

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**



Giảng viên hướng dẫn: ThS. Nguyễn Thị Kim Phụng

Sinh viên thực hiện:

Đặng Minh Quân 19520867

**BÁO CÁO ĐỒ ÁN**

**KHO DỮ LIỆU VÀ OLAP**

**ĐỀ TÀI**

**Xây dựng kho dữ liệu Movie Industry**

**TP. Hồ Chí Minh, tháng 6 năm 2022**

**LỜI CẢM ƠN**

Trên thực tế không có sự thành công nào mà không gắn liền với những sự hỗ trợ, giúp đỡ dù ít hay nhiều, dù trực tiếp hay gián tiếp của người khác. Với lòng biết ơn sâu sắc nhất, đầu tiên em xin gởi lời cảm ơn chân thành đến tập thể quý Thầy Cô Trường Đại học Công nghệ thông tin – Đại học Quốc gia TP.HCM và quý Thầy Cô khoa Hệ thống thông tin đã giúp cho em có những kiến thức cơ bản làm nền tảng để thực hiện đề tài này.

Đặc biệt em xin gửi lời cảm ơn chân thành tới Cô Nguyễn Thị Kim Phụng - giảng viên lý thuyết môn Kho dữ liệu và OLAP đã tận tình giúp đỡ, trực tiếp chỉ bảo, hướng dẫn em trong suốt quá trình làm đồ án môn học. Nhờ đó, em đã tiếp thu được nhiều kiến thức bổ ích trong việc vận dụng cũng như kỹ năng làm đồ án. Nếu không có những lời hướng dẫn, dạy bảo của cô thì em nghĩ đồ án này rất khó có thể hoàn thiện được. Một lần nữa, em xin chân thành cảm ơn Cô.

Xuất phát từ mục đích học tập tìm hiểu về kho dữ liệu, phương pháp xây dựng kho dữ và phân tích dữ liệu trên kho dữ liệu Movie Industry. Dựa trên những kiến thức được Cô cung cấp trên trường kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới, em đã cố gắng thực hiện đồ án một cách tốt nhất. Trong thời gian một học kỳ thực hiện đề tài, em đã vận dụng những kiến thức nền tảng đã tích lũy đồng thời kết hợp với việc học hỏi và nghiên cứu những kiến thức mới. Từ đó, em vận dụng tối đa những gì đã thu thập được để hoàn thành một báo cáo đồ án tốt nhất. Tuy nhiên, trong quá trình thực hiện, em không tránh khỏi những thiếu sót. Chính vì vậy, em rất mong nhận được những sự góp ý từ phía Cô nhằm hoàn thiện những kiến thức mà em đã học tập và là hành trang để em thực hiện tiếp các đề tài khác trong tương lai.

Thành phố Hồ Chí Minh, tháng 6 năm 2022

Sinh viên thực hiện

**NHẬN XÉT CỦA GIẢNG VIÊN**

**…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………**

**…………………………………………………………………………………………**

**MỤC LỤC**

[LỜI CẢM ƠN 1](#_Toc106053212)

[NHẬN XÉT CỦA GIẢNG VIÊN 2](#_Toc106053213)

[MỤC LỤC 3](#_Toc106053214)

[CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU 5](#_Toc106053215)

[1. PHÁT BIỂU VỀ DỮ LIỆU 5](#_Toc106053216)

[1.1. Mô tả về dữ liệu 5](#_Toc106053217)

[1.2. Thuộc tính của kho dữ liệu 5](#_Toc106053218)

[1.3. Kho dữ liệu đã xử lý 6](#_Toc106053219)

[2. XÂY DỰNG KHO DỮ LIỆU 11](#_Toc106053220)

[2.1. Sơ đồ bông tuyết minh hoạ 11](#_Toc106053221)

[2.2. Dim\_Genre 11](#_Toc106053222)

[2.3. Dim\_Rating 11](#_Toc106053223)

[2.4. Dim\_Country 11](#_Toc106053224)

[2.5. Dim\_Film 12](#_Toc106053225)

[2.6. Dim\_Time 12](#_Toc106053226)

[2.7. Fact\_Table 12](#_Toc106053227)

[3. NỘI DUNG 15 CÂU TRUY VẤN 13](#_Toc106053228)

[CHƯƠNG 2. TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS) 15](#_Toc106053229)

[1. TẠO MỚI MỘT DATABASE 15](#_Toc106053230)

[2. TẠO PROJECT SSIS TRONG VISUAL STUDIO 15](#_Toc106053231)

[3. TẠO BẢNG DIM VÀ BẢNG FACT 19](#_Toc106053232)

[3.1. Tạo các bảng Dim 19](#_Toc106053233)

[3.2. Tạo bảng Fact 34](#_Toc106053234)

[CHƯƠNG 3. PHÂN TÍCH DỮ LIỆU TRONG KHO (SSAS) 38](#_Toc106053235)

[1. TẠO PROJECT SSAS TRONG VISUAL STUDIO 38](#_Toc106053236)

[2. XÁC ĐỊNH DATA SOURCE 39](#_Toc106053237)

[3. XÁC ĐỊNH DATA SOURCE VIEW 42](#_Toc106053238)

[4. XÂY DỰNG CÁC CUBE VÀ CÁC MEASURE 45](#_Toc106053239)

[5. PHÂN CẤP TRONG BẢNG DIM 52](#_Toc106053240)

[5.1. Bảng Dim\_Time 52](#_Toc106053241)

[6. CHẠY PROJECT SSAS 59](#_Toc106053242)

[7. QUÁ TRÌNH PHÂN TÍCH DỮ LIỆU BẰNG NGÔN NGỮ MDX 62](#_Toc106053243)

[CHƯƠNG 4. QUÁ TRÌNH LẬP BÁO BIỂU (SSRS) 83](#_Toc106053244)

[1. QUÁ TRÌNH LẬP BÁO BIỂU BẰNG CÔNG CỤ MICROSOFT SSRS (Report Builder) 83](#_Toc106053245)

[1.1. Tạo project SSRS 83](#_Toc106053246)

[1.2. Tạo kết nối với kho dữ liệu 86](#_Toc106053247)

[1.3. Thực hiện lập báo biểu trên Report Builder 89](#_Toc106053248)

[CHƯƠNG 5. QUÁ TRÌNH DATA MINING 102](#_Toc106053249)

[1. QUÁ TRÌNH THỰC HIỆN DATA MINING VỚI THUẬT TOÁN LUẬT KẾT HỢP (MICROSOFT ASSOCIATION RULES) 102](#_Toc106053250)

