**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**



**KHO DỮ LIỆU VÀ OLAP**

**GVHD:** ThS. Nguyễn Thị Kim Phụng

**Lớp:** IS217.N22.HTCL - VN

**Sinh viên:**

Nguyễn Cẩm Tú 20520837

Nguyễn Thị Thu Thủy 20520797

**THÀNH PHỐ HỒ CHÍ MINH, THÁNG 7 NĂM 2023**

# NHẬN XÉT CỦA GIẢNG VIÊN

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

# MỤC LỤC

[NHẬN XÉT CỦA GIẢNG VIÊN 2](#_Toc170080331)

[MỤC LỤC 3](#_Toc170080332)

[I. MÔ TẢ BỘ DỮ LIỆU 4](#_Toc170080333)

[1.1. Mô tả bộ dữ liệu 4](#_Toc170080334)

[1.1. Nguồn dữ liệu 4](#_Toc170080335)

[1.2. Bảng mô tả 4](#_Toc170080336)

[1.2. Mô tả thuộc tính 5](#_Toc170080337)

[1.3. Sơ đồ hình sao 7](#_Toc170080338)

[II. SSIS 8](#_Toc170080339)

[2.1. Tạo DataWarehouse 8](#_Toc170080340)

[2.2. Tạo Project và thực hiện kết nối CSDL 8](#_Toc170080341)

[2.3. Thực thi project 21](#_Toc170080342)

[III. SSAS VÀ REPORT 22](#_Toc170080343)

[3.1. Tạo các Data Sourses, Cubes, các Dimension 22](#_Toc170080344)

[3.1.1 Data Sources 22](#_Toc170080345)

[3.1.2 Data Sources View 24](#_Toc170080346)

[3.1.3 Cubes 28](#_Toc170080347)

[3.2. Truy vấn 31](#_Toc170080348)

[3.2.1. Truy vấn Roll-Up: 31](#_Toc170080349)

[3.2.2. Truy vấn Drill-Down: 33](#_Toc170080350)

[3.2.3. Truy vấn Slice and Dice: 33](#_Toc170080351)

[3.2.4. Truy vấn Pivot: 35](#_Toc170080352)

[3.2.5. Slice and Dice kết hợp Drill-Down: 36](#_Toc170080353)

[3.2.6. Thao tác lọc cột/dòng (manual): 37](#_Toc170080354)

[IV. SSRS 38](#_Toc170080355)

[4.1. Report dạng matrix 38](#_Toc170080356)

[4.2. Report dạng chart 44](#_Toc170080357)

[4.3. Power BI 45](#_Toc170080358)

[V. MDX 47](#_Toc170080359)

[BẢNG PHÂN CÔNG CÔNG VIỆC TRONG NHÓM 52](#_Toc170080360)

# MÔ TẢ BỘ DỮ LIỆU

## Mô tả bộ dữ liệu

### Nguồn dữ liệu

* Bộ dữ liệu viết về trang sách Goodreads với khoảng 20.000 cuốn sách có sẵn trên kho lưu trữ của trang mạng và gồm nhiều bộ dữ liệu được thu thập.
* Nguồn bộ dữ liệu được lấy từ: [Goodreads Book Datasets With User Rating 2M | Kaggle](https://www.kaggle.com/datasets/bahramjannesarr/goodreads-book-datasets-10m?select=book1000k-1100k.csv&fbclid=IwAR3oXWkASdINglN7Hhi_rJAg_jAlDQpt5q66v-NbMR1DAb5rMw6GbxFspgM) và dữ liệu có sẵn ở định dạng CSV gồm 20000 dòng dữ liệu và 19 cột.
* Đường dẫn drive:

<https://drive.google.com/drive/folders/1d3kLM24WFXLdC2k03z7Gvdt-D8Omz9WZ?usp=sharing>

