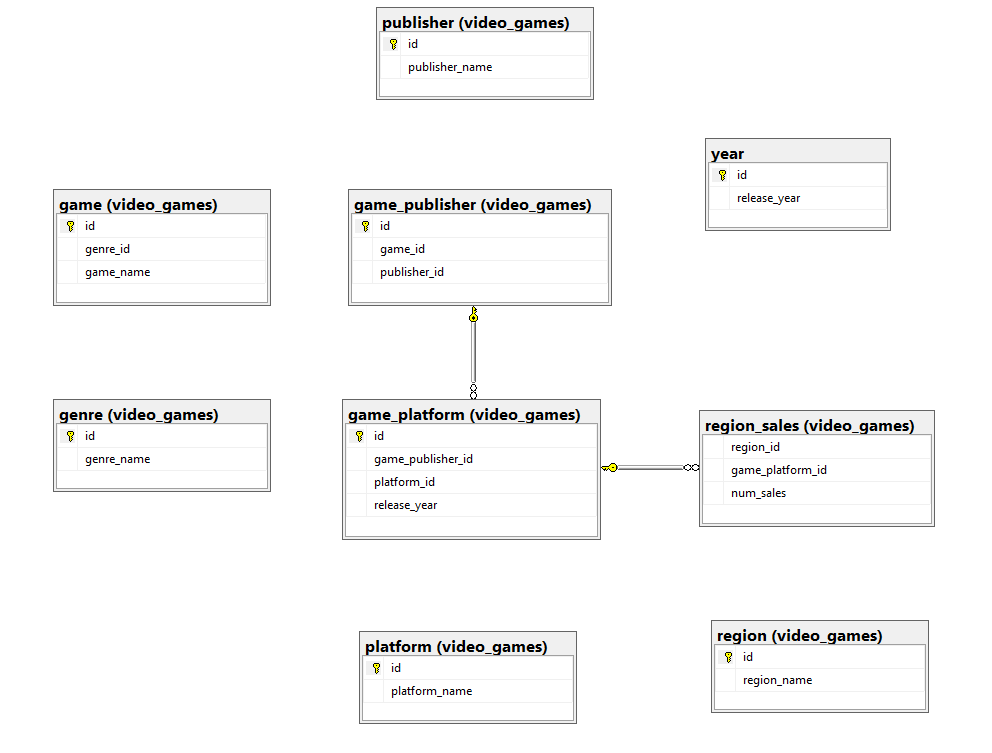


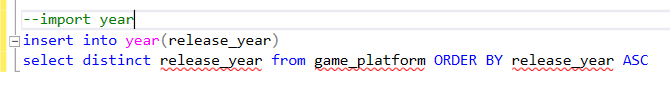




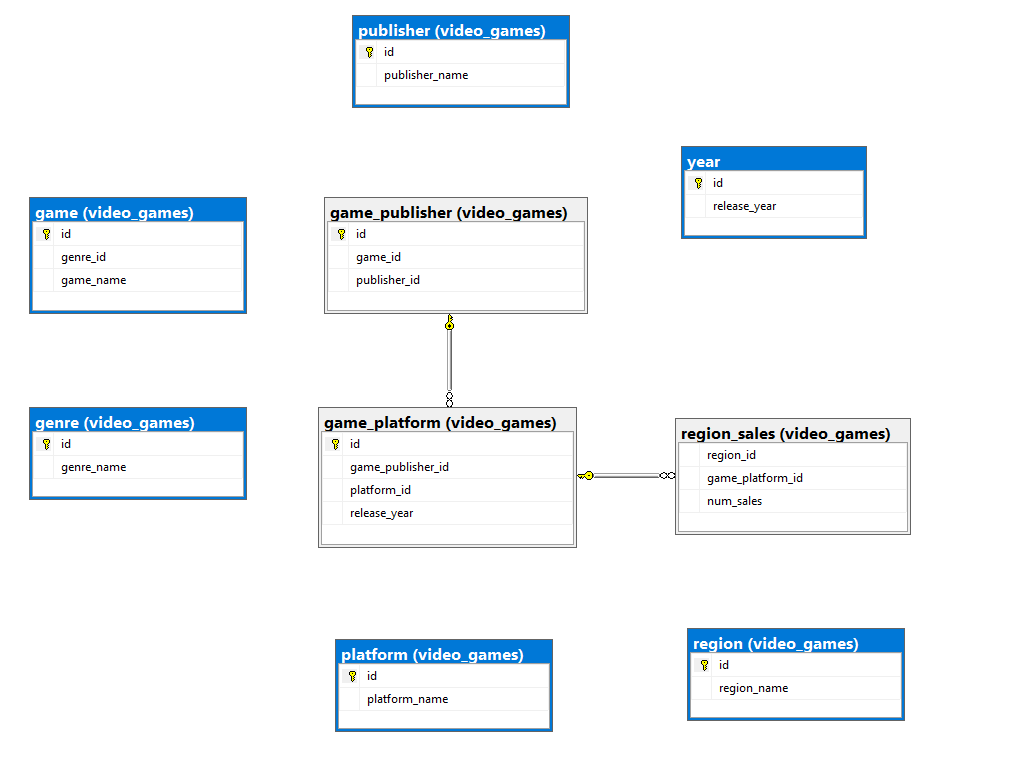
**2. Chuyển Database copy về DW**

**B1: Xoá các khoá ngoại, thêm table year**

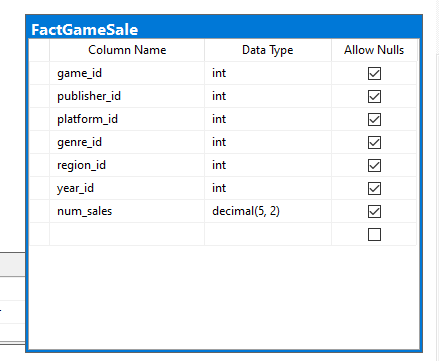




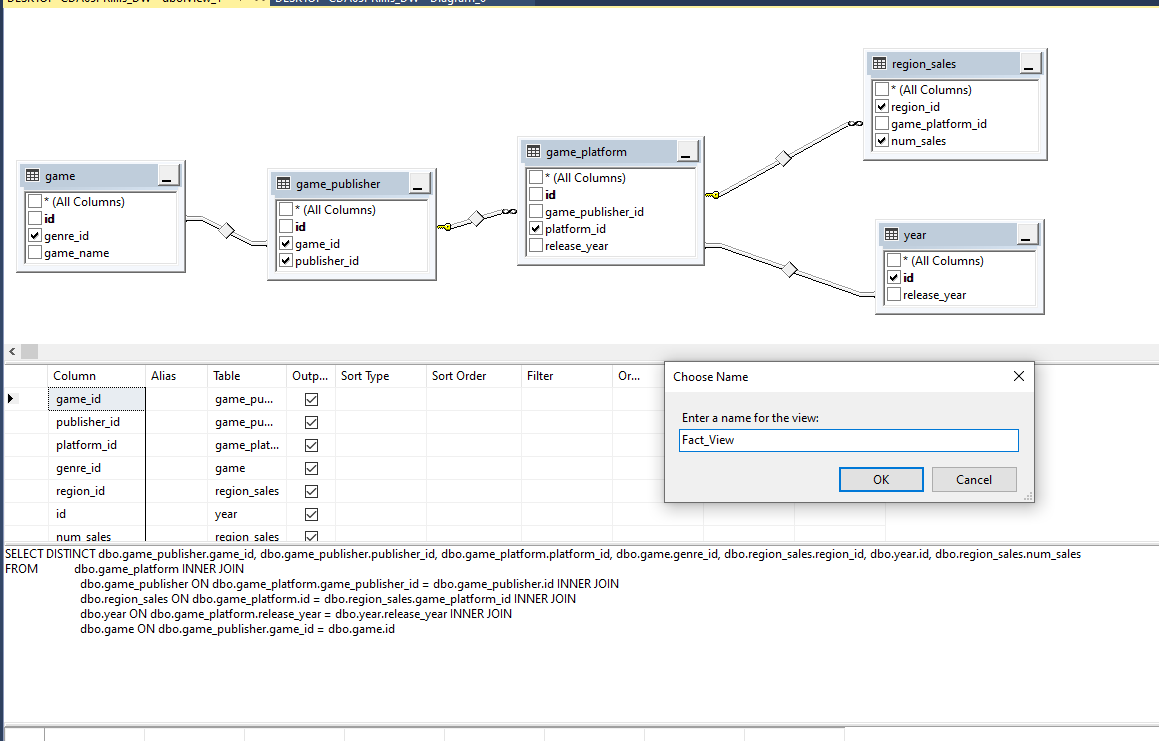
Các table có màu là đã nhập, chúng ta sẽ chuyển 3 table còn lại thành dim fact