[1.1. Thiết lập cấu trúc data mining 102](#_Toc106053251)

[1.2. Deploy project data mining 109](#_Toc106053252)

[1.3. Phân tích mining bằng luật kết hợp và đưa ra tập luật 112](#_Toc106053253)

[1.4. So sánh với các thuật toán khác 113](#_Toc106053254)

[CHƯƠNG 6. TÀI LIỆU THAM KHẢO 118](#_Toc106053255)

# CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU

1. PHÁT BIỂU VỀ DỮ LIỆU
   1. Mô tả về dữ liệu

* Kho dữ liệu Movie Industry là một kho dữ liệu thu nhập ngành công nghiệp điện ảnh thế giới từ năm 1980 – 2022.
* Thông qua kho dữ liệu người dùng có thể biết được thông tin về tên phim, đạo diễn, diễn viên, thông tin về kinh phi và doanh thu lợi nhuận mà phim đạt được, thông tin về số điểm rating của phim, …
* Kho dữ liệu được xây dựng với hướng chủ đề ngành công nghiệp điện ảnh (Movie Industry)
* Kho dữ liệu gồm 7668 dòng và 15 thuộc tính.
* Link dataset: <https://www.kaggle.com/datasets/danielgrijalvas/movies>
  1. Thuộc tính của kho dữ liệu

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa của thuộc tính |
| 1 | Name | Text | Tên của phim |
| 2 | Rating | Text | Nhãn xếp hạng của phim |
| 3 | Genre | Text | Thể loại chính của phim |
| 4 | Year | Number | Năm phát hành của phim |
| 5 | Released | Date | Ngày phát hành của phim |
| 6 | Score | Number | Số điểm đánh giá của phim |
| 7 | Votes | Number | Số lượng đánh giá của phim |
| 8 | Director | Text | Đạo diễn của phim |
| 9 | Writer | Text | Biên kịch của phim |
| 10 | Star | Text | Diễn viên chính của phim |
| 11 | Country | Text | Nước sản xuất của phim |
| 12 | Budget | Number | Kinh phí sản xuất của phim |
| 13 | Gross | Number | Doanh thu lợi nhuận của phim |
| 14 | Company | Text | Công ty sản xuất của phim |
| 15 | Runtime | Number | Thời lượng của phim |

* 1. Kho dữ liệu đã xử lý
* Ta sử dụng Jupyter Notebook xử lý ngôn ngữ Python để tiến hành xử lý dữ liệu
* Ta tiến hành import các thư viện cần thiết

Graphical user interface, text, application

Description automatically generated

Hình 1. 1 Import các thư viện

* Tiếp theo, ta đọc dữ liệu vào 1 dataframe

Graphical user interface, text, application, email

Description automatically generated

Hình 1. 2 Đọc dữ liệu

Table

Description automatically generated

Hình 1. 3 Kiểm tra dữ liệu đã đọc

Graphical user interface, table

Description automatically generated

Hình 1. 4 Xem xét các thuộc tính của dataset

* Ta thống kê các cột dữ liệu có giá trị null

Table

Description automatically generated with medium confidence

Hình 1. 5 Thống kê số lượng giá trị null của các thuộc tính

* Sau đó, ta tiến hành xoá những dữ liệu null đó để tránh ảnh hưởng tới quá trình phân tích

Graphical user interface, application

Description automatically generated

Hình 1. 6 Loại bỏ các giá trị NA

* Cuối cùng, ta tiến hành kiểm tra lại xem còn giá trị null nào không.

Text

Description automatically generated with medium confidence

Hình 1. 7 Kiểm tra lại dữ liệu sau khi xoá giá trị null

Graphical user interface, application

Description automatically generated

Hình 1. 8 Xem kích cỡ dữ liệu

* Sau khi xử lý dữ liệu, ta được 5421 dòng và 15 thuộc tính để sử dụng cho phân tích đề tài.

1. XÂY DỰNG KHO DỮ LIỆU
   1. Sơ đồ bông tuyết minh hoạ

Graphical user interface, diagram

Description automatically generated

Hình 1. 9 Sơ đồ bông tuyết minh hoạ

* 1. Dim\_Genre

|  |  |  |  |
| --- | --- | --- | --- |
| Khoá | Tên thuộc tính | Kiểu dữ liệu | Mô tả thuộc tính |
| Khoá chính | genre\_id | int | Mã thể loại |
|  | genre | varchar | Tên thể loại |

* 1. Dim\_Rating

|  |  |  |  |
| --- | --- | --- | --- |
| Khoá | Tên thuộc tính | Kiểu dữ liệu | Mô tả thuộc tính |
| Khoá chính | rating\_id | int | Mã nhãn dán |
|  | rating | varchar | Tên nhãn dán |

* 1. Dim\_Country

|  |  |  |  |
| --- | --- | --- | --- |
| Khoá | Tên thuộc tính | Kiểu dữ liệu | Mô tả thuộc tính |
| Khoá chính | country\_id | int | Mã quốc gia |
|  | country | varchar | Tên quốc gia |

* 1. Dim\_Film

|  |  |  |  |
| --- | --- | --- | --- |
| Khoá | Tên thuộc tính | Kiểu dữ liệu | Mô tả thuộc tính |
| Khoá chính | film\_id | int | Mã bộ phim |
| Khoá ngoại | country\_id | int | Mã quốc gia |
| Khoá ngoại | rating\_id | int | Mã nhãn dán |
| Khoá ngoại | genre\_id | int | Mã thể loại |
|  | name | varchar | Tên bộ phim |
|  | director | varchar | Đạo diễn |
|  | writer | varchar | Biên kịch |
|  | star | varchar | Diễn viên chính |
|  | company | varchar | Công ty sản xuất |

* 1. Dim\_Time

|  |  |  |  |
| --- | --- | --- | --- |
| Khoá | Tên thuộc tính | Kiểu dữ liệu | Mô tả thuộc tính |
| Khoá chính | released | Datetime | Ngày khởi chiếu |
|  | year | int | Năm |
|  | month | int | Tháng |
|  | day | int | Ngày |
|  | quarter | int | Quý |

* 1. Fact\_Table

|  |  |  |  |
| --- | --- | --- | --- |
| Khoá | Tên thuộc tính | Kiểu dữ liệu | Mô tả thuộc tính |
| Khoá chính | fact\_id | int | Mã fact |
| Khoá ngoại | released | Datetime | Ngày khởi chiếu |
| Khoá ngoại | film\_id | int | Mã bộ phim |
|  | score | float | Số điểm |
|  | votes | int | Số lượng đánh giá |
|  | budget | int | Kinh phí |
|  | gross | int | Doanh thu |
|  | runtime | int | Thời lượng |