### Bảng mô tả

*Bảng mô tả các trường dữ liệu:*

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên trường dữ liệu** | **Mô tả** | **Kiểu dữ liệu** |
| **1** | ID | Mã sách | Int |
| **2** | Name | Tựa đề sách | String |
| **3** | Authors | Tên tác giả | String |
| **4** | ISBN | Mã số tiêu chuẩn quốc cho sách (International Standard Book Number) | String |
| **5** | Rating | Điểm đánh giá trung bình | Float |
| **6** | PublishYear | Năm xuất bản | Int |
| **7** | PublishMonth | Tháng xuất bản | Int |
| **8** | PublishDay | Ngày xuất bản | Int |
| **9** | Publisher | Tên nhà xuất bản | String |
| **10** | RatingDist5 | Số lượt đánh giá 5 sao  (VD: 5:1 tức là có 1 lượt đánh 5 sao) | String |
| **11** | RatingDist4 | Số lượt đánh giá 4 sao | String |
| **12** | RatingDist3 | Số lượt đánh giá 3 sao | String |
| **13** | RatingDist2 | Số lượt đánh giá 2 sao | String |
| **14** | RatingDist1 | Số lượt đánh giá 1 sao | String |
| **15** | RatingDistTotal | Tổng số lượt đánh giá | String |
| **16** | CountsOfReview | Tổng số lượt đánh giá (dạng văn bản) | Int |
| **17** | Language | Ngôn ngữ | String |
| **18** | pagesNumber | Tổng số trang sách | Int |
| **19** | Description | Miêu tả của quyển sách | String |

* Trên bộ dữ liệu là 20 cột nhưng có 2 cột trùng tên nhau đó là cột CountsOfReview và cột Count of text reviews nên chúng em chỉ chọn 1 cột CountsOfReview để sử dụng

## Mô tả thuộc tính

*Bảng Dim\_Book*

|  |  |  |
| --- | --- | --- |
| **Tên trường dữ liệu** | **Mô tả** | **Kiểu dữ liệu** |
| ID | Mã sách | Int |
| Name | Tựa đề sách | String |

*Bảng Dim\_Authors*

|  |  |  |
| --- | --- | --- |
| **Tên trường dữ liệu** | **Mô tả** | **Kiểu dữ liệu** |
| ID\_Authors | Mã tác giả | Int |
| Authors | Tên tác giả | String |

*Bảng Dim\_PublishDate*

|  |  |  |
| --- | --- | --- |
| **Tên trường dữ liệu** | **Mô tả** | **Kiểu dữ liệu** |
| PublishDate | Ngày xuất bản | Date |
| PublishYear | Năm xuất bản | Int |
| PublishMonth | Tháng xuất bản | Int |
| PublishDay | Ngày xuất bản | Int |

*Bảng Dim\_Publisher*

|  |  |  |
| --- | --- | --- |
| **Tên trường dữ liệu** | **Mô tả** | **Kiểu dữ liệu** |
| ID\_Publisher | Mã nhà xuất bản | int |
| Publisher | Tên nhà xuất bản | String |

*Bảng Dim\_Language*

|  |  |  |
| --- | --- | --- |
| **Tên trường dữ liệu** | **Mô tả** | **Kiểu dữ liệu** |
| ID\_Language | Mã ngôn ngữ | Int |
| Language | Ngôn ngữ | String |

*Bảng Fact*

|  |  |  |
| --- | --- | --- |
| **Tên trường dữ liệu** | **Mô tả** | **Kiểu dữ liệu** |
| ID | Mã sách | Int |
| ID\_Authors | Mã tác giả | Int |
| PublishDate | Ngày xuất bản | Date |
| ID\_Pulisher | Mã nhà xuất bản | Int |
| ID\_Language | Mã ngôn ngữ | Int |
| ISBN | Mã số tiêu chuẩn quốc cho sách (International Standard Book Number) | String |
| Rating | Điểm đánh giá trung bình | Float |
| RatingDist5 | Số lượt đánh giá 5 sao  (VD: 5:1 tức là có 1 lượt đánh 5 sao) | String |
| RatingDist4 | Số lượt đánh giá 4 sao | String |
| RatingDist3 | Số lượt đánh giá 3 sao | String |
| RatingDist2 | Số lượt đánh giá 2 sao | String |
| RatingDist1 | Số lượt đánh giá 1 sao | String |
| RatingDistTotal | Tổng số lượt đánh giá | String |
| pagesNumber | Tổng số trang sách | Int |
| Description | Miêu tả của quyển sách | String |