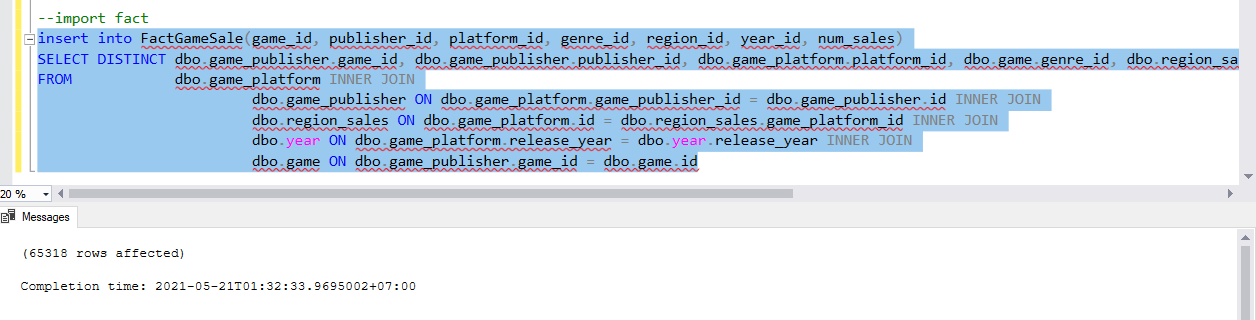


**B2: Tạo fact table**

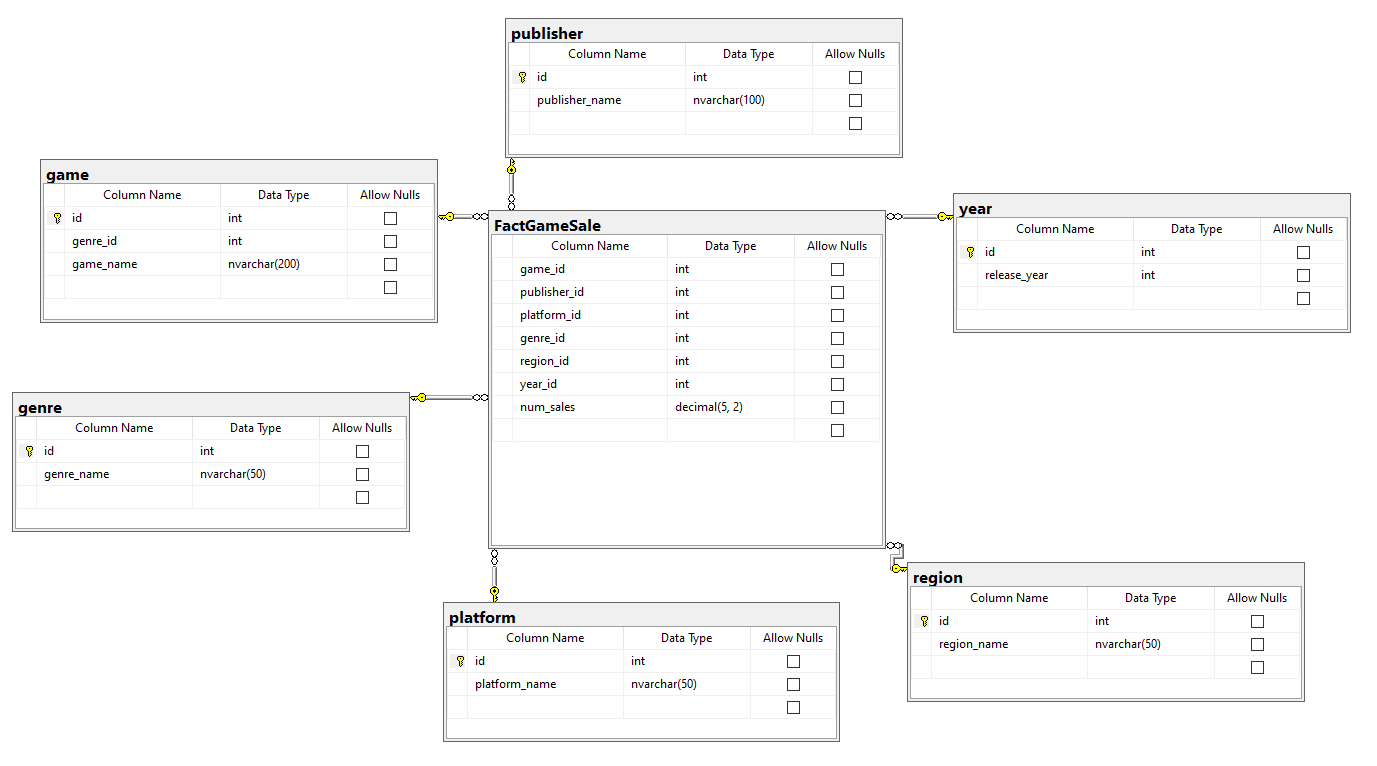


**B3: Insert dữ liệu cho fact table**

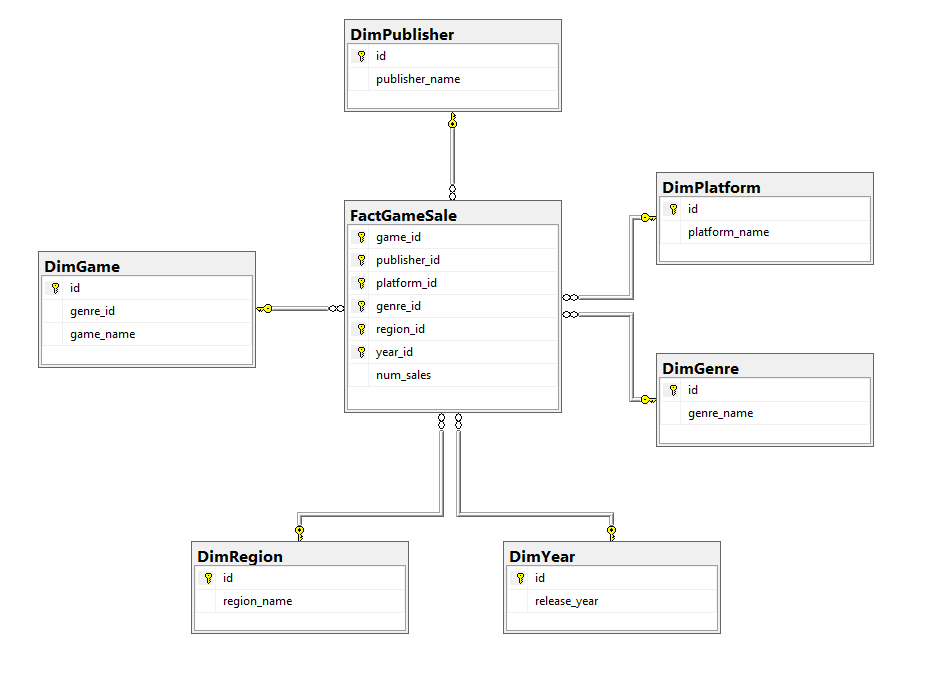




**B4: Đổi tên các table, nối lại các ràng buộc:**

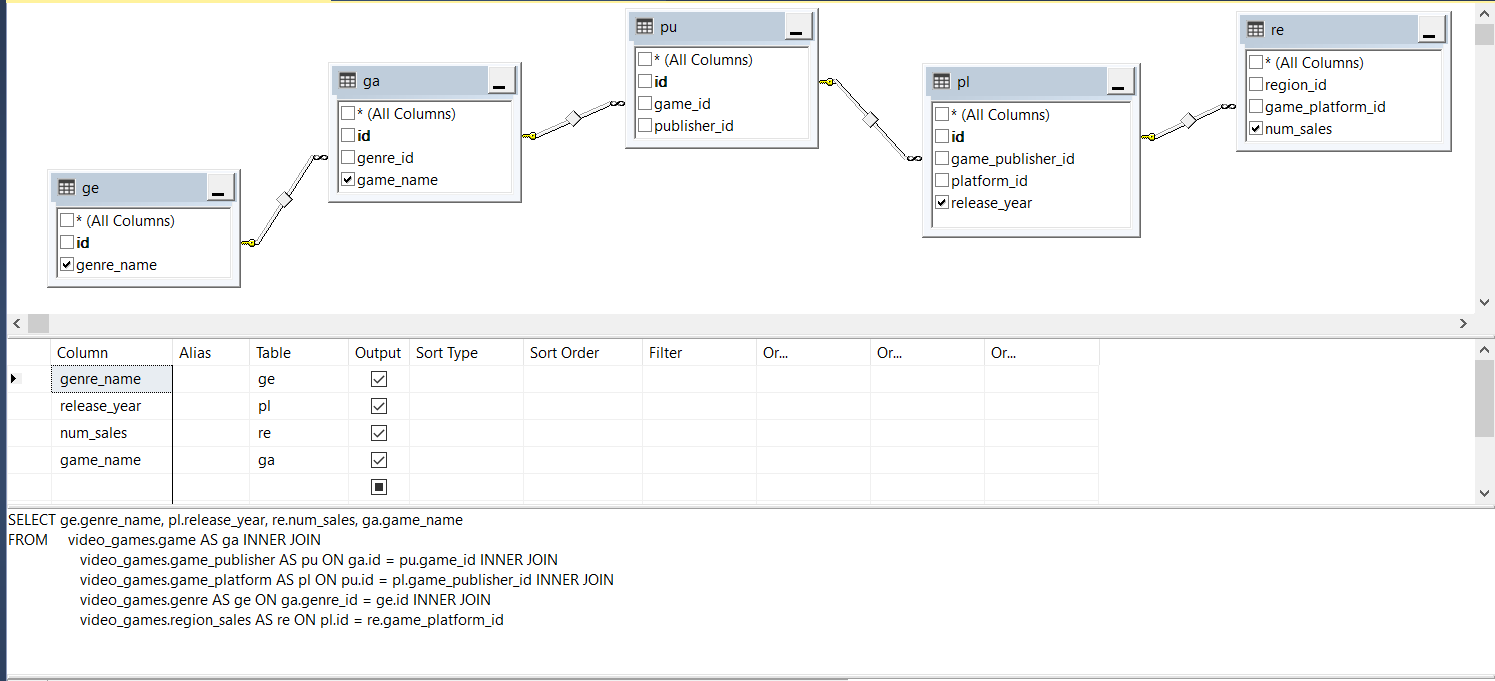


**B5: Set khoá chính cho fact**



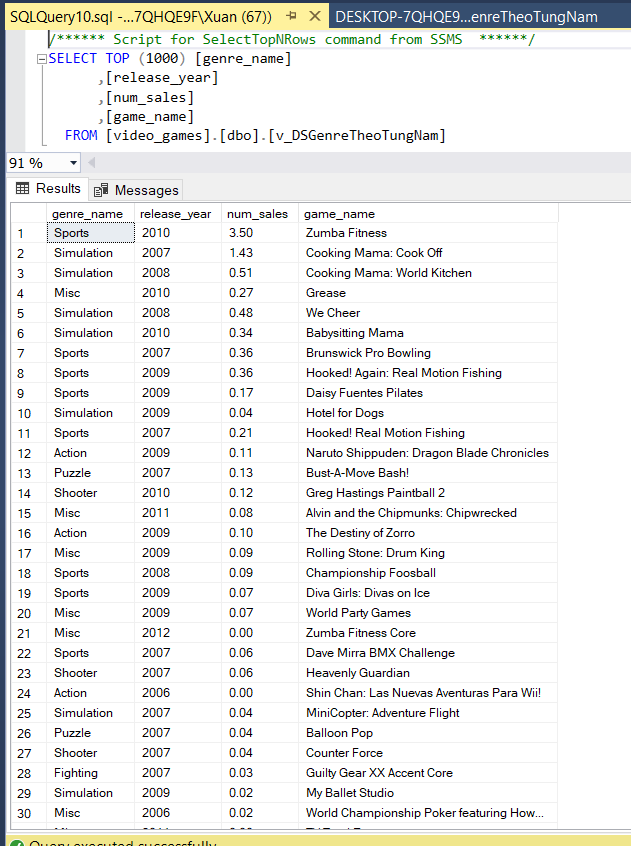
* + 1. Thiết lập các Views tính toán cần thiết để nạp dữ liệu từ DB gốc vào các Factors của phân hệ DWH

**v\_DSGenreTheoTungNam**



|  |
| --- |
| SELECT ge.genre\_name, pl.release\_year, re.num\_sales, ga.game\_name  FROM video\_games.game AS ga INNER JOIN  video\_games.game\_publisher AS pu ON ga.id = pu.game\_id INNER JOIN  video\_games.game\_platform AS pl ON pu.id = pl.game\_publisher\_id INNER JOIN  video\_games.genre AS ge ON ga.genre\_id = ge.id INNER JOIN  video\_games.region\_sales AS re ON pl.id = re.game\_platform\_id |

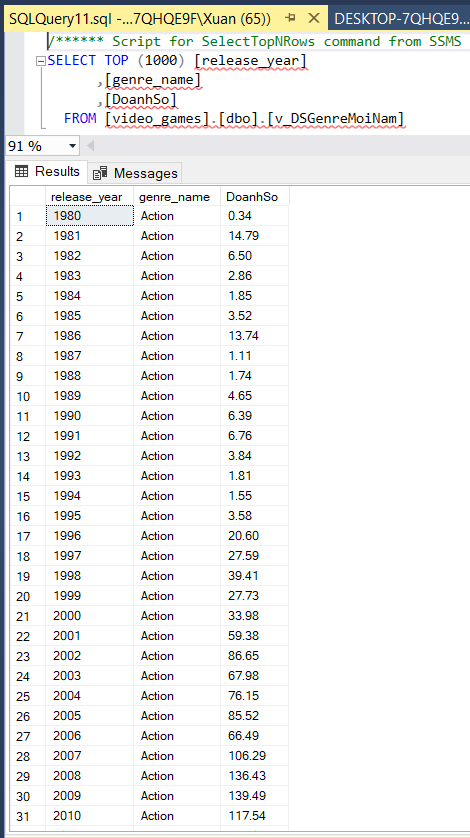
KẾT QUẢ



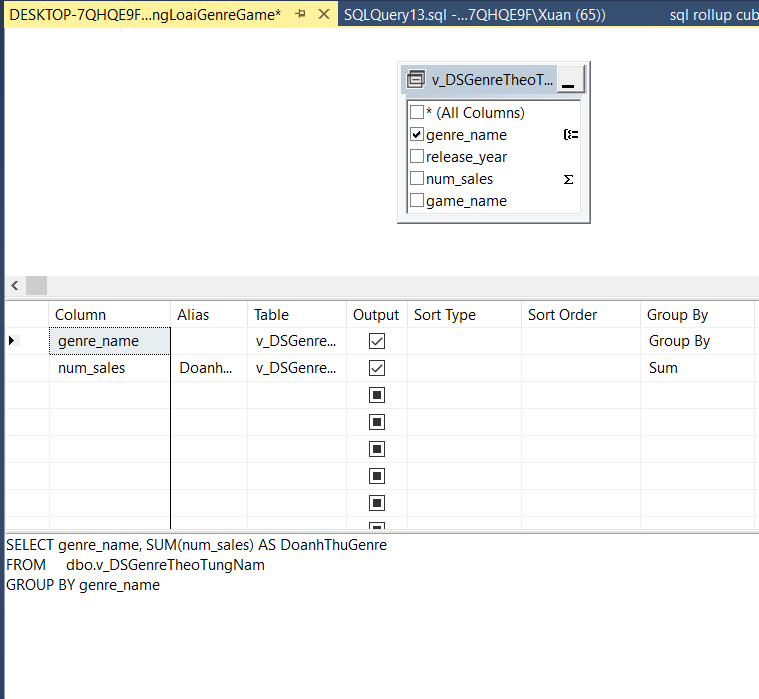
**v\_DSGenreMoiNam**

|  |
| --- |
| SELECT release\_year, genre\_name, SUM(num\_sales) AS DoanhSo  FROM dbo.v\_DSGenreTheoTungNam  GROUP BY release\_year, genre\_name |