1. NỘI DUNG 15 CÂU TRUY VẤN

a. Cho biết tổng doanh thu đạt được của các bộ phim ở nước Mĩ (United State) theo từng năm.

b. Cho biết thông tin bộ phim đạt được doanh thu cao nhất trong quý 2 năm 2016 của nước Anh (United Kingdom).

c. Đếm số lượng bộ phim được sản xuất theo từng quốc gia.

d. Liệt kê bộ phim có kinh phí đầu tư thấp nhất và bộ phím có kinh phí đầu tư cao nhất và in ra số điểm của 2 bộ phim đó.

e. Thống kê kinh phí đầu tư, doanh thu và thông tin các bộ phim thành công (kinh phí < doanh thu).

f. Đưa ra top 3 thể loại có số lượng phim nhiều nhất trong năm 2016.

g. Tìm ra đạo diễn có bộ phim đạt doanh thu cao nhất.

h. Đưa ra top 3 bộ phim được yêu thích nhất 3 năm gần đây (dựa theo số điểm).

i. Thống kê các bộ phim có thời lượng chiếu trên 180 phút và có số điểm đạt được trên 8.

j. Theo từng tháng, quý, năm liệt kê số lượng bộ phim được khởi chiếu trong 3 năm gần đây nhất.

k. Liệt kê top 5 những bộ phim có kinh phí đầu tư cao nhất hoặc top 3 có số điểm cao nhất.

l. Đưa ra top 3 diễn viên có số lượng bộ phim được vào vai nhiều nhất.

m. Liệt kê những bộ phim có doanh thu từ $1000000 đến $3000000.

n. Đưa ra 3 thể loại được quan tâm nhất (dựa theo số lượng vote).

o. Đưa ra top 3 bộ phim được yêu thích nhất mọi thời đại

# CHƯƠNG 2. TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

1. TẠO MỚI MỘT DATABASE

* Mở SQL Server và thực hiện tạo một databasse để lưu datawarehouse với tên là Movie\_Industry.

Graphical user interface, text, application, email

Description automatically generated

Hình 2. 1 Tạo mới database

1. TẠO PROJECT SSIS TRONG VISUAL STUDIO

* Chọn File -> New -> Project hộp thoại sẽ hiện ra như hình

Text

Description automatically generated

Hình 2. 2 Tạo mới project SSIS trong Visual Studio

* Sau đó chọn công cụ Integration Service Project, chọn đường dẫn lưu project, đặt tên và nhấn OK.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 2. 3 Màn hình project SSIS trong Visual Studio

* Ta tạo thêm 1 Flat File Connection để tạo kết nối tới file dữ liệu trong Connection Manager. Nhấn chuột phải vào Connection Manager và chọn New Flat File Connection.

Graphical user interface, text, application, email

Description automatically generated

Hình 2. 4 Tạo Flat File Connection

* Ta nhấn Columns để kiểm tra dữ liệu. Sau đó nhấn OK để tạo connection.

Graphical user interface

Description automatically generated

Hình 2. 5 Kiểm tra dữ liệu đầu vào

* Tiếp theo, ta tạo 1 OLE DB Connection để tạo kết nối tới database Movie\_Industry trong Connection Manager.

Graphical user interface, application, email

Description automatically generated

Hình 2. 6 Tạo OLE DB Connection

* Ta điền tên server của SQL Server và chọn database Movie\_Industry rồi nhấn OK để tạo connection.

1. TẠO BẢNG DIM VÀ BẢNG FACT
   1. Tạo các bảng Dim
   * Đầu tiên, ta kéo Data Flow Task để thực hiện quá trình đổ dữ liệu cho các Dim và Execute SQL Task để làm sạch dữ liệu sau mỗi lần chạy lại rồi nối chúng lại với nhau theo trình tự.

A screenshot of a computer

Description automatically generated

Hình 2. 7 Control Flow của project SSIS

* + Sau đó ta nhấn vào Data Flow Dim Task để tiến hành đổ dữ liệu các bảng Dim ở bên ngoài chỉ chứa khoá chính (Dim\_Genre, Dim\_Rating, Dim\_Country, Dim\_Time).

Graphical user interface, diagram

Description automatically generated

Hình 2. 8 Data Flow của các bảng Dim bên ngoài

* + Tại đây, ta kéo 1 Flat File Source vào Data Flow.

Graphical user interface, text, application

Description automatically generated

Hình 2. 9 Màn hình của Flat File Source Editor

* + Flat file connection manager thì ta sẽ chọn connection mới tạo ở trên. Nhấn Columns để kiểm tra các thuộc tính có đầy đủ hay không rồi nhấn OK.

Graphical user interface, table

Description automatically generated

Hình 2. 10 Kiểm tra các thuộc tính của Flat File Source

* + Tiếp theo, ta kéo 1 Multicast vào để chia dữ liệu cho các bảng Dim. Ta kéo 1 Sort vào để tiến hành sort dữ liệu cho bảng dim.

Graphical user interface

Description automatically generated

Hình 2. 11 Sort dữ liệu

* + Trong hộp thoại Sort, ta chọn thuộc tính sort cho bảng Dim. Như hình thì ta đang xử lý cho bảng Dim\_Genre nên sẽ sort theo thuộc tính genre và nhấn Remove rows… để loại bỏ các giá trị trùng và chỉ lấy giá trị độc lập.
  + Cuối cùng, ta kéo OLE DB Destination để tạo 1 điểm đích chứa dữ liệu đổ vào.

Graphical user interface, application

Description automatically generated

Hình 2. 12 Màn hình của OLE DB Destination Editor

* + Ta nhấn New để tạo mới 1 table đại diện cho Dim và chọn các thuộc tính của bảng Dim đó.

Graphical user interface, application

Description automatically generated

Hình 2. 13 Tạo mới table trong database để lưu dữ liệu

* + Tại đây, ta tạo thêm 1 thuộc tính id làm khoá chính và cho nó tăng tự động rồi nhấn OK để tạo table.
  + Ta nhấn Mappings để kiểm tra xem dữ liệu đã được map qua chưa. Kiểm tra xong thì nhấn OK.

Graphical user interface

Description automatically generated

Hình 2. 14 Kiểm tra mapping dữ liệu

* + Vậy là ta đã tạo xong 1 flow để đổ dữ liệu cho 1 bảng Dim. Đối với những bảng Dim còn lại thì ta cứ tiến hành Sort và đổ ra OLE DB Destination như trên.
  + Riêng với bảng Dim \_Time thì trước khi Sort ta kéo Derived Columns để tách ngày tháng năm quý từ thuộc tính “released”.