## Sơ đồ hình sao

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

# SSIS

## Tạo DataWarehouse

Ảnh có chứa văn bản

Mô tả được tạo tự động

Click chuột phải vào Databases 🡪 New Database

## Tạo Project và thực hiện kết nối CSDL

***Bước 1:*** Tạo mới Project ‘Integration Services Project’

Sau khi tạo được project, sẽ hiển thị màn hình như sau:

Ảnh có chứa văn bản, phần mềm, Phần mềm đa phương tiện, Biểu tượng máy tính

Mô tả được tạo tự động

***Bước 2:*** Thêm các toolbox trong thanh SSIS Toolbox với các tool: Excel Source, Multicast, Sort và OLE DB Destination.

Ảnh có chứa văn bản, phần mềm, Biểu tượng máy tính, Phần mềm đa phương tiện

Mô tả được tạo tự động

***Bước 3:*** Chọn dữ liệu đầu vào cho Excel Source:

Ảnh có chứa văn bản, phần mềm, Biểu tượng máy tính, Phần mềm đa phương tiện

Mô tả được tạo tự động

***Bước 4:*** Chỉnh sửa các Sort Transformation

Sort BOOK Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Sort AUTHORẢnh có chứa văn bản, màn hình, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

Sort PUBLISHDATE

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Sort PUBLISHER

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Sort LANGUAGE

Ảnh có chứa văn bản, màn hình, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

**CÁC BẢNG DIM:**

Để thực hiện chỉnh sửa được OLE DB Destination chúng ta cần phải kết nối với CSDL SQL Server -> Chọn New.

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Tại Provide: Chọn Native OLE DB\....OLE DB Provider for SQL Server

Tại Server name: Chọn CSDL mà chúng ta sử dụng để thiết lập kết nối

Tại Select or enter a database name: Chọn database đã tạo trước đó.

Cuối cùng, kiểm tra kết nối -> Chọn Test connection -> Test connection successful là kết nối thành công.

Ảnh có chứa văn bản, phần mềm, ảnh chụp màn hình, Biểu tượng máy tính

Mô tả được tạo tự động

Sau khi kết nối xong, bắt đầu tạo các bảng DIM

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Tạo lần lượt các bảng DIM. Sau đó kiểm tra Mappings, tức là kiểm tra dữ liệu bảng vừa tạo có khớp với dữ liệu mình muốn đưa vào từ file Excel hay chưa.

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, màn hình, ảnh chụp màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Sau khi chỉnh sửa xong Data Flow, thì chúng ta tiến hành chỉnh sửa Control Flow, kết nối cơ sở dữ liệu SQL Server

Ảnh có chứa văn bản, phần mềm, Biểu tượng máy tính, Phần mềm đa phương tiện

Mô tả được tạo tự động

Chọn Database để kết nối đến

Tại SQLStatement, thêm các câu lệnh delete dữ liệu các bảng DIM.

Vì khi chạy project, các bảng DIM đã được tạo tự động trên csdl SQL Server. Do đó, khi muốn tiến hành chỉnh sửa hay cập nhật dữ liệu mới và phải chạy lại project một lần nữa, thì các bảng DIM đã tạo trước đó sẽ bị trùng nhau. Vì thế cần có câu lệnh “delete from table” để tránh trường hợp bị trùng khi rebuild project mà không cần phải xóa các bảng DIM trên SQL bằng tay.