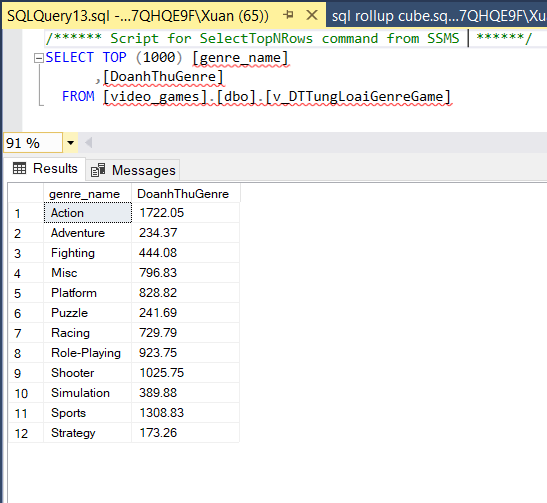
KẾT QUẢ



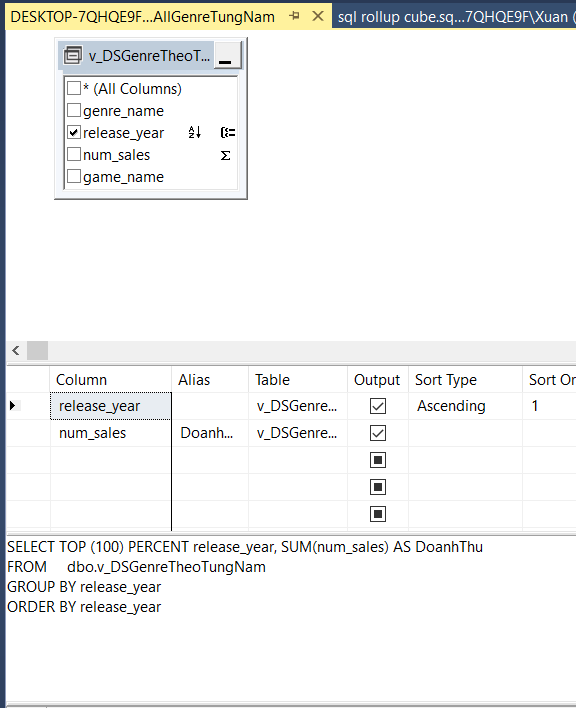
**v\_DoanhThuTungLoaiGenre**



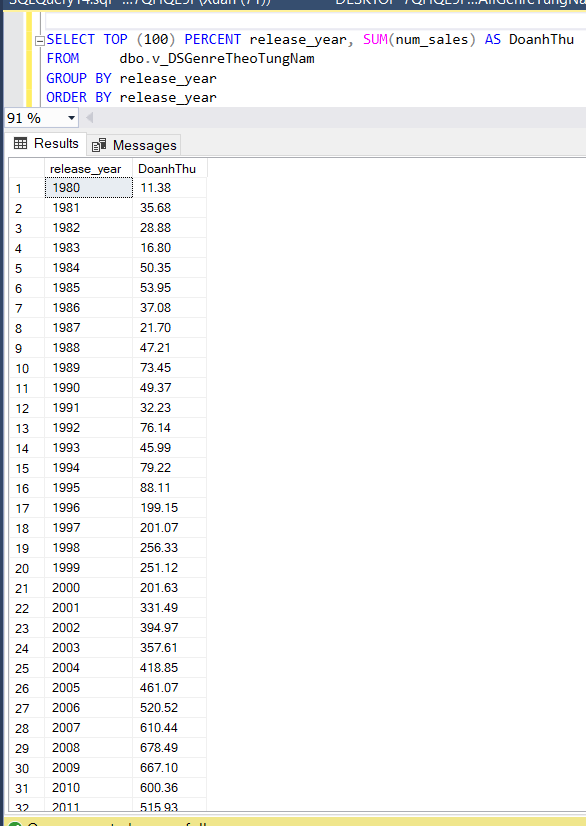
|  |
| --- |
| SELECT genre\_name, SUM(num\_sales) AS DoanhThuGenre  FROM dbo.v\_DSGenreTheoTungNam  GROUP BY genre\_name |



**Vv\_DTAllGenreTungNam**

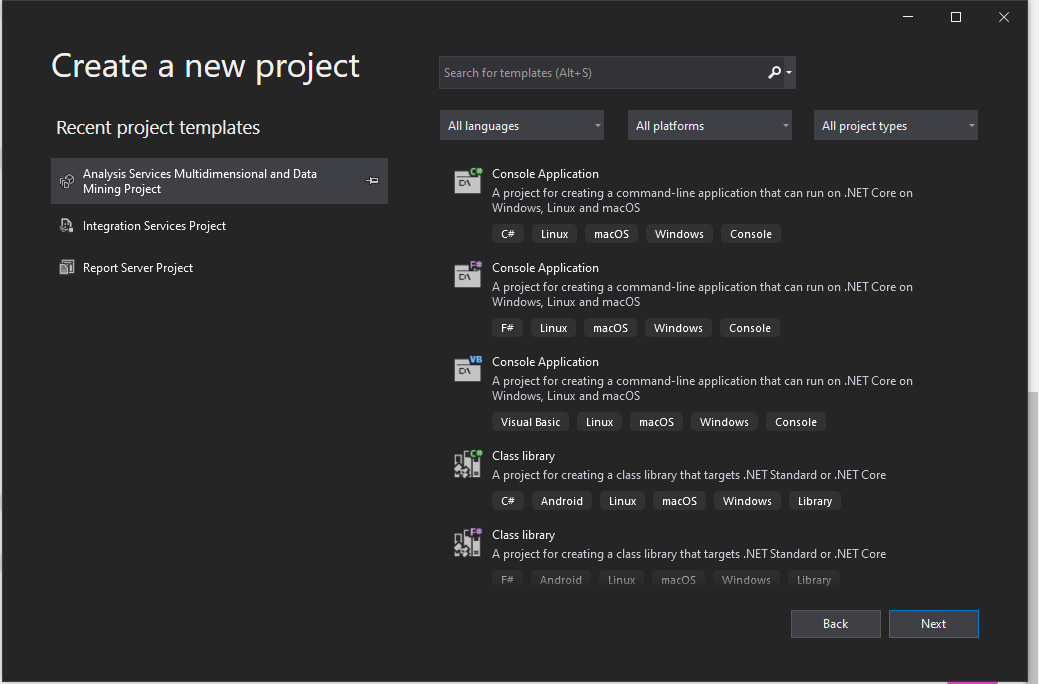


|  |
| --- |
| SELECT TOP (100) PERCENT release\_year, SUM(num\_sales) AS DoanhThu  FROM dbo.v\_DSGenreTheoTungNam  GROUP BY release\_year  ORDER BY release\_year |