Graphical user interface, text, application

Description automatically generated

Hình 2. 15 Chia ngày tháng năm từ Derived Column

* + Tại đây, ta kéo các Date/Time Functions mà ta muốn tách từ thuộc tính “released” và đặt tên cho cột được tách như hình. Sau khi tách và Sort xong, ta tạo OLE DB Destination và tạo 1 table Dim\_Time rồi kiểm tra dữ liệu đã được map chưa trong Columns.

Graphical user interface

Description automatically generated

Hình 2. 16 Kiểm tra mapping dữ liệu

* + Sau khi hoàn thành các flow đổ dữ liệu cho bảng Dim thì ta quay lại Execute SQL Task để thêm các câu lệnh làm sạch dữ liệu cho bảng Dim.

Graphical user interface, application

Description automatically generated

Hình 2. 17 Màn hình của Execute SQL Task Editor

* + Rồi ta nhấn Start để tiến hành đổ dữ liệu cho các bảng Dim.

A screenshot of a video game

Description automatically generated

Hình 2. 18 Chạy thành công Data Flow của các bảng Dim bên ngoài

* + Vậy là ta đã đổ dữ liệu cho các bảng Dim bên ngoài. Tiếp theo, ta tiến hành đổ dữ liệu cho bảng Dim\_Film để nối với các bảng Dim ở bên ngoài lại.
  + Ta nhấn vào Data Flow Dim Task 1 để đổ dữ liệu cho bảng Dim\_Film.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 2. 19 Data Flow của bảng Dim Film

* + Ta cũng kéo 1 Flat File Source vào và chọn connection manager rồi kiểm tra các thuộc tính.
  + Sau đó, ta kéo các OLE DB Source vào để chứa dữ liệu của các bảng Dim đã tạo ở trên.
  + Ta sẽ sử dụng Merge Join để nối các bảng Dim lại với nhau theo sơ đồ bông tuyết ở trên. Điều kiện để sử dụng được Merge Join là 2 nguồn dữ liệu phải được Sort theo cùng 1 thuộc tính cùng kiểu dữ liệu.
  + Vì vậy, ta kéo các Sort vào cho các nguồn dữ liệu và tiến hành Sort theo thuộc tính đặc trưng của các bảng Dim.

Graphical user interface, table

Description automatically generated

Hình 2. 20 Merge Join dữ liệu của 2 bảng Dim

* + Sau khi Sort xong, ta nhấn vào Merge Join. Tại đây, ta sẽ merge join theo cùng 1 thuộc tính. Theo hình thì ta đang nối với Dim\_Genre nên ta sẽ join theo thuộc tính “genre” và chọn các thuộc tính. Ta sẽ chọn thuộc tính “genre\_id” đại diện cho thể loại phim và làm khoá ngoại cho bảng Dim\_Film để nối được các Dim lại với nhau thông qua khoá ngoại và khoá chính.
  + Rồi ta tiến hành Sort và Merge Join cho các bảng Dim còn lại. Ta có hình tổng quát như sau.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 2. 21 Data Flow của bảng Dim Film

* + Sau khi Merge Join đủ các bảng Dim thì ta kéo 1 OLE DB Destination vào để đổ dữ liệu cho bảng Dim\_Film.

Graphical user interface, text, application

Description automatically generated

Hình 2. 22 Tạo table trong database để lưu dữ liệu

* + Tại đây, ta chọn các thuộc tính cho bảng Dim\_Film và tạo 1 id tăng tự động làm khoá chính rồi nhấn Mappings để kiểm tra.
  + Sau khi hoàn thành xong flow đổ dữ liệu cho bảng Dim\_Film thì ta thêm câu lệnh làm sạch dữ liệu ở Execute SQL Task rồi nhấn Start để đổ dữ liệu.

Diagram

Description automatically generated

Hình 2. 23 Chạy thành công Data Flow của bảng Dim Film

* + Ta đã đổ dữ liệu cho bảng Dim\_Film xong.
  1. Tạo bảng Fact
  + Ta tiến hành đổ dữ liệu cho bảng Fact\_Table. Nhấn vào Data Flow Fact Task.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 2. 24 Data Flow của bảng Fact

* + Ta tiến hành thực hiện các bước Sort và Merge Join như đã làm ở bảng Dim\_Film theo hình trên.
  + Sau khi đã hoàn thành flow đổ dữ liệu bảng Fact\_Table thì ta cũng thêm câu lệnh làm sạch dữ liệu cho bảng Fact\_Table ở Execute SQL Task rồi nhấn Start.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 2. 25 Chạy thành công Data Flow của bảng Fact

* + Vậy là ta đã hoàn thành xong tiến trình SSIS cho các bảng Dim và bảng Fact. Tiếp theo, ta vào SQL Server và tạo mới 1 diagram trên database Movie\_Industry rồi vẽ mối quan giữa các bảng Dim và bảng Fact để phục vụ cho tiến trình SSAS.

Graphical user interface

Description automatically generated

Hình 2. 26 Sơ đồ bông tuyết sau khi xong quá trình SSIS

* + Ta nối các khoá ngoại với khoá chính lại với nhau để ra được sơ đồ bông tuyết như đã minh hoạ ở trên.

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

1. TẠO PROJECT SSAS TRONG VISUAL STUDIO

* Chọn File -> New -> Project như hình.

Text

Description automatically generated

Hình 3. 1 Tạo mới Project SSAS trong Visual Studio

* Sau đó chọn công cụ Analysis Service, chọn đường dẫn lưu project, đặt tên và nhấn OK.

A screenshot of a computer

Description automatically generated

Hình 3. 2 Màn hình của project SSAS trong Visual Studio

1. XÁC ĐỊNH DATA SOURCE

* Góc bên phải phần Solution Explorer nhấn chuột phải vào Data Sources và chọn New Data Source…

Graphical user interface, application

Description automatically generated

Hình 3. 3 Tạo mới Data Source

* Một cửa sổ hiện ra rồi nhấn Next.

Graphical user interface, text, application, Word

Description automatically generated

Hình 3. 4 Màn hình của Data Source Wizard

* Chọn kết nối đã tạo từ tiến trình SSIS rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 3. 5 Chọn Data Connection

* Điền tài khoản với username là username trong phần Login của database Movie\_Industry ở SQL Server và mật khẩu là mật khẩu của máy.