Ảnh có chứa văn bản, phần mềm, Biểu tượng máy tính, Phần mềm đa phương tiện

Mô tả được tạo tự động

Lưu ý: Chỉnh font chữ về UTF8 bằng cách DefaultCodePage thành 65001

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

## Thực thi project

Nhấn start để bắt đầu thực thi project

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

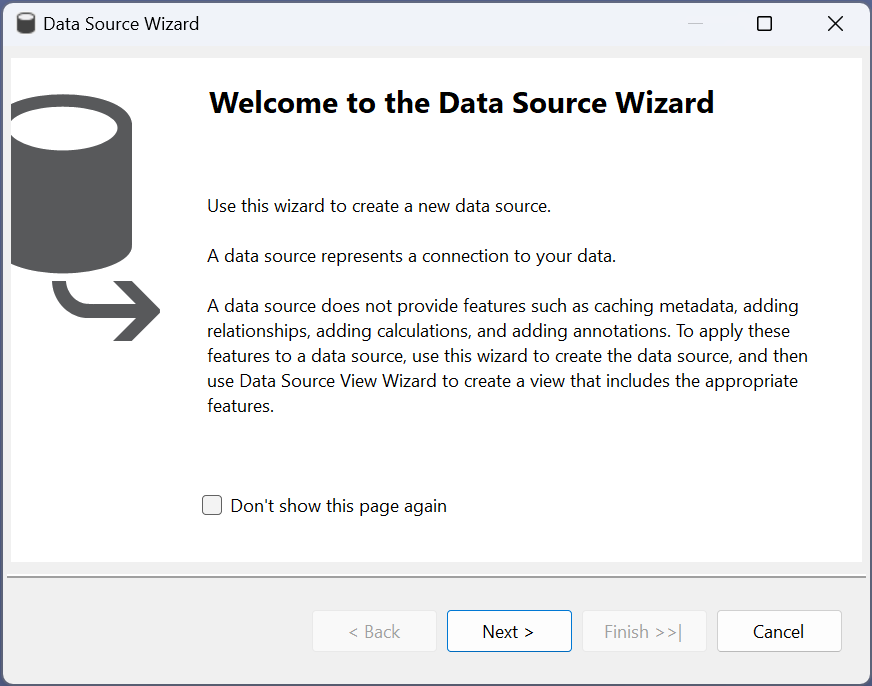
Mô tả được tạo tự động

Sau khi thực thi project, tiến hành kiểm tra dữ liệu tại SQL Server

# SSAS VÀ REPORT

## Tạo các Data Sourses, Cubes, các Dimension

### Data Sources



Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

### Data Sources View

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

* **Cài đặt Relationship cho Goodread Book.dsv**

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

### Cubes

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

* **Kết quả sau khi tạo Cubes**

Ảnh có chứa văn bản, phần mềm, Biểu tượng máy tính, Hệ điều hành

Mô tả được tạo tự động

* **Kết quả sau khi thực thi**

Ảnh có chứa văn bản, phần mềm, máy tính, Phần mềm đa phương tiện

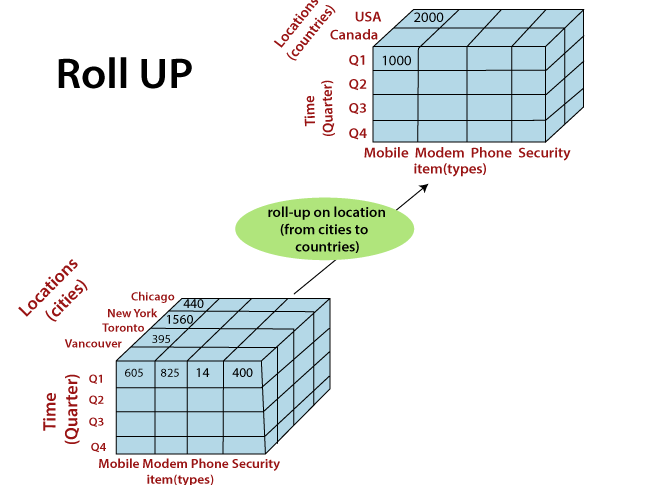
Mô tả được tạo tự động

## Truy vấn

* **Report dạng table**

### Truy vấn Roll-Up:

Truy vấn Roll-Up cho phép người dùng tổng hợp dữ liệu từ mức chi tiết lên mức tổng quát hơn. Nó cho phép xem dữ liệu theo một cách tổng quát hơn, bằng cách tóm tắt dữ liệu từ các chi tiết cụ thể thành các mức tổng quát cao hơn. Ví dụ, từ dữ liệu bán hàng chi tiết hàng ngày, bạn có thể tổng hợp lên thành dữ liệu bán hàng hàng tháng hoặc hàng năm.