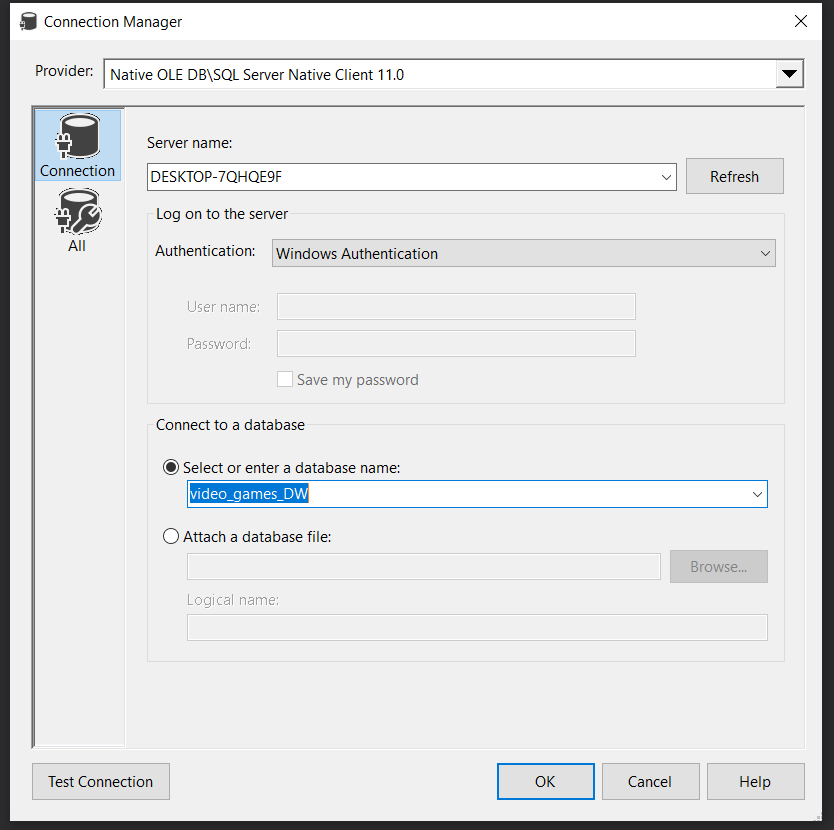


* + 1. Triển khai phân hệ DWH lên BIDS

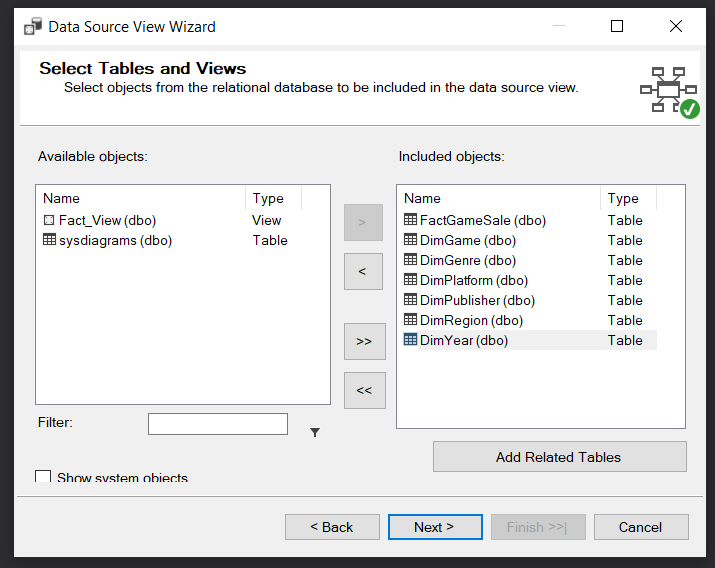
**B1: Sử dụng BIDS  để tạo view :**

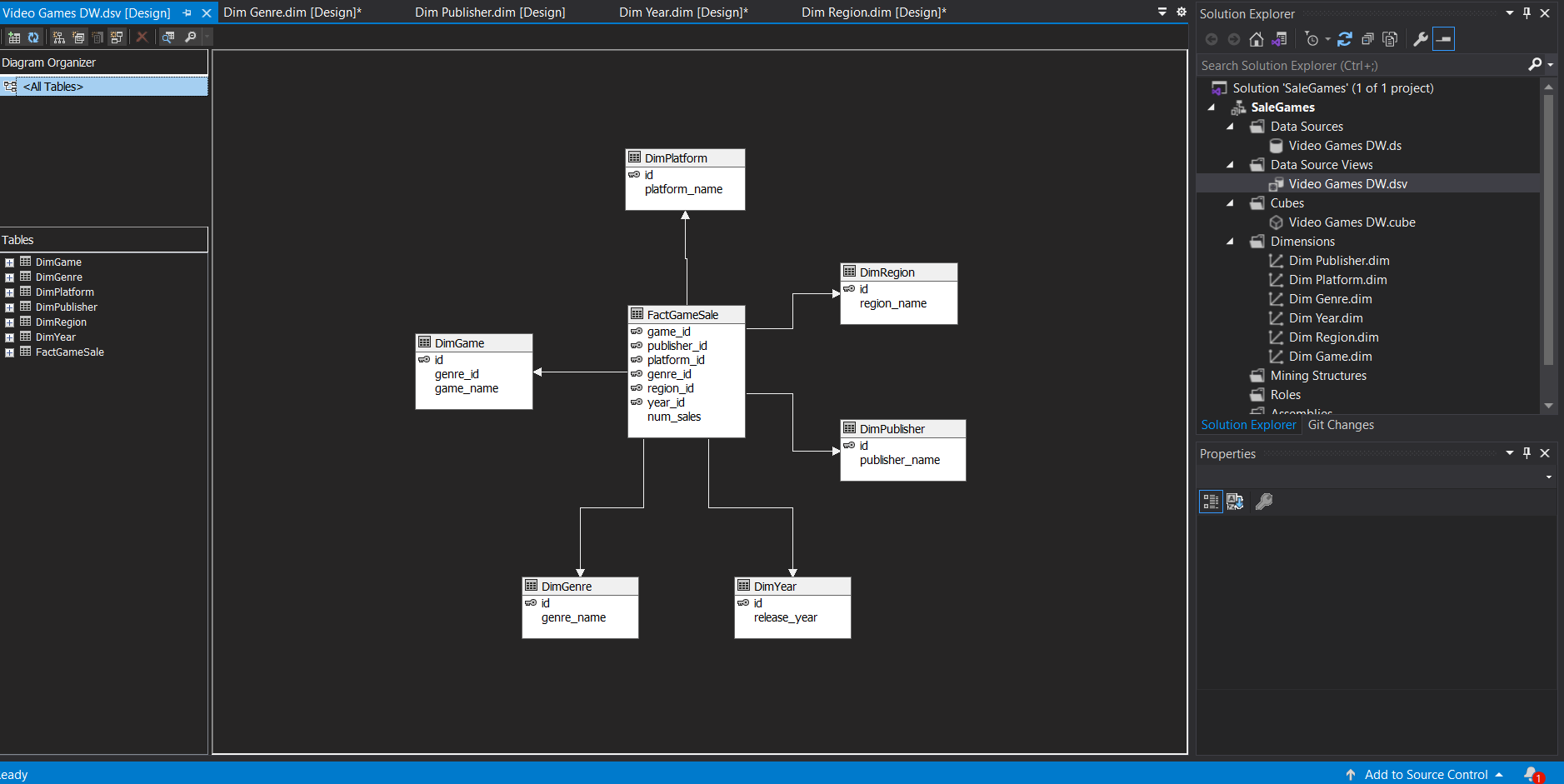


**B2. Chọn data source là video\_games\_DW**

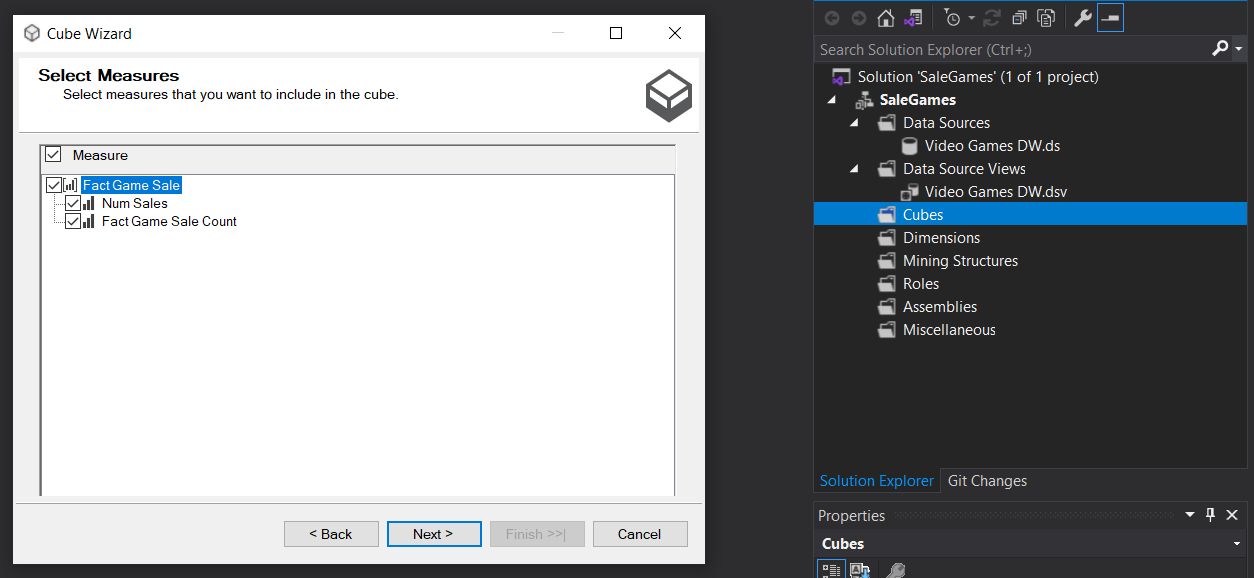


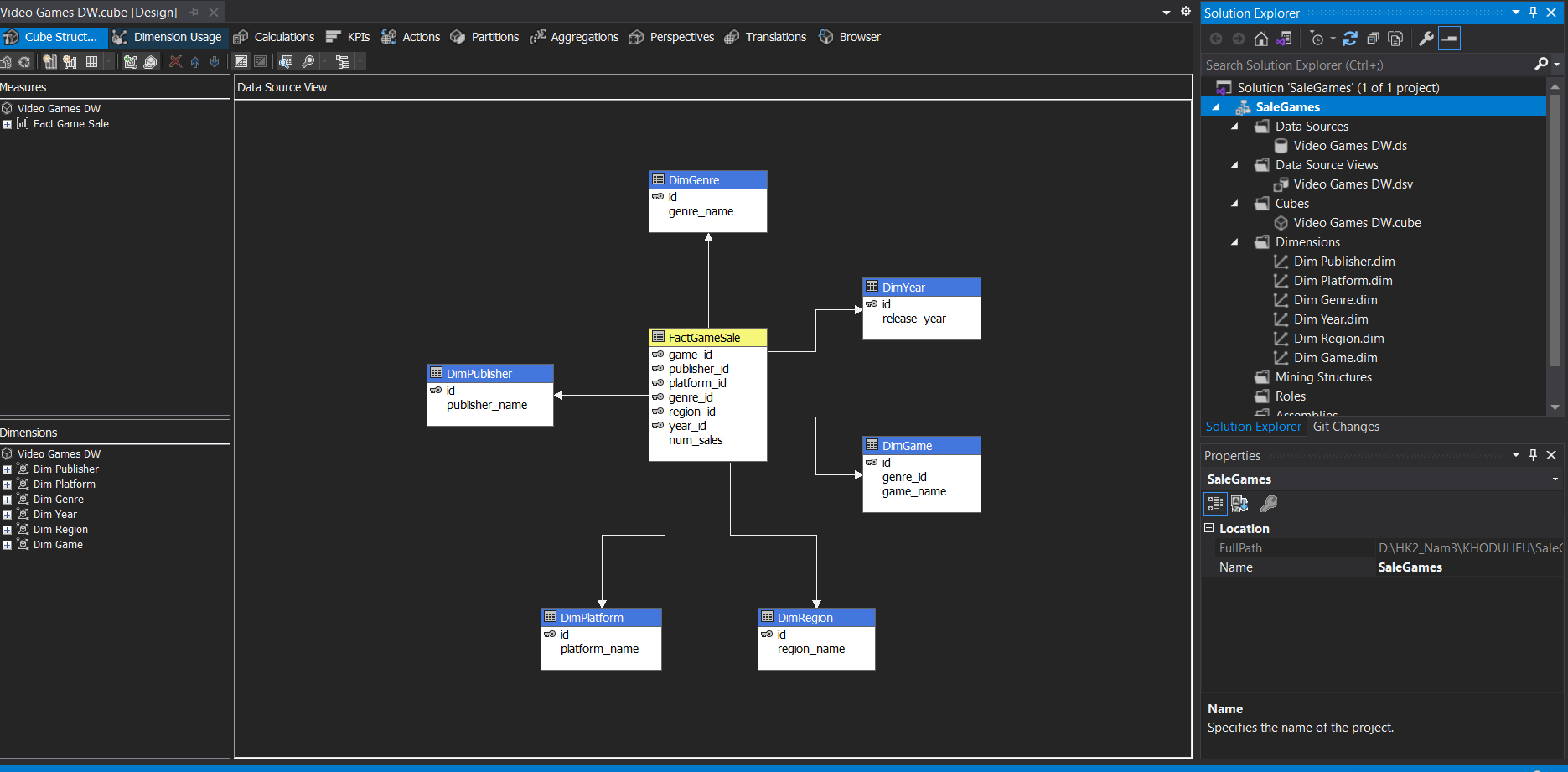