Graphical user interface, text, application, email

Description automatically generated

Hình 3. 6 Nhập tài khoản Windows

* Nhấn Next rồi Finish để hoàn thành việc tạo Data Source.

1. XÁC ĐỊNH DATA SOURCE VIEW

* Góc bên phải phần Solution Explorer nhấn chuột phải vào Data Source Views và chọn New Data Source View…

A screenshot of a computer

Description automatically generated with medium confidence

Hình 3. 7 Tạo mới Data Source View

* Một cửa sổ hiện ra rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 3. 8 Màn hình của Data Source View Wizard

* Chọn Data Source Movie Industry mới tạo ở trên rồi nhấn Next.

Graphical user interface, text

Description automatically generated

Hình 3. 9 Chọn Data Source

* Chọn bảng Fact và Dim cho quá trình phân tích rồi nhấn Next.

Graphical user interface

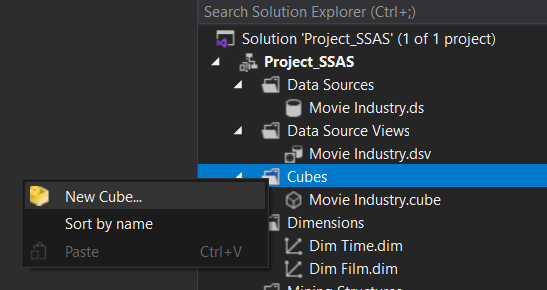
Description automatically generated

Hình 3. 10 Chọn các Table

* Nhấn Finish để kết thúc quá trình tạo Data Source View

1. XÂY DỰNG CÁC CUBE VÀ CÁC MEASURE

* Góc bên phải phần Solution Explorer nhấn chuột phải vào Cubes rồi chọn New Cube…



Hình 3. 11 Tạo mới Cube

* Một cửa sổ hiện ra rồi nhấn Next.

Graphical user interface, application

Description automatically generated

Hình 3. 12 Màn hình của Cube Wizard

* Trong cửa sổ Cube Wizard chọn Use existing tables rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 3. 13 Chọn phương thức tạo Cube

* Chọn Fact\_Table để phân chia các measure group.

Graphical user interface, text, application

Description automatically generated

Hình 3. 14 Chọn Measure Group Tables

* Chọn những thuộc tính định lượng đề xuất rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 3. 15 Chọn các Measures

* Chọn danh sách Dimension rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 3. 16 Chọn các Dimensions

* Nhấn Finish để hoàn thành quá trình tạo Cube.

Graphical user interface, text, application, email

Description automatically generated

Hình 3. 17 Hoàn thành việc tạo Cube

A screenshot of a computer

Description automatically generated with medium confidence

Hình 3. 18 Màn hình của cube mới tạo

1. PHÂN CẤP TRONG BẢNG DIM
   1. Bảng Dim\_Time
   * Kéo những thuộc tính cần phân cấp qua cửa sổ hierachies

Text

Description automatically generated

Hình 3. 19 Phân cấp dữ liệu bảng Dim Time

* + Sau đó, ta chỉnh khoá dòng và tên dòng của thuộc tính Day. Vì thuộc tính Day là thuộc tính cấp nhỏ nhất nên sẽ lấy khoá dòng gồm chính nó và những thuộc tính cấp cao hơn.

Graphical user interface

Description automatically generated

Hình 3. 20 Chỉnh khoá dòng của thuộc tính Day

* + Tạo tên dòng của thuộc tính Day

Graphical user interface

Description automatically generated

Hình 3. 21 Tạo tên dòng cho thuộc tính Day

* + Tiếp theo, ta chỉnh khoá dòng và tên dòng của thuộc tính Month. Vì thuộc tính Month là thuộc tính cấp nhỏ hơn Quarter và Year nên sẽ lấy khoá dòng gồm chính nó và Quarter, Year.

Graphical user interface, application

Description automatically generated

Hình 3. 22 Chỉnh khoá dòng của thuộc tính Month

* + Tạo tên dòng của thuộc tính Month

Graphical user interface, application, Word

Description automatically generated

Hình 3. 23 Tạo tên dòng cho thuộc tính Month

* + Ta chỉnh khoá dòng và tên dòng của thuộc tính Quarter. Vì thuộc tính Quarter là thuộc tính cấp nhỏ hơn Year nên sẽ lấy khoá dòng gồm chính nó và Year.

Graphical user interface, application

Description automatically generated

Hình 3. 24 Chỉnh khoá dòng của thuộc tính Quarter

* + Tạo tên dòng của thuộc tính Quarter

Graphical user interface, application

Description automatically generated

Hình 3. 25 Tạo tên dòng cho thuộc tính Quarter

* + Cuối cùng, ta qua tab Attribute Relationships để liên kết các mối quan hệ phân cấp và tiến hành phân cấp như hình:

Graphical user interface

Description automatically generated

Hình 3. 26 Mối quan hệ của các thuộc tính trong bảng Dim Time

1. CHẠY PROJECT SSAS

* Nhấn chuột phải vào Cube đã tạo ở trên rồi nhấn Process như hình:

A screenshot of a computer

Description automatically generated with medium confidence

Hình 3. 27 Process Cube

* Sau đó hộp thoại xuất hiện và nhấn Run.

Graphical user interface, text, application, email

Description automatically generated

Hình 3. 28 Process Cube

Graphical user interface, text, application

Description automatically generated

Hình 3. 29 Process Cube

* Và thế là ta đã chạy thành công SSAS.

1. QUÁ TRÌNH PHÂN TÍCH DỮ LIỆU BẰNG NGÔN NGỮ MDX

a. Cho biết tổng doanh thu đạt được của các bộ phim ở nước Mĩ (United State) theo từng năm.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

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

NON EMPTY { ([Dim Time].[Year].CHILDREN ) } ON ROWS

FROM [Movie Industry]

WHERE [Dim Film].[Country].&[United States];

- Kết quả câu truy vấn:

Table

Description automatically generated with medium confidence

Hình 3. 30 Kết quả MDX của câu a

- Kết quả trên PowerBI:

Table

Description automatically generated

Hình 3. 31 Kết quả PowerBI của câu a

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

b. Cho biết thông tin bộ phim đạt được doanh thu cao nhất trong quý 2 năm 2016 của nước Anh (United Kingdom).