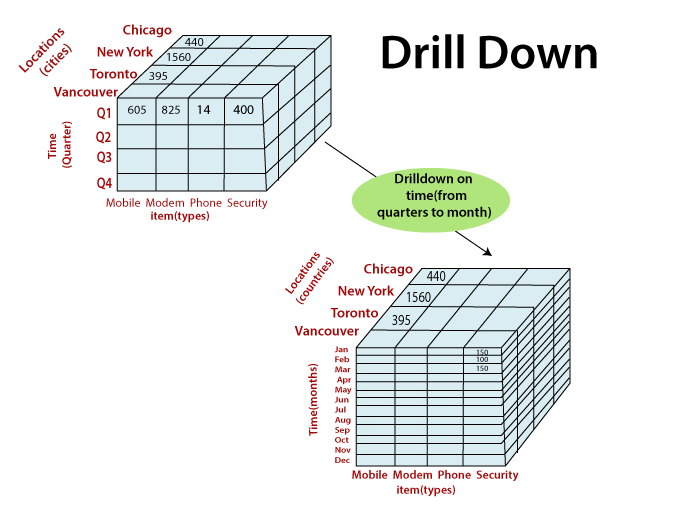
****

--1. Hiển thị tổng số lượng đánh giá của các quyển sách theo từng ngôn ngữ



### Truy vấn Drill-Down:

Trái ngược với Roll-Up, truy vấn Drill-Down cho phép người dùng xem dữ liệu chi tiết hơn từ mức tổng quát. Người dùng có thể "drill down" vào dữ liệu tổng quát để xem chi tiết dữ liệu ở mức cấu trúc thấp hơn. Ví dụ, từ dữ liệu bán hàng hàng năm, bạn có thể drill down để xem dữ liệu bán hàng hàng tháng hoặc hàng ngày.

****

--2. Cho biết quyển sách có số lượng đánh giá cao nhất:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

### Truy vấn Slice and Dice:

Truy vấn Slice and Dice cho phép người dùng cắt (slice) và cắt gọt (dice) dữ liệu để xem theo các tiêu chí hay quan điểm khác nhau. Slice là việc chọn một phần của dữ liệu dựa trên một số tiêu chí nhất định. Dice là việc cắt gọt dữ liệu từ một slice đã chọn và tiếp tục lựa chọn thêm các tiêu chí khác. Ví dụ, từ dữ liệu bán hàng, có thể cắt (slice) dữ liệu chỉ cho một khu vực cụ thể và sau đó cắt gọt (dice) dữ liệu chỉ cho một khoảng thời gian cụ thể trong khu vực đó.

--3. Cho biết quyển sách có số lượng rating tổng > 10000 vào năm 1990

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

--4. Cho biết tác giả có số lượng rating trong năm 2000 lớn hơn 5000

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--5. Hiển thị những quyển sách có cùng nhà xuất bản ra chúng trong năm 2008

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

### Truy vấn Pivot:

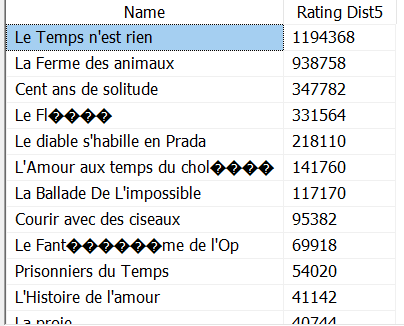
Truy vấn Pivot cho phép người dùng xoay dữ liệu từ dạng dọc thành dạng ngang hoặc ngược lại, tạo ra một bảng dữ liệu mới với các dữ liệu được nhìn từ góc độ khác nhau. Pivot cho phép chuyển đổi dữ liệu giữa các trục ngang và dọc. Ví dụ, từ dữ liệu bán hàng theo sản phẩm và thời gian, bạn có thể pivot để xem dữ liệu bán hàng theo thời gian và từng sản phẩm.