**B3. Tạo source view :**



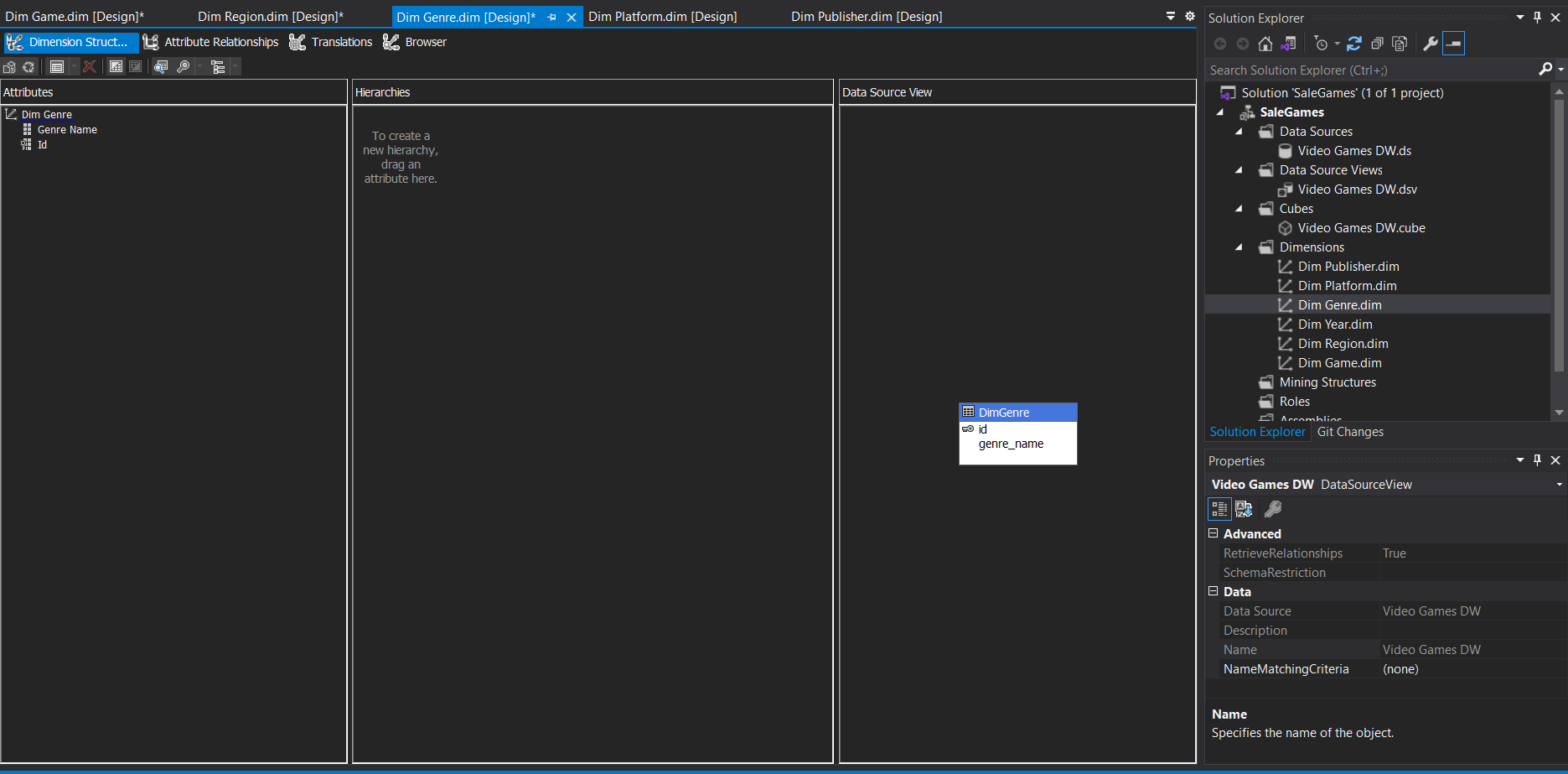


**B4. Tạo cube :**

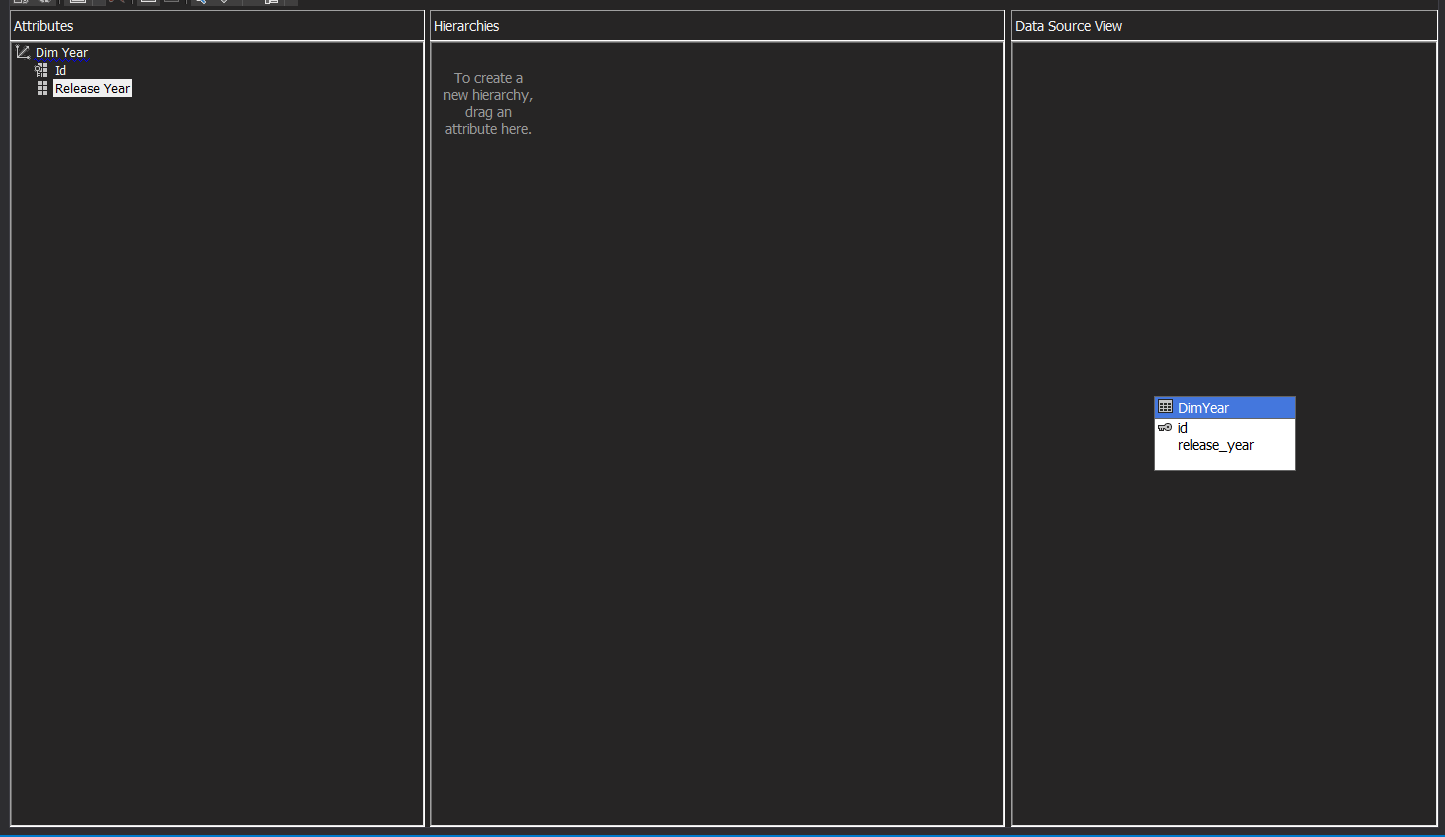




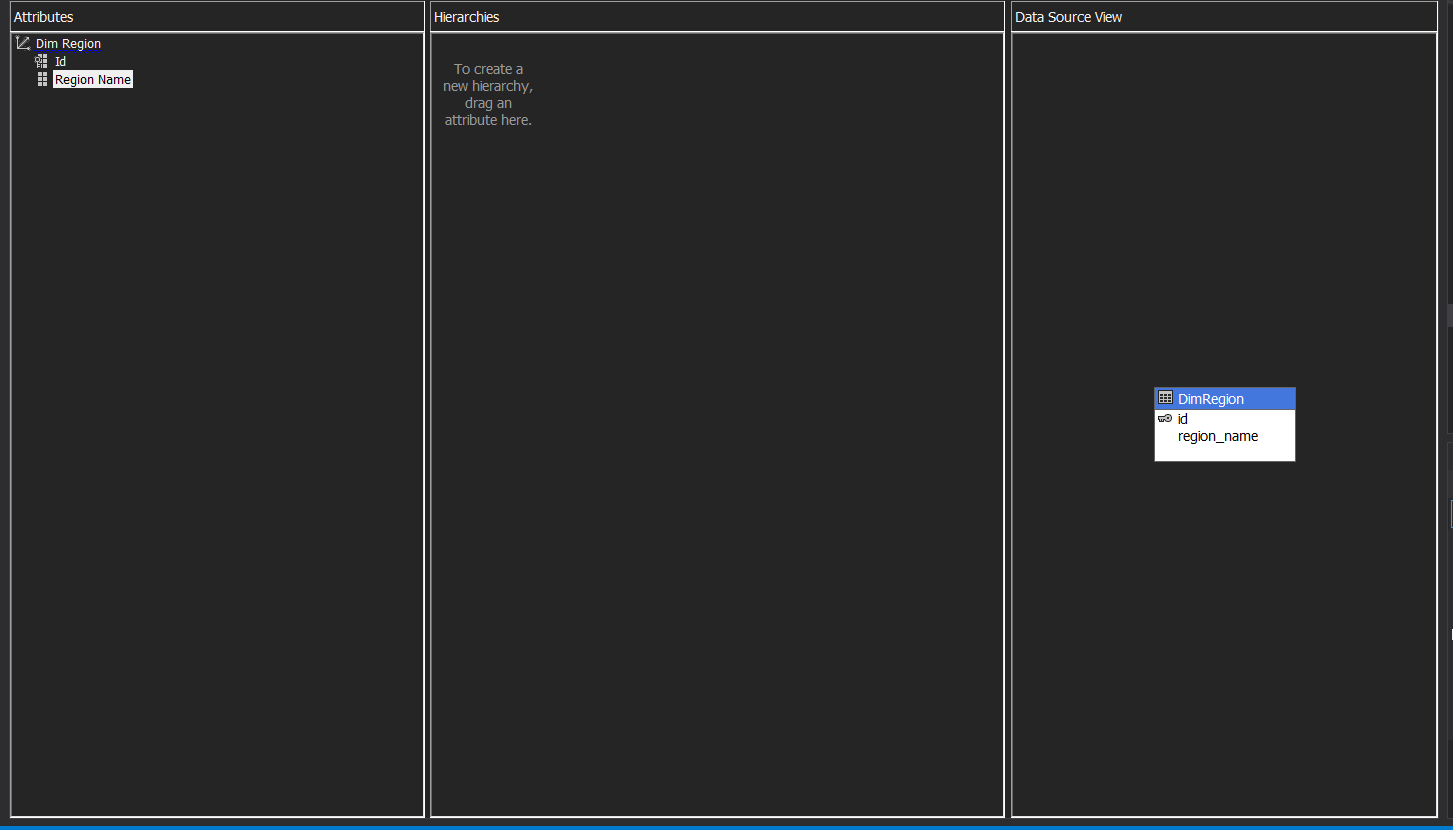
**Chọn thuộc tính Genre\_name trong DimGenre**



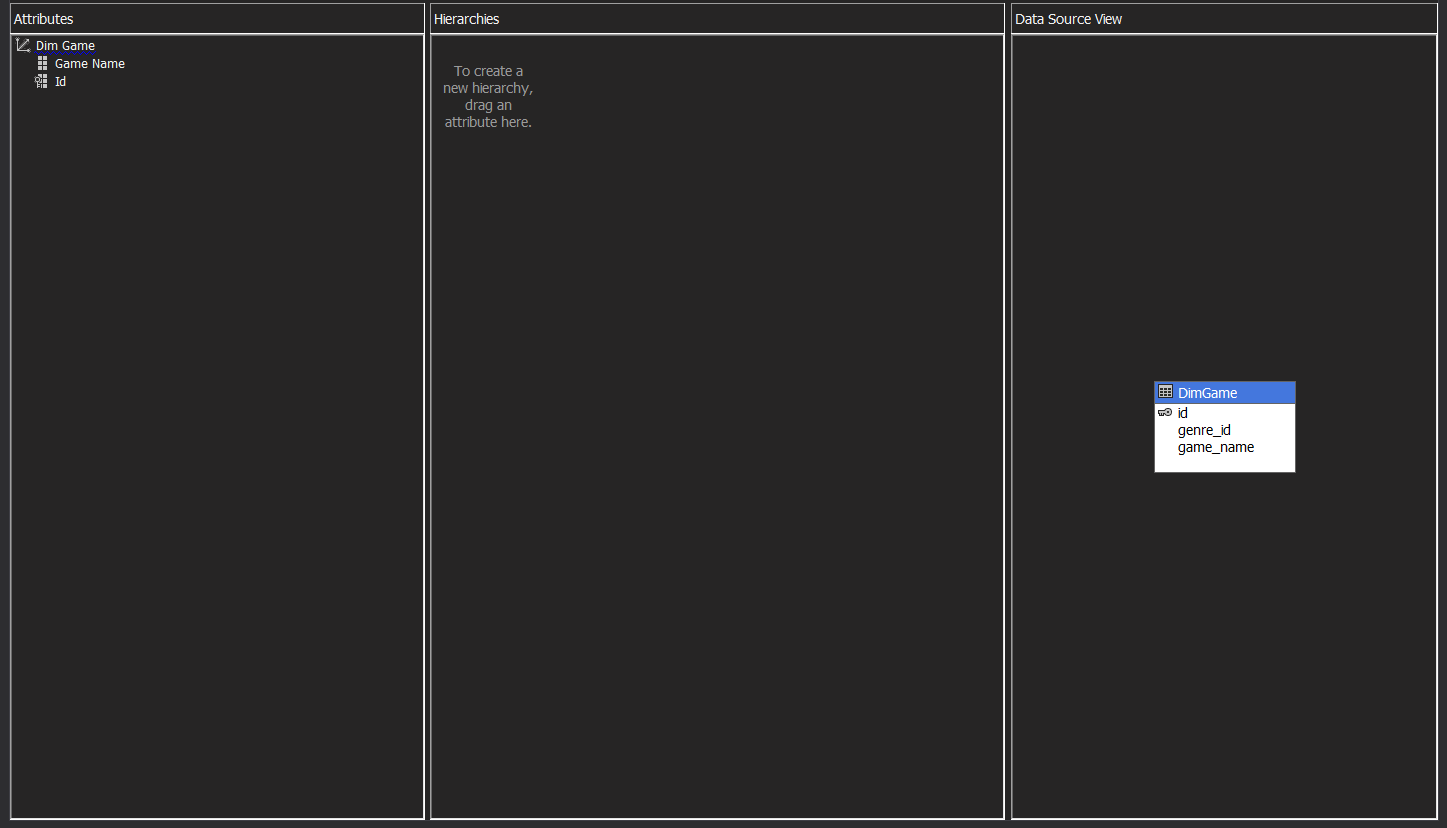
Chọn thuộc tính Release\_Year trong DimYear



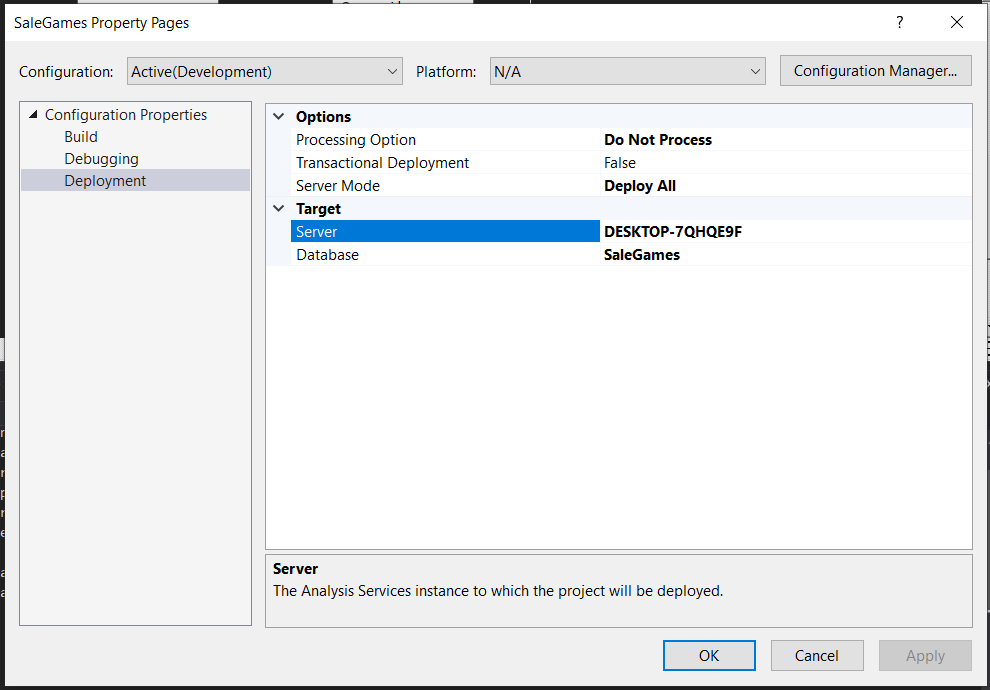
Chọn thuộc tính Region\_name trong DimRegion