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

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

{TOPCOUNT( [Dim Film].[Name].CHILDREN \* [Dim Film].[Rating].CHILDREN \* [Dim Film].[Genre].CHILDREN, 1, [Measures].[Gross] )} ON ROWS

FROM [Movie Industry]

WHERE ({[Dim Film].[Country].&[United Kingdom]}, {[Dim Time].[Quarter].&[2]&[2016]});

- Kết quả câu truy vấn:

Table

Description automatically generated

Hình 3. 32 Kết quả MDX của câu b

- Kết quả trên PowerBI:

Table

Description automatically generated with medium confidence

Hình 3. 33 Kết quả PowerBI của câu b

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

c. Đếm số lượng bộ phim được sản xuất theo từng quốc gia.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT NON EMPTY { [Measures].[Fact Table Count] } ON COLUMNS,

NON EMPTY { [Dim Film].[Country].CHILDREN } ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Table

Description automatically generated

Hình 3. 34 Kết quả MDX của câu c

- Kết quả trên PowerBI:

Table

Description automatically generated

Hình 3. 35 Kết quả PowerBI của câu c

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

d. Liệt kê bộ phim có kinh phí đầu tư thấp nhất và bộ phím có kinh phí đầu tư cao nhất và in ra số điểm của 2 bộ phim đó.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT { [Measures].[Budget], [Measures].[Score] } ON COLUMNS,

{UNION(TOPCOUNT( NONEMPTY([Dim Film].[Name].CHILDREN), 1, [Measures].[Score]),

BOTTOMCOUNT( NONEMPTY( [Dim Film].[Name].CHILDREN), 1, [Measures].[Score] ))} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Graphical user interface, application

Description automatically generated

Hình 3. 36 Kết quả MDX của câu d

- Kết quả trên PowerBI:

Graphical user interface, text

Description automatically generated with medium confidence

Hình 3. 37 Kết quả PowerBI của câu d

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

e. Thống kê kinh phí đầu tư, doanh thu và thông tin các bộ phim thành công (kinh phí < doanh thu).

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT NON EMPTY [Dim Film].[Genre].CHILDREN \* [Dim Film].[Name].CHILDREN

HAVING [Measures].[Gross] > [Measures].[Budget] ON ROWS,

NON EMPTY { [Measures].[Budget], [Measures].[Gross] } ON COLUMNS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Table

Description automatically generated

Hình 3. 38 Kết quả MDX của câu e

- Kết quả trên PowerBI:

Table

Description automatically generated

Hình 3. 39 Kết quả PowerBI của câu e

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

f. Đưa ra top 3 thể loại có số lượng phim nhiều nhất trong năm 2016.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT NON EMPTY { [Measures].[Fact Table Count] } ON COLUMNS,

NON EMPTY {TOPCOUNT( [Dim Film].[Genre].CHILDREN, 3, [Measures].[Fact Table Count] )} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Table

Description automatically generated

Hình 3. 40 Kết quả MDX của câu f

- Kết quả trên PowerBI:

Table

Description automatically generated

Hình 3. 41 Kết quả PowerBI của câu f

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

g. Tìm ra đạo diễn có bộ phim đạt doanh thu cao nhất.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

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

{TOPCOUNT( [Dim Film].[Director].CHILDREN \* [Dim Film].[Name].CHILDREN, 1, [Measures].[Gross] )} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Table

Description automatically generated

Hình 3. 42 Kết quả MDX của câu g

- Kết quả trên PowerBI:

Graphical user interface, table

Description automatically generated

Hình 3. 43 Kết quả PowerBI của câu g

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

h. Đưa ra top 3 bộ phim được yêu thích nhất 3 năm gần đây (dựa theo số điểm).

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT { [Measures].[Score] } ON COLUMNS,

{TOPCOUNT( [Dim Film].[Name].CHILDREN \* [Dim Film].[Rating].CHILDREN

\* [Dim Film].[Genre].CHILDREN \* [Dim Time].[Year].CHILDREN, 3, [Measures].[Score] )} ON ROWS

FROM (

SELECT { [Measures].[Score] } ON COLUMNS,

{TAIL( NONEMPTY([Dim Time].[Year].CHILDREN), 3)} ON ROWS

FROM [Movie Industry]);

- Kết quả câu truy vấn:

Table

Description automatically generated

Hình 3. 44 Kết quả MDX của câu h

- Kết quả trên PowerBI:

Table

Description automatically generated

Hình 3. 45 Kết quả PowerBI của câu h

=> Nhận xét: Kết quả câu truy vấn MDX có khác với PowerBI do Top N của PowerBI có sử dụng With Ties để lấy thêm những dữ liệu có giá trị giống với dữ liệu nằm trong top, tránh tình trạng thiếu sót những dữ liệu đó.

i. Thống kê các bộ phim có thời lượng chiếu trên 180 phút và có số điểm đạt được trên 8.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT NON EMPTY FILTER( [Dim Film].[Name].CHILDREN,

[Measures].[Runtime] > 180 and [Measures].[Score] > 8.0) ON ROWS,

NON EMPTY { [Measures].[Runtime], [Measures].[Score] } ON COLUMNS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Graphical user interface, text, application

Description automatically generated

Hình 3. 46 Kết quả MDX của câu i

- Kết quả trên PowerBI:

Table

Description automatically generated with medium confidence

Hình 3. 47 Kết quả PowerBI của câu i

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

j. Theo từng tháng, quý, năm liệt kê số lượng bộ phim được khởi chiếu trong 3 năm gần đây nhất.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT NON EMPTY { [Measures].[Fact Table Count] } ON COLUMNS,

{DRILLDOWNLEVEL(

DRILLDOWNLEVEL(

DRILLDOWNLEVEL(

[Dim Time].[PhanCap])

)

)

} ON ROWS

FROM (

SELECT { [Measures].[Fact Table Count] } ON COLUMNS,

{TAIL( NONEMPTY([Dim Time].[Year].CHILDREN), 3)} ON ROWS

FROM [Movie Industry]);

- Kết quả câu truy vấn:

Table

Description automatically generated with medium confidence

Hình 3. 48 Kết quả MDX của câu j

- Kết quả trên PowerBI:

Table

Description automatically generated

Hình 3. 49 Kết quả PowerBI của câu j

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

k. Liệt kê top 5 những bộ phim có kinh phí đầu tư cao nhất hoặc top 3 có số điểm cao nhất.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

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

{UNION(TOPCOUNT( [Dim Film].[Name].CHILDREN , 5, [Measures].[Budget] ),

TOPCOUNT( [Dim Film].[Name].CHILDREN , 3, [Measures].[Score] ))} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Graphical user interface, text

Description automatically generated

Hình 3. 50 Kết quả MDX của câu k

- Kết quả trên PowerBI:

Table

Description automatically generated with low confidence

Hình 3. 51 Kết quả PowerBI của câu k

=> Nhận xét: Kết quả câu truy vấn MDX có khác với PowerBI do Top N của PowerBI có sử dụng With Ties để lấy thêm những dữ liệu có giá trị giống với dữ liệu nằm trong top, tránh tình trạng thiếu sót những dữ liệu đó.

l. Đưa ra top 3 diễn viên có số lượng bộ phim được vào vai nhiều nhất.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT NON EMPTY { [Measures].[Fact Table Count] } ON COLUMNS,

{TOPCOUNT( [Dim Film].[Star].CHILDREN, 3, [Measures].[Fact Table Count] )} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Table

Description automatically generated

Hình 3. 52 Kết quả MDX của câu l

- Kết quả trên PowerBI:

Table

Description automatically generated

Hình 3. 53 Kết quả PowerBI của câu l

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

m. Liệt kê những bộ phim có doanh thu từ $1000000 đến $3000000.