--6. Cho biết quyển sách có bao nhiêu trang trong các quyển sách có lượt rating 5 > 100 trong năm 2000:

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

--7. Hiển thị tổng số lượng rating 5 sao của các quyển sách có ngôn ngữ là tiếng Pháp (fre) theo thứ tự giảm dần:



--8. Hiển thị tổng số lượng rating 3 sao của các quyển sách có ngôn ngữ là tiếng Đức (gre) theo thứ tự tăng dần:

Ảnh có chứa văn bản, ảnh chụp màn hình, số, phần mềm

Mô tả được tạo tự động

--9. Hiển thị tổng số lượng đánh giá, tổng số lượt review dạng văn bản của từng nhà văn theo từng ngôn ngữ

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

### Slice and Dice kết hợp Drill-Down:

--10.Hiển thị top 10 cuốn sách có lượng đánh giá cao nhất, trong đó ở mỗi cuốn sách, chỉ lấy thông tin cuốn sách với đánh giá cao nhất:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

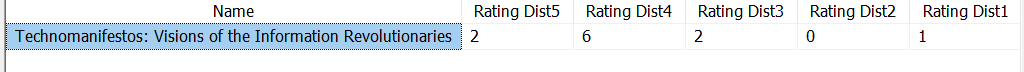
Mô tả được tạo tự động

--11. Cho biết top 5 quyển sách có lượng rating 5 sao cao nhất, trong các công việc cố tổng rating > 10000:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--12. Hiển thị các lượt rating lần lượt 5 sao, 4 sao,...,1 sao của những quyển sách của nhà văn Adam Brate:



--13. Hiển thị top 1 quyển sách trong danh sách số lượng trang sách giảm dần của nhà văn Agatha Christie

*Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động*

### Thao tác lọc cột/dòng (manual):

--Lọc danh sách sách chỉ hiển thị các cột như tên sách, tác giả và đánh giá trung bình.