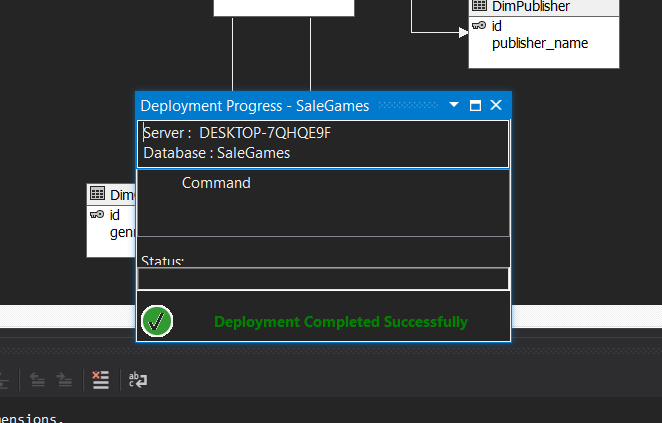


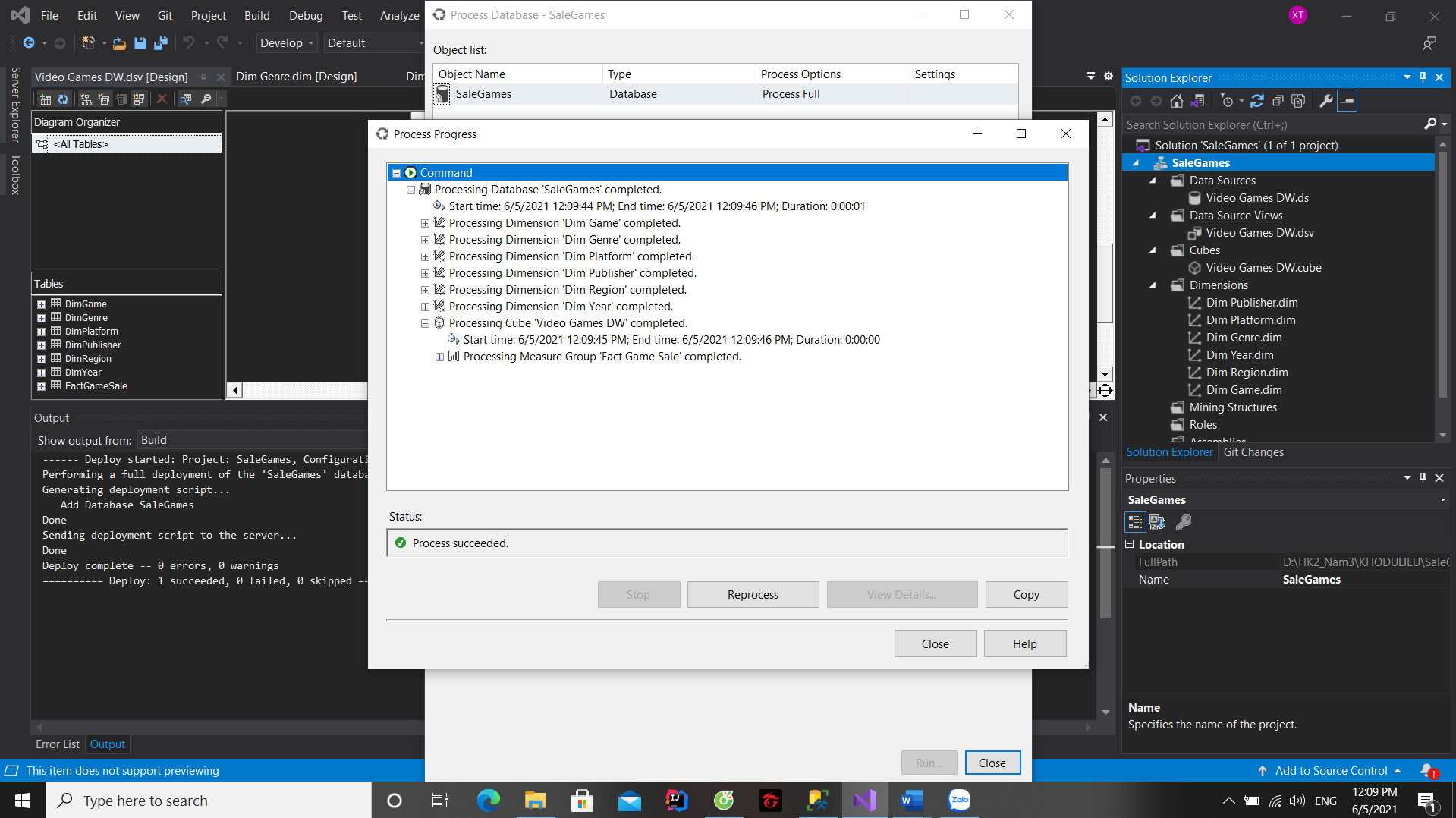
Chọn thuộc tính Game\_name trong DimGame

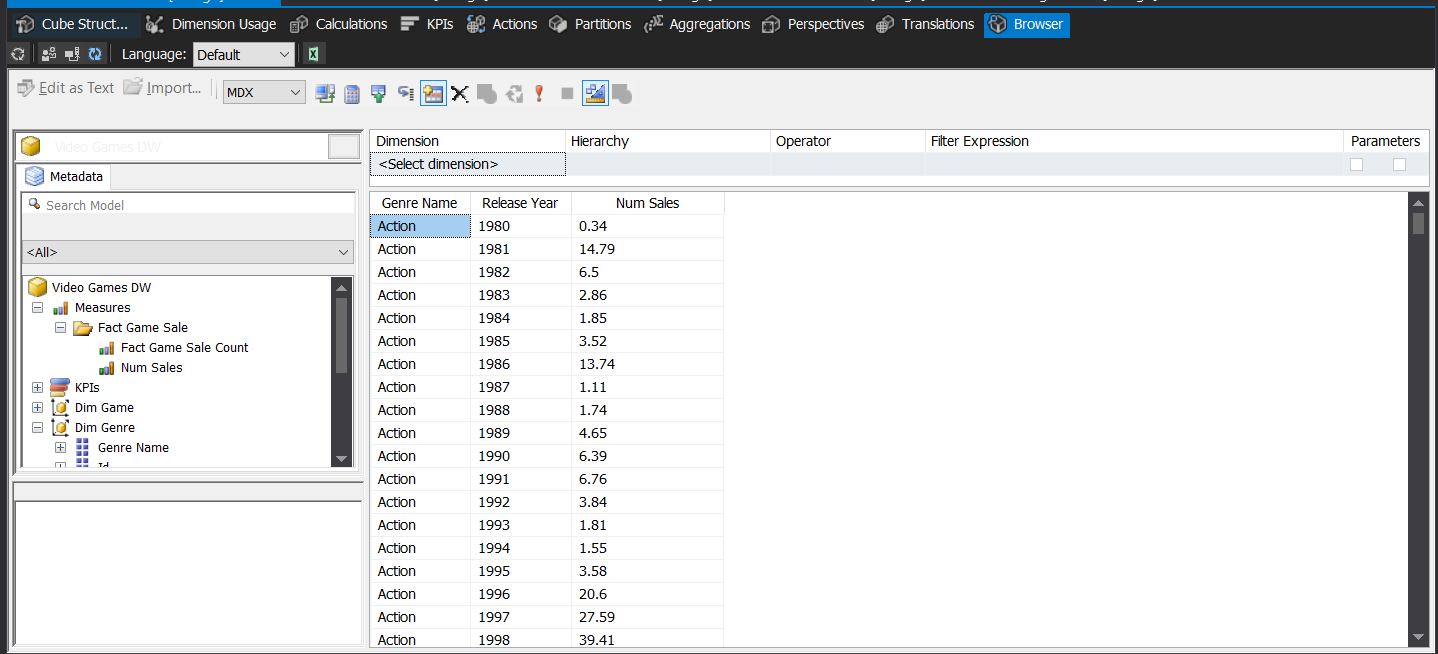


**B6. Deloy và process thành công**





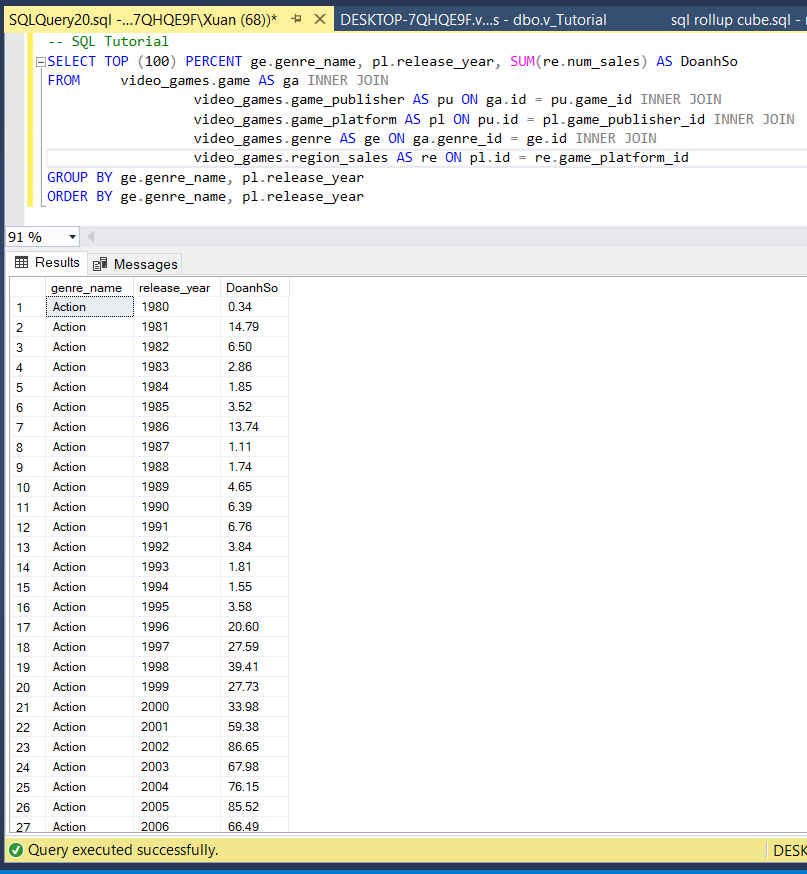




* + 1. Khai thác sử dụng phân hệ DWH dùng các vấn tin SQL

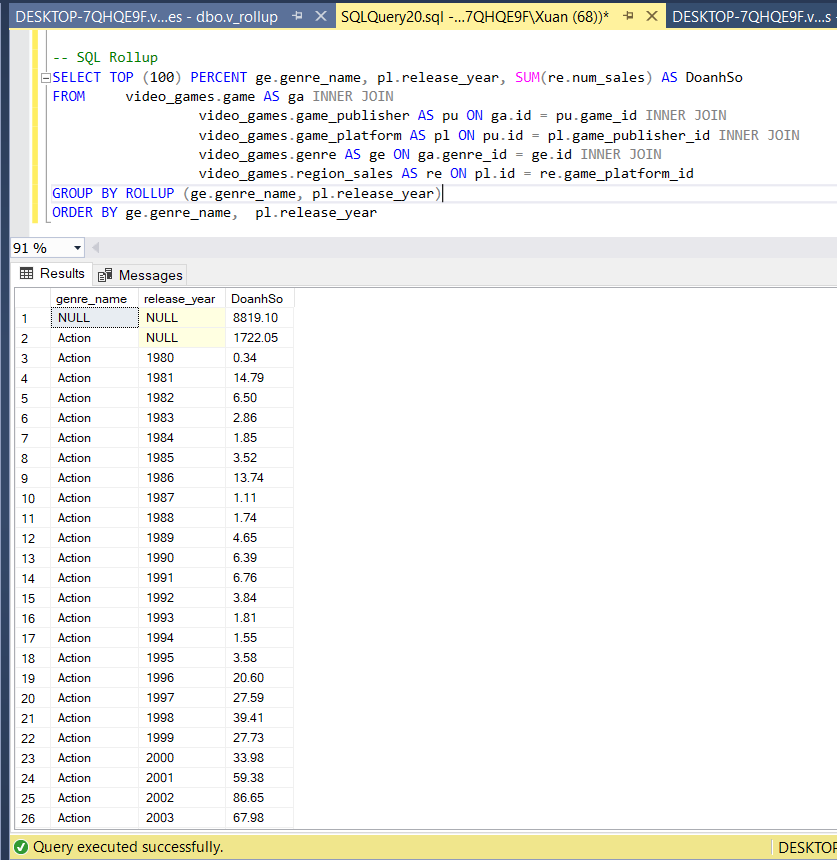
Câu hỏi phân tích 1: Cho biết Doanh Số của các thể loại game theo từng năm ?

|  |
| --- |
| -- SQL Tutorial  SELECT TOP (100) PERCENT ge.genre\_name, pl.release\_year, SUM(re.num\_sales) AS DoanhSo  FROM video\_games.game AS ga INNER JOIN  video\_games.game\_publisher AS pu ON ga.id = pu.game\_id INNER JOIN  video\_games.game\_platform AS pl ON pu.id = pl.game\_publisher\_id INNER JOIN  video\_games.genre AS ge ON ga.genre\_id = ge.id INNER JOIN  video\_games.region\_sales AS re ON pl.id = re.game\_platform\_id  GROUP BY ge.genre\_name, pl.release\_year  ORDER BY ge.genre\_name, pl.release\_year |



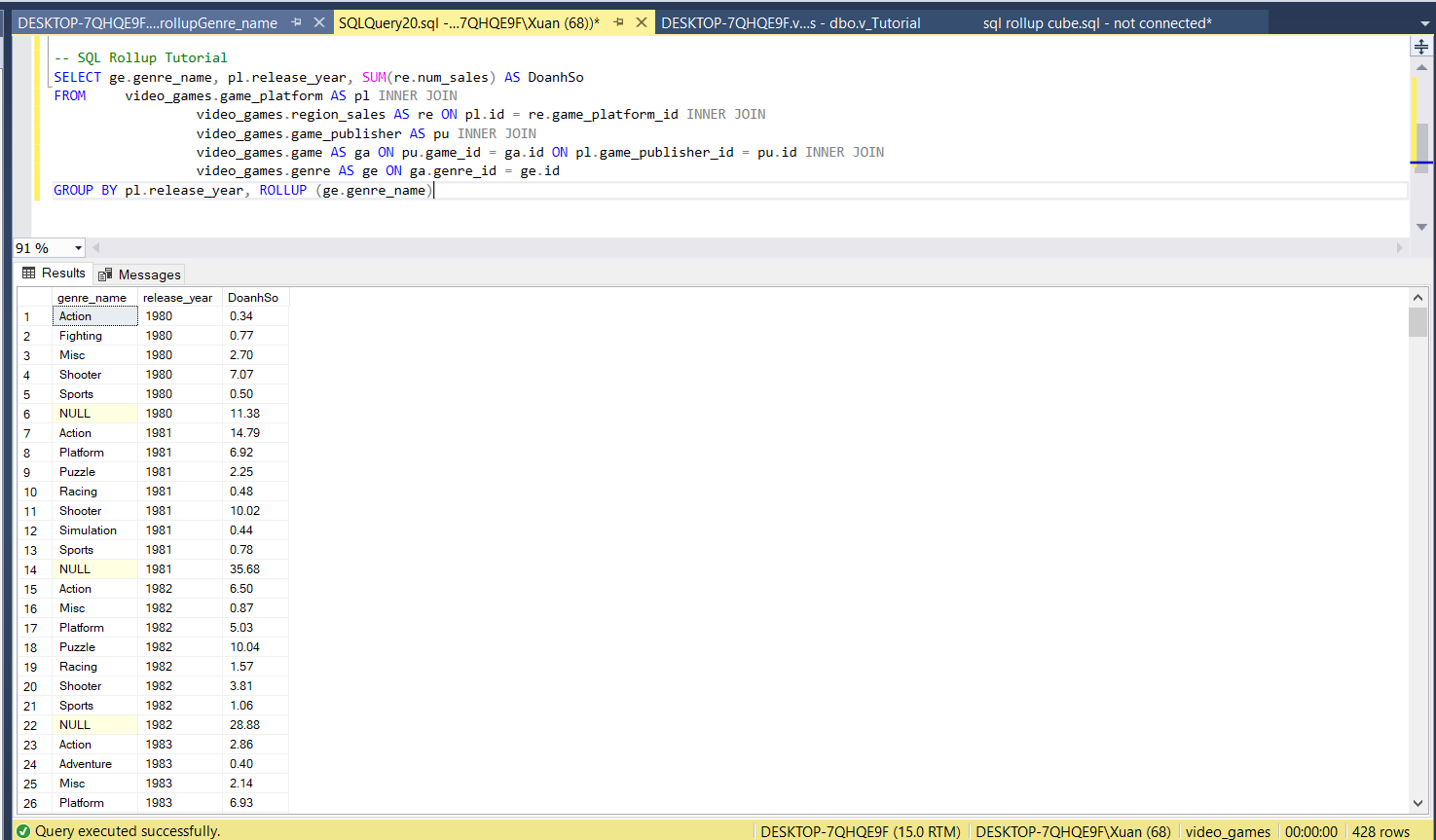
Câu hỏi phân tích 2: Cho biết Doanh thu của các thể loại game theo từng năm và tổng doanh thu của toàn thị trường?