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

SELECT { [Measures].[Gross] } ON COLUMNS,

{INTERSECT(

{FILTER([Dim Film].[Name].CHILDREN , [Measures].[Gross] >= 1000000)}

,{FILTER([Dim Film].[Name].CHILDREN , [Measures].[Gross] <= 3000000)}

)} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Table

Description automatically generated with low confidence

Hình 3. 54 Kết quả MDX của câu m

- Kết quả trên PowerBI:

Graphical user interface

Description automatically generated with low confidence

Hình 3. 55 Kết quả PowerBI của câu m

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

n. Đưa ra 3 thể loại được quan tâm nhất (dựa theo số lượng vote).

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

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

{TOPCOUNT( [Dim Film].[Genre].CHILDREN, 3, [Measures].[Votes] )} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Graphical user interface, table

Description automatically generated

Hình 3. 56 Kết quả MDX của câu n

- Kết quả trên PowerBI:

Text

Description automatically generated

Hình 3. 57 Kết quả PowerBI của câu n

=> Nhận xét: Kết quả câu truy vấn MDX giống với PowerBI.

o. Đưa ra top 3 bộ phim được yêu thích nhất mọi thời đại

- Câu lệnh truy vấn bằng ngôn ngữ MDX:

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

{TOPCOUNT( [Dim Film].[Name].CHILDREN \* [Dim Film].[Rating].CHILDREN

\* [Dim Film].[Genre].CHILDREN \* [Dim Time].[Year].CHILDREN, 3, [Measures].[Score] )} ON ROWS

FROM [Movie Industry];

- Kết quả câu truy vấn:

Graphical user interface, application

Description automatically generated

Hình 3. 58 Kết quả MDX của câu o

- Kết quả trên PowerBI:

Graphical user interface, table

Description automatically generated

Hình 3. 59 Kết quả PowerBI của câu o

=> Nhận xét: Kết quả câu truy vấn MDX có khác với PowerBI do Top N của PowerBI có sử dụng With Ties để lấy thêm những dữ liệu có giá trị giống với dữ liệu nằm trong top, tránh tình trạng thiếu sót những dữ liệu đó.

# CHƯƠNG 4. QUÁ TRÌNH LẬP BÁO BIỂU (SSRS)

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

* Chọn File -> New -> Project như hình.

Graphical user interface, text

Description automatically generated

Hình 4. 1 Tạo mới project SSRS trong Visual Studio

* Sau đó chọn công cụ Report Server Project, chọn đường dẫn lưu project, đặt tên và nhấn OK.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 4. 2 Màn hình của project SSRS trong Visual Studio

* Tại cửa sổ Solution Explorer phía bên phải, nhấn chuột phải vào phần Reports -> Add -> New Item.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 4. 3 Tạo mới Item

* Chọn Report -> đặt tên -> nhấn Add.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 4. 4 Chọn Report

A screenshot of a computer

Description automatically generated

Hình 4. 5 Mà hình Report

* Tab Design: Để thiết kế Report
* Tab Preview: Để show kết quả Report
  1. Tạo kết nối với kho dữ liệu
* Tại thanh Report Data -> Chọn Data Source -> Chọn Add Data Source.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 4. 6 Tạo mới Data Source

Graphical user interface, application

Description automatically generated

Hình 4. 7 Màn hình của Data Source Properties

* Trường Type chọn Microsoft SQL Server Analysis Services -> Chọn Build và cấu hình như minh hoạ (tuỳ tên) -> nhấn Test Connection. Nếu thành công nhấn OK.

Graphical user interface, text, application, email

Description automatically generated

Hình 4. 8 Điền Server name và chọn database

Graphical user interface, text, application, chat or text message

Description automatically generated

Hình 4. 9 Kiểm tra kết nối

Graphical user interface, text, application, chat or text message

Description automatically generated

Hình 4. 10 Data Source mới tạo

* 1. Thực hiện lập báo biểu trên Report Builder

a. Báo biểu 1 (Report Grouping): Đưa doanh số, lợi nhuận, tổng số lượng mua của từng loại sản phẩm, và các tiểu mục thể loại.

* Thêm datasets cần thiết cho báo biểu 1 đặt tên là DataQuery1 bằng cách nhấn chuột phải vào Datasets -> Add Dataset.

Graphical user interface, text, application, chat or text message

Description automatically generated

Hình 4. 11 Tạo mới Dataset

Graphical user interface, application

Description automatically generated

Hình 4. 12 Màn hình của Dataset Properties

* Trong Datasets Properties -> Trường Query. Tại trường Name đặt tên là DataQuery1 và chọn chế độ Use a dataset embedded in my report. Trường Data Source thì chọn Data Source đã tạo ở trên.

Graphical user interface, application

Description automatically generated

Hình 4. 13 Chọn Data Source đã tạo ở trên

* Chọn Query Designer và hộp thoại sẽ hiện ra.

Graphical user interface, text, application

Description automatically generated

Hình 4. 14 Màn hình của Query Designer

* Chọn Design Mode và nhập câu MDX để lập báo biểu rồi nhấn OK.

Graphical user interface, application

Description automatically generated

Hình 4. 15 Nhập câu MDX để làm report

* Qua trường Toolbox kéo table qua tab Design rồi kéo những thuộc tính vào table rồi định dạng Report.

Graphical user interface, text, application

Description automatically generated

Hình 4. 16 Kéo table vào tab Design

* Tuỳ chọn Grouping bằng các bước sau: Nhấn chuột phải vào Detail -> Add Group -> Parent Group.

Graphical user interface, application

Description automatically generated

Hình 4. 17 Thêm Grouping cho report

* Chọn Group by [Genre].