|  |
| --- |
| -- SQL Rollup  SELECT TOP (100) PERCENT ge.genre\_name, pl.release\_year, SUM(re.num\_sales) AS DoanhSo  FROM video\_games.game AS ga INNER JOIN  video\_games.game\_publisher AS pu ON ga.id = pu.game\_id INNER JOIN  video\_games.game\_platform AS pl ON pu.id = pl.game\_publisher\_id INNER JOIN  video\_games.genre AS ge ON ga.genre\_id = ge.id INNER JOIN  video\_games.region\_sales AS re ON pl.id = re.game\_platform\_id  GROUP BY ROLLUP (ge.genre\_name, pl.release\_year)  ORDER BY ge.genre\_name, pl.release\_year |



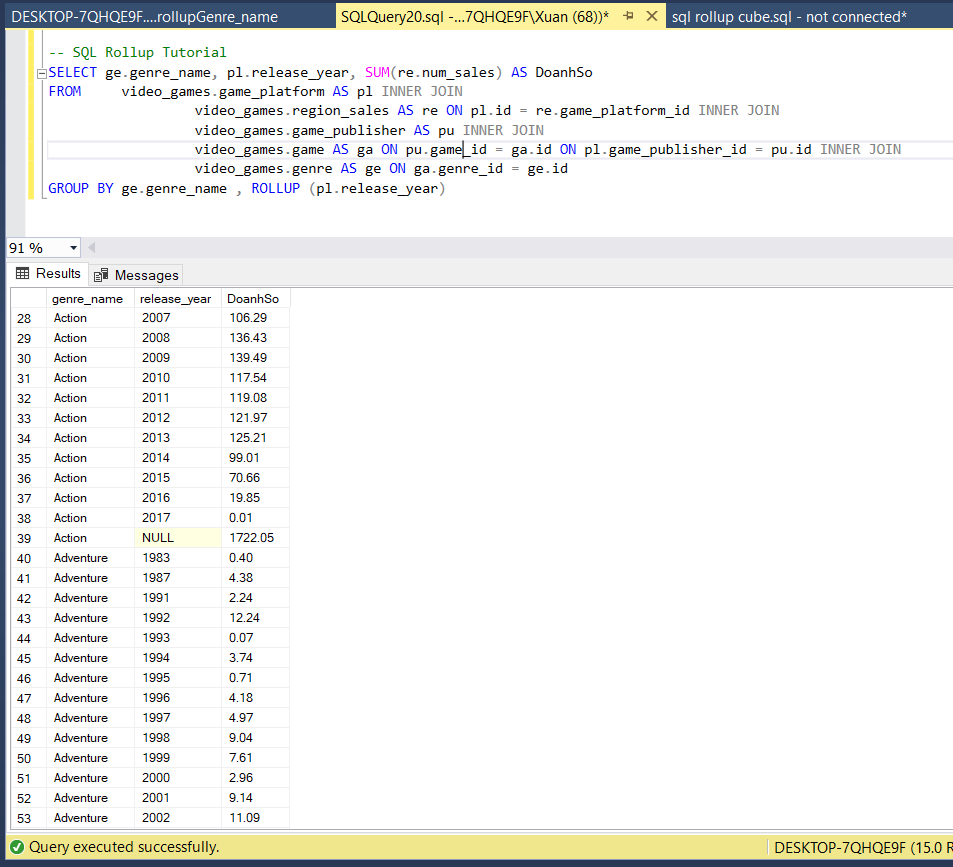
+ Tổng doanh thu của tất cả genre theo từng năm

|  |
| --- |
| -- SQL Rollup Tutorial  SELECT ge.genre\_name, pl.release\_year, SUM(re.num\_sales) AS DoanhSo  FROM video\_games.game\_platform AS pl INNER JOIN  video\_games.region\_sales AS re ON pl.id = re.game\_platform\_id INNER JOIN  video\_games.game\_publisher AS pu INNER JOIN  video\_games.game AS ga ON pu.game\_id = ga.id ON pl.game\_publisher\_id = pu.id INNER JOIN  video\_games.genre AS ge ON ga.genre\_id = ge.id  GROUP BY pl.release\_year, ROLLUP (ge.genre\_name) |



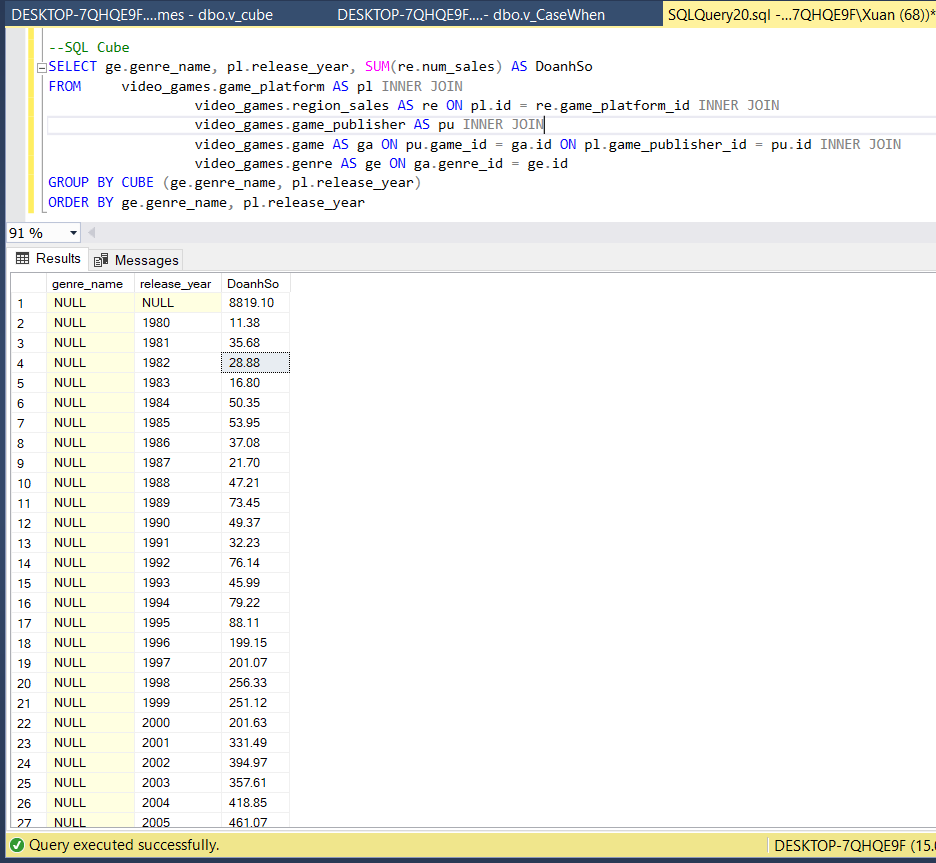
+ Tổng doanh thu của từng thể loại game qua tất cả các năm

|  |
| --- |
| -- SQL Rollup Tutorial  SELECT ge.genre\_name, pl.release\_year, SUM(re.num\_sales) AS DoanhSo  FROM video\_games.game\_platform AS pl INNER JOIN  video\_games.region\_sales AS re ON pl.id = re.game\_platform\_id INNER JOIN  video\_games.game\_publisher AS pu INNER JOIN  video\_games.game AS ga ON pu.game\_id = ga.id ON pl.game\_publisher\_id = pu.id INNER JOIN  video\_games.genre AS ge ON ga.genre\_id = ge.id  GROUP BY ge.genre\_name , ROLLUP (pl.release\_year) |

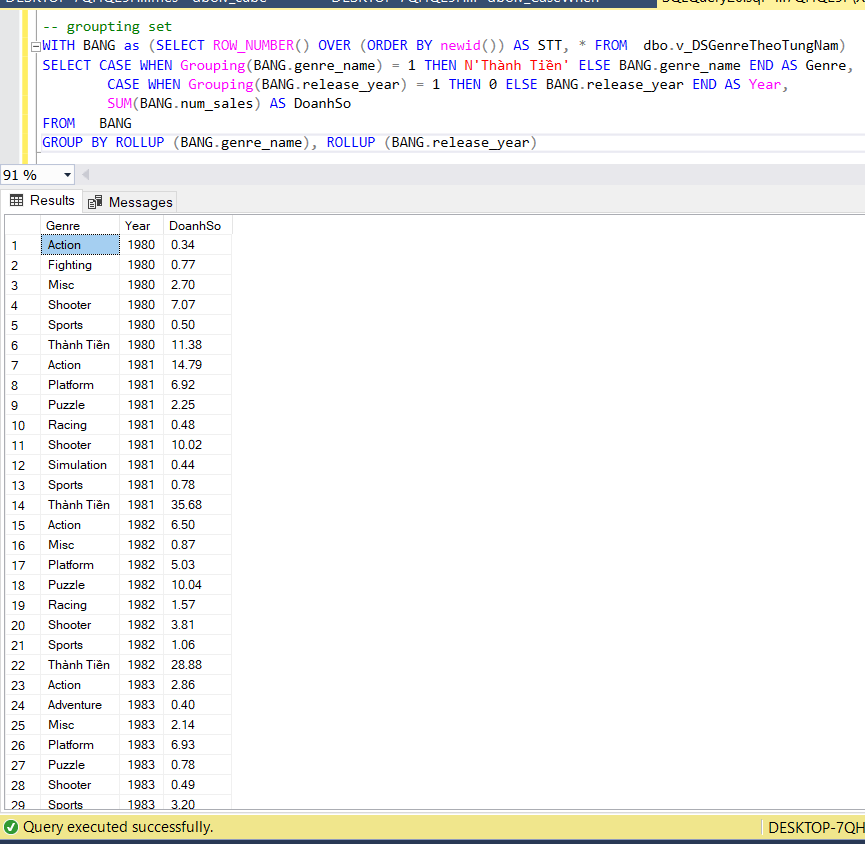


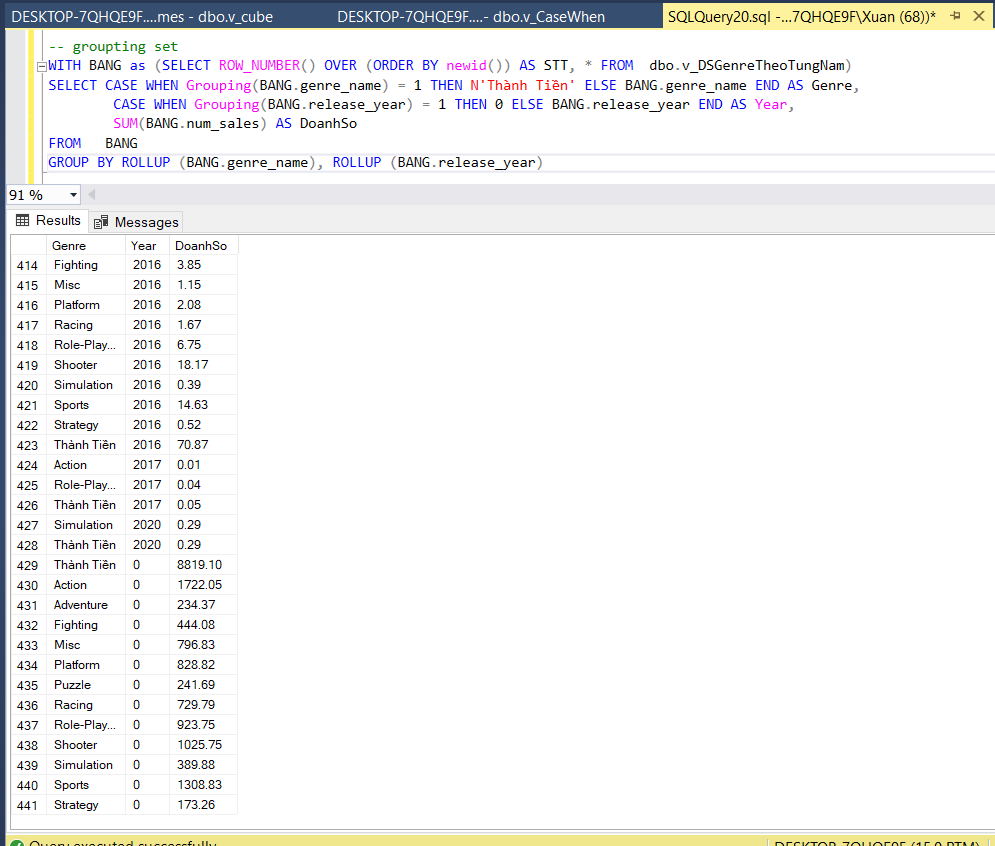
Câu hỏi phân tích 3: Cho biết Doanh thu của các thể loại game từng năm, tổng doanh của toàn thị trường theo từng năm, tổng doanh thu của từng thể loại game, tổng doanh thu của toàn thị trường?