Graphical user interface, application

Description automatically generated

Hình 4. 18 C họn thuộc tính Grouping

* In báo cáo ra file PDF bằng cách qua tab Preview -> Save file with PDF.

Table

Description automatically generated

Hình 4. 19 Màn hình report

* Đăng Report lên Server. Cài đặt và tìm kiếm Report Server Configuration Manager trong công cụ tìm kiếm của Windows. Chọn Server Name mặc định của SQL server -> Nhấn Connect.

Graphical user interface

Description automatically generated

Hình 4. 20 Kết nối tới Report Server

* Tại phần Web Service URL, để những thông số mặc định rồi nhấn Apply.

Graphical user interface, application

Description automatically generated

Hình 4. 21 Màn hình của Web Service URL

* Vào phần Database chọn Change Database -> Create new report server database. Nhấn Next đến khi hoàn thành -> nhấn Finish.

Graphical user interface, text, application

Description automatically generated

Hình 4. 22 Màn hình của Database

* Connect Report Builder với Report Server. Nhấn chuột phải Project\_SSRS rồi chọn Properties.

Graphical user interface, text, application

Description automatically generated

Hình 4. 23 Điền ServerURL trong Project SSRS Properties

* Tại trường TargetServerURL, nhập URL của Web Service URL trong Report Server Configuration Manager rồi nhấn OK. Sau đó, nhấn chuột phải Project SSRS và chọn deploy.
* Sau khi deploy thành công thì tiến hành truy cập đường link của Web Portal URL trong Report Server Configuration Manager và chọn báo biểu đã tạo ở trên và cho ra kết quả như hình.

Graphical user interface, application

Description automatically generated

Hình 4. 24 Deploy Report lên SQL Server Reporting Services

# CHƯƠNG 5. QUÁ TRÌNH DATA MINING

1. QUÁ TRÌNH THỰC HIỆN DATA MINING VỚI THUẬT TOÁN LUẬT KẾT HỢP (MICROSOFT ASSOCIATION RULES)
   1. Thiết lập cấu trúc data mining
   * Góc bên phải phần Solution Explorer nhấn chuột phải Mining Structures rồi chọn New Mining Structure…

A screenshot of a computer

Description automatically generated with medium confidence

Hình 5. 1 Tạo mới Mining Structure

* + Một hộp thoại xuất hiện rồi nhấn Next.

Graphical user interface, text, application, email

Description automatically generated

Hình 5. 2 Màn hình của Data Mining Wizard

* + Chọn From existing cube rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 5. 3 Chọn thuật toán Mining

* + Chọn thuật toán muốn chạy rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 5. 4 Chọn Source Cube Dimension

* + Chọn Dim rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 5. 5 Chọn Case Key

* + Chọn Case Key rồi nhấn Next.

Graphical user interface, text, application

Description automatically generated

Hình 5. 6 Chọn Case Level Columns

* + Chọn thuộc tính để chạy mining rồi nhấn Next

Graphical user interface, text, application

Description automatically generated

Hình 5. 7 Phân loại Mining Model Column Usage

* + Chọn thuộc tính Input và Predict rồi nhấn Next.

Graphical user interface, text, application, email

Description automatically generated

Hình 5. 8 Điều chỉnh tỉ lệ testing

* + Tại đây chọn tỉ lệ train:test là 70:30 rồi nhấn Next và nhấn Finish để hoàn thành quá trình thiết lập cấu trúc data mining.
  1. Deploy project data mining
  + Nhấn chuột phải vào Mining Structure mới tạo rồi chọn Process.

A screenshot of a computer

Description automatically generated with medium confidence

Hình 5. 9 Process Mining Strucure

* + Một hộp thoại hiện lên rồi nhấn Run.

Graphical user interface, text, application, email

Description automatically generated

Hình 5. 10 Process Mining Structure

Text

Description automatically generated

Hình 5. 11 Process Mining Structure

* + Vậy là ta đã chạy process xong Mining Structure.
  1. Phân tích mining bằng luật kết hợp và đưa ra tập luật
  + Nhấn chuột phải vào Mining Structure đã tạo rồi chọn Browse để xem kết quả.

Graphical user interface, text

Description automatically generated

Hình 5. 12 Các luật từ Mining Structure

* + Rút ra được tập luật:
    - “Nếu thể loại phim là Crime thì nhãn dán là nhãn R (87.4%)
    - “Nếu thể loại phim là Mystery thì nhãn dán là nhãn R (83.3%)
    - “Nếu thể loại phim là Fantasy thì nhãn dán là nhãn R (77.8%)
    - “Nếu thể loại phim là Horror thì nhãn dán là nhãn R (76.8%)
    - “Nếu thể loại phim là Animation thì nhãn dán là nhãn PG (64.5%)
    - “Nếu thể loại phim là Thriller thì nhãn dán là nhãn PG-13 (60%)
  1. So sánh với các thuật toán khác
  + Nhấn chuột phải vào ô trống chọn New Mining Model ở phần Mining Models

A screenshot of a computer

Description automatically generated

* + Chọn thuật toán Microsoft Clustering

Graphical user interface, application

Description automatically generated

* + Tiếp tục tạo thêm thuật toán Microsoft Neural Network

Graphical user interface, application

Description automatically generated

A screenshot of a computer

Description automatically generated

* + Tiến hành Process và Browse

Graphical user interface, text, application

Description automatically generated

* + So sánh và đánh giá giữa các thuật toán bằng đồ thị Lift
    - Trong phần Mining Accuracy Chart chọn Lift Chart để xem đồ thị Lift

Graphical user interface, chart

Description automatically generated

* + - Thuật toán Association Rules có phần trăm chính xác 30.75% với số điểm 0.59
    - Thuật toán Clustering có phần trăm chính xác 25.89% với số điểm 0.51
    - Thuật toán Neural Network có phần trăm chính xác 29.64% với số điểm 0.55
* Kết luận: Ta ưu tiên chọn thuật toán Association Rules.

# CHƯƠNG 6. TÀI LIỆU THAM KHẢO

[1]: Slide bài giảng, ThS. Nguyễn Thị Kim Phụng . [Access Time: 15/05/2022]

[2]: “Movie Industry”, Kaggle, <https://www.kaggle.com/datasets/danielgrijalvas/movies> . [Access Time: 15/05/2022]

[3]: “MDX Query Fundamentals (Analysis Services)”, Microsoft, <https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/mdx/mdx-query-fundamentals-analysis-services?view=asallproducts-allversions> . [Access Time: 15/05/2022]