|  |
| --- |
| --SQL Cube  SELECT ge.genre\_name, pl.release\_year, SUM(re.num\_sales) AS DoanhSo  FROM video\_games.game\_platform AS pl INNER JOIN  video\_games.region\_sales AS re ON pl.id = re.game\_platform\_id INNER JOIN  video\_games.game\_publisher AS pu INNER JOIN  video\_games.game AS ga ON pu.game\_id = ga.id ON pl.game\_publisher\_id = pu.id INNER JOIN  video\_games.genre AS ge ON ga.genre\_id = ge.id  GROUP BY CUBE (ge.genre\_name, pl.release\_year)  ORDER BY ge.genre\_name, pl.release\_year |



|  |
| --- |
| -- groupting set  WITH BANG as (SELECT ROW\_NUMBER() OVER (ORDER BY newid()) AS STT, \* FROM dbo.v\_DSGenreTheoTungNam)  SELECT CASE WHEN Grouping(BANG.genre\_name) = 1 THEN N'Thành Tiền' ELSE BANG.genre\_name END AS Genre,  CASE WHEN Grouping(BANG.release\_year) = 1 THEN 0 ELSE BANG.release\_year END AS Year,  SUM(BANG.num\_sales) AS DoanhSo  FROM BANG  GROUP BY ROLLUP (BANG.genre\_name), ROLLUP (BANG.release\_year) |

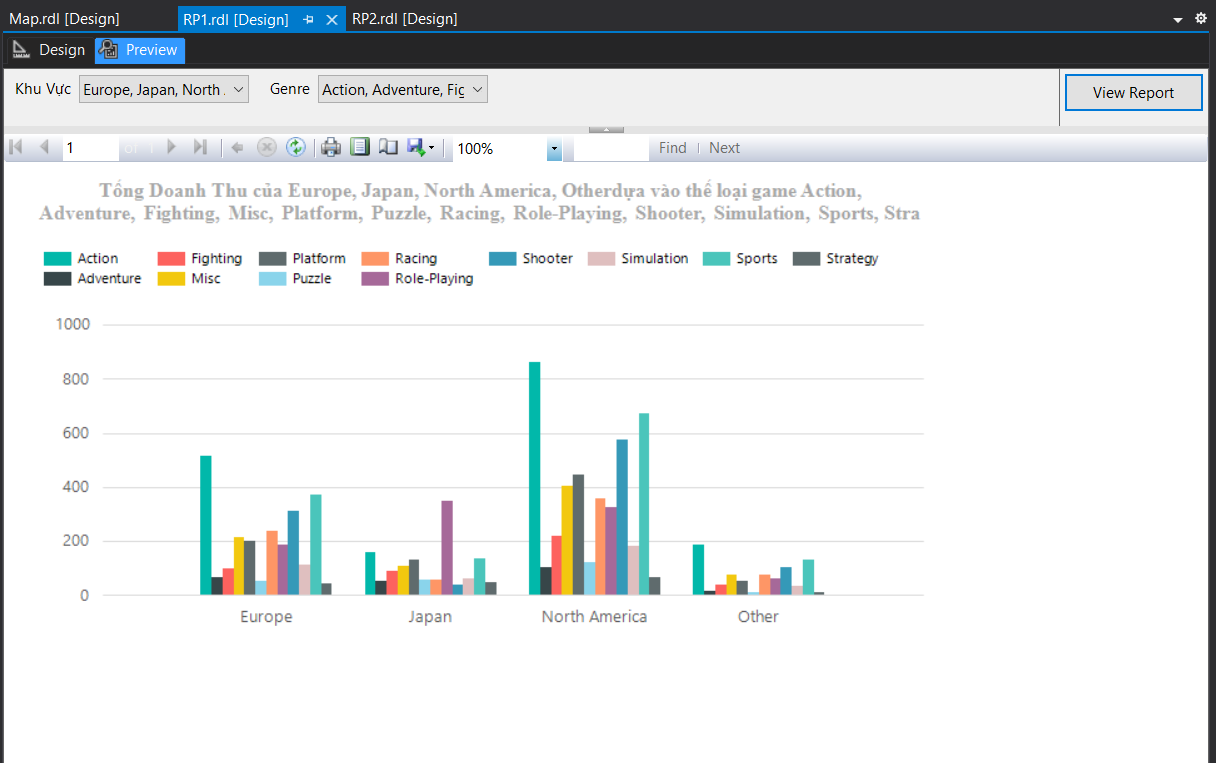




* + 1. Trực quan hóa dữ liệu với SSRS

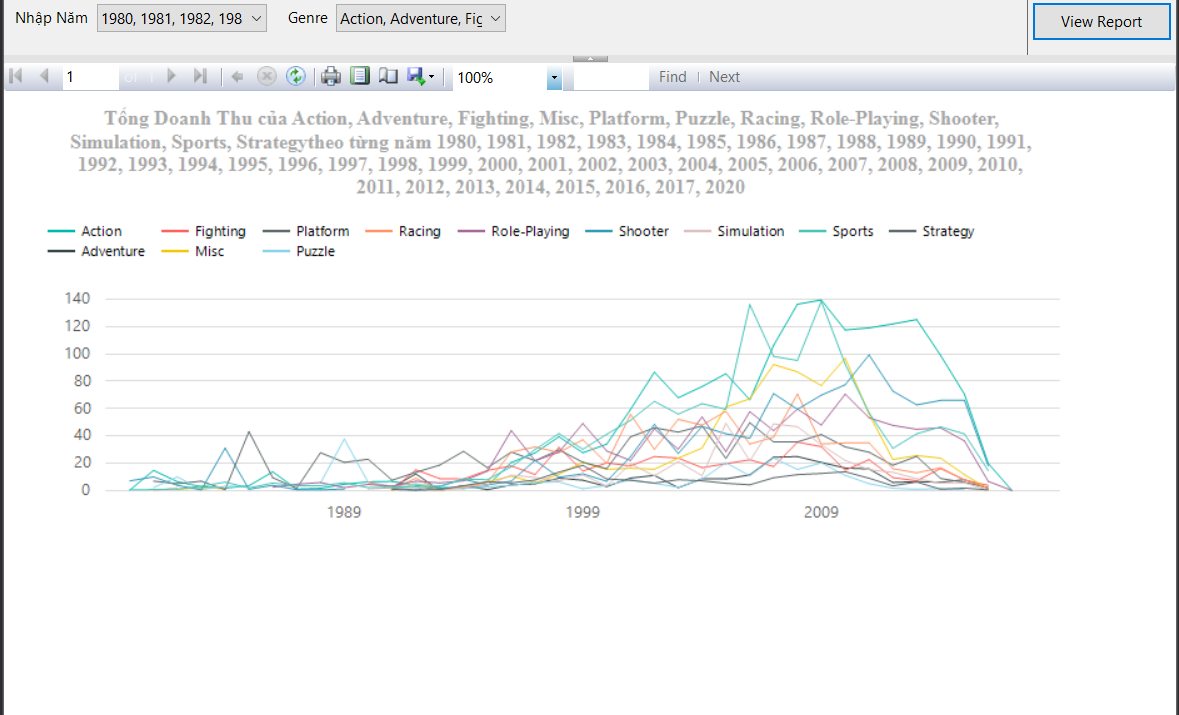
**Bài toán 1:Thống kê doanh thu của từng khu vực theo thể loại game qua các năm**

Với thống kê này ,để vẽ chart thì sẽ chọn các Dim :DimGenre và DimRegion và Fact



Ở biểu đồ thứ nhất cho thấy khu vực North America có tổng doanh thu với tất cả thể loại genre cao nhất so với 3 khu vực còn lại từ năm 1980 đến năm 2020. Ở khu vực NA thì doanh thu của thể loại game về action là chiếm tỉ lệ cao nhất. Qua biểu đồ đó cho thấy các nhà phát hành sẽ tập trung làm game ở khu vực NA để cho các người có nhu cầu sử dụng nhiều hơn.

**Bài toán 2:Thống kê doanh thu của Genre theo các khu vực**



Ở biểu đồ thứ hai cho thấy doanh khu của các thể loại game theo từng giai đoạn từ năm 1980 đến năm 2020. Biểu đồ cho thấy doanh thu của các thể loại game phát triển mạnh vào năm 2009. Và thể loại game Action chiếm tỉ lệ doanh thu cao nhất so với tất cả những thể loại game còn lại

1. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN
   1. Kết luận
      1. Những kết quả đạt được

ĐỒ ÁN HỌC PHẦN đã giải quyết được vấn đề gì. Gạch đầu dòng nói những ý đã làm được.

* + 1. Hạn chế

Nêu những vấn đề còn tồn tại.

Nêu nhưng nội dung THEO yêu cầu của ĐỒ ÁN HỌC PHẦN nhưng bản thân chưa thực hiện được, hoặc đã thực hiện, nhưng chưa tốt và chưa hài lòng.

Cố gắng tự khen mình, không nên chê, không “vạch áo cho người khác xem lưng” (nói khéo để bỏ qua vấn đề còn vướng mắc): nhưng KHÔNG khen ẩu

* 1. Hướng phát triển
     1. Hướng khắc phục các hạn chế

Nêu hướng khắc phục tất cả những tồn tại, hạn chế đã nêu trên (trong phần 5.1.2. của kết luận).

+ Có bao nhiêu hạn chế là bấy nhiêu hướng khắc phục; không bỏ qua hạn chế nào

+ Hướng khắc phục chỉ là giải pháp đề xuất, không nhất thiết đúng và cũng nhất thiết là thực hiện được (chỉ là giải pháp = phương pháp giải quyết vấn đề do cá nhân đề xuất); nhưng KHÔNG đề xuất những giải pháp mâu thuẫn, vô lý, quá phi thực tế, . . .

* + 1. Hướng mở rộng ĐỒ ÁN HỌC PHẦN

+ Hướng mở rộng để nghiên cứu thêm cho ĐỒ ÁN HỌC PHẦN.

+ (chú ý: khác với 5.1.2 hạn chế) : Nêu những nội dung có thể thực hiện thêm NGOÀI những yều cầu của ĐỒ ÁN HỌC PHẦN nhằm tăng giá trị của ĐỒ ÁN HỌC PHẦN về lý luận hoặc/và thực tiễn.

# DANH MỤC TÀI LIỆU THAM KHẢO

1. Võ Xuân Thể (2021), *Bài giảng học phần Kho dữ liệu*, TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM (lưu hành nội bộ).

# CÁC PHỤ LỤC

(nếu có)

# ĐÁNH GIÁ KẾT QUẢ HỌC TẬP HỌC PHẦN

ĐỒ ÁN HP THAY CHO BÀI THI KẾT THÚC HỌC PHẦN

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Hình thức KT** | **Nội dung** | **Chuẩn đầu ra đánh giá** | **Trình độ năng lực** | **Phương pháp đánh giá** | **Công cụ đánh giá** | **Tỉ lệ (%)** |
| **BÀI TẬP LẬP TRÌNH** | | | | | |  |
| BL#1 | Bài tập lập trình theo từng Buổi thực hành = tính bình quân  Kiểm tra viết | G1.1 | 3 | Bài tập  Kiểm tra | Bài tập  Kiểm tra | 50 |
| G2.1 | 3 |
| G3.1  G3.2 | 4  3 |
| G4.1 | 5 |
| **ĐỒ ÁN HỌC PHẦN (Project)** | | | | | |  |
| ĐA#1 | Phát triển một hệ thống KHO DỮ LIỆU đơn giản:  + Đề tài Nhóm có phân công phân hệ cụ thể cho từng thành viên thuộc nhóm  + Xây dựng và thực nghiệm Hệ thống Kho dữ liệu theo nhóm và phân hệ cá nhân  => báo cáo vào ngày thi. | G1.1 | 3 | Báo cáo, Demo và thuyết trình | Rubric | 50 |
| G2.1 | 3 |
| G3.1 | 4 |
| G3.2  G4.1 | 3  5 |
| G2.1 | 4 |
| G3.1 | 4 |
| G3.2  G4.1 | 3  5 |