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

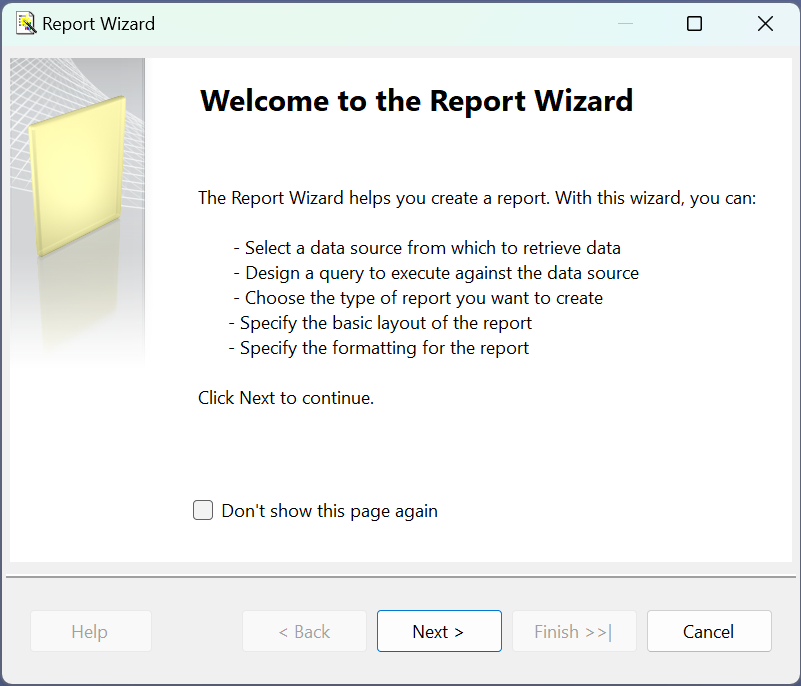
# SSRS

## Report dạng matrix

Tạo một project mới

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động



Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

Bấm Edit. Sau đó nhập Server name để kết nối đến database

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

Ảnh có chứa văn bản, phần mềm, Biểu tượng máy tính, Phần mềm đa phương tiện

Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, phần mềm

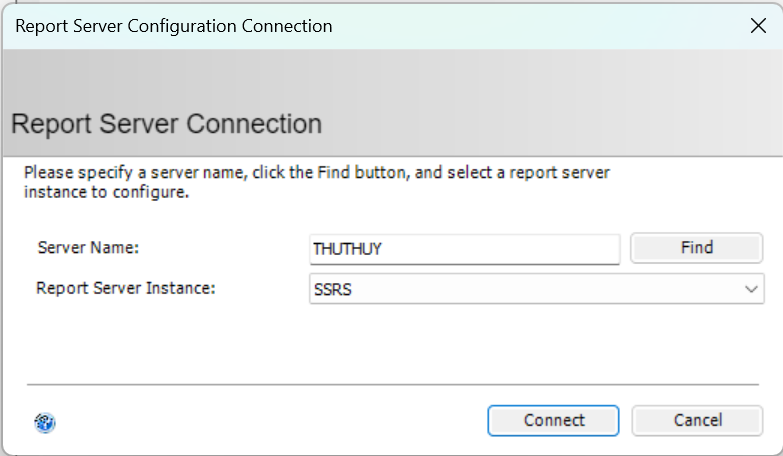
Mô tả được tạo tự động

Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động

Sau khi Design the Query thì tiến hành Design the Report

**Kết nối với**



Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

Câu 1: Thống kê tổng số rating theo ngày xuất bản

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

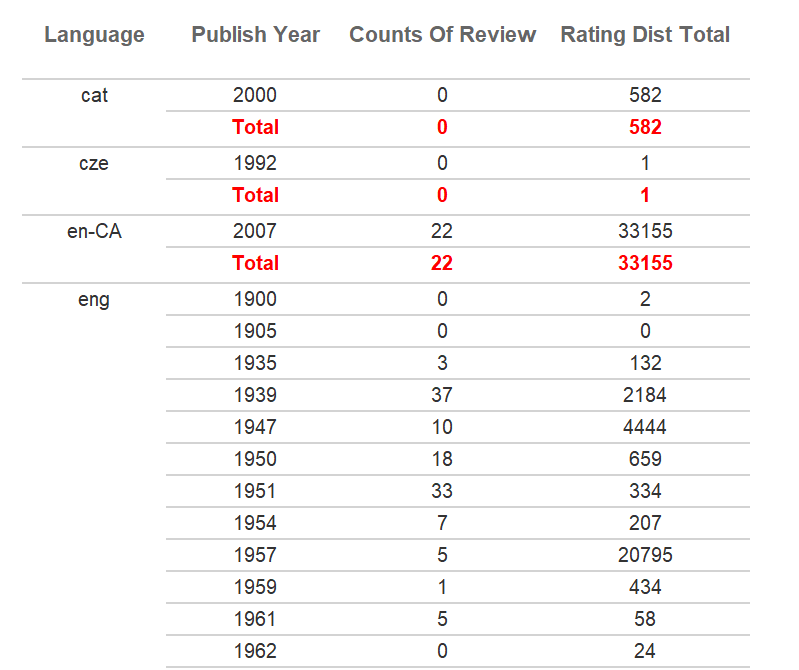
Mô tả được tạo tự động

Excel

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

Câu 2: Thống kê số đánh giá và tổng số lượt review dạng văn bản của từng ngôn ngữ theo năm phát hành



## Report dạng chart

Câu 1:

Chart trên Visual studio

Ảnh có chứa văn bản, hàng, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

Chart trên Excel

Ảnh có chứa văn bản, ảnh chụp màn hình, Sơ đồ, hàng

Mô tả được tạo tự động

Câu 2:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

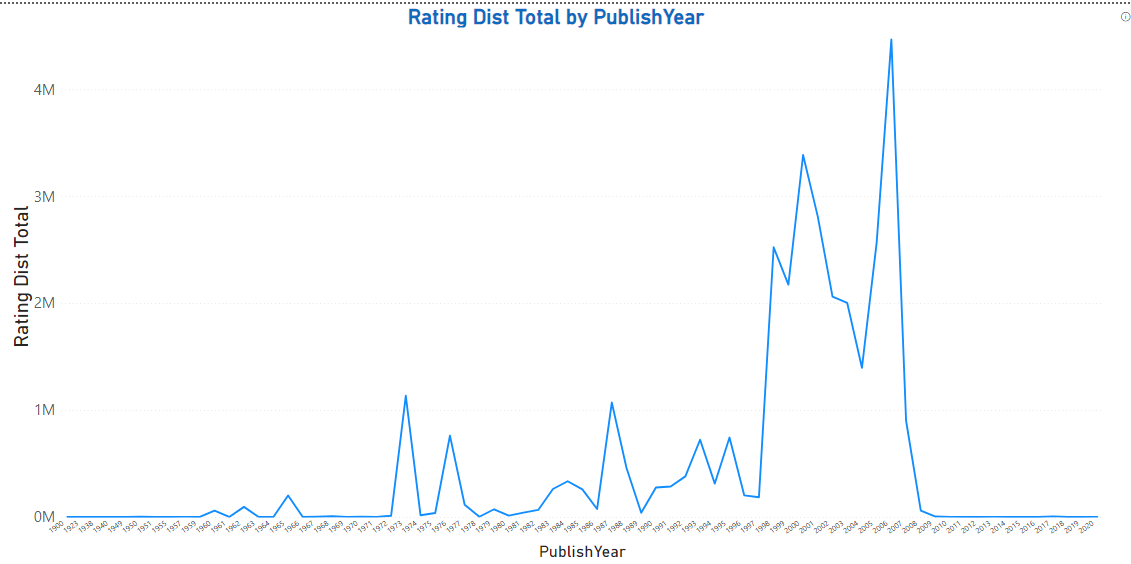
Mô tả được tạo tự động

Chart trên Excel

Ảnh có chứa văn bản, ảnh chụp màn hình, hàng, Sơ đồ

Mô tả được tạo tự động

## Power BI



Thống kê tổng số Rating theo ngày xuất bản

A picture containing text, screenshot, line, plot

Description automatically generated

A picture containing text, screenshot, line, plot

Description automatically generated

Thống kê số đánh giá và tổng số lượt review dạng văn bản theo từng ngôn ngữ theo năm phát hành

# MDX

--Truy vấn Roll-Up:

--1. Hiển thị tổng số lượng đánh giá của các quyển sách theo từng ngôn ngữ

SELECT [Measures].[Rating Dist Total] ON 0,

NON EMPTY [DIM BOOK].[Name] \* [DIM LANGUAGE].[Language].MEMBERS ON 1

FROM [SSIS];

Ảnh có chứa văn bản, số, Phông chữ, chữ viết tay

Mô tả được tạo tự động

--Truy vấn Drill-Down:

--2. Cho biết quyển sách có số lượng đánh giá cao nhất:

SELECT [Measures].[Rating Dist Total] ON 0,

TOPCOUNT ([DIM BOOK].[Name].CHILDREN, 1, [Measures].[Rating Dist Total]) ON 1

FROM [SSIS];

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

--Truy vấn Slice and Dice:

--3. Cho biết quyển sách có số lượng rating tổng > 10000 vào năm 1990

SELECT [Measures].[Rating Dist Total] ON 0,

NON EMPTY [DIM BOOK].[Name].CHILDREN

HAVING [Measures].[Rating Dist Total] > 10000 ON 1

FROM [SSIS]

WHERE [DIM PUBLISHDATE].[Publish Year].&[1990];

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--4. Cho biết tác giả có số lượng rating trong năm 2000 lớn hơn 5000

SELECT [Measures].[Rating Dist Total] ON 0,

FILTER ([DIM AUTHOR].[Authors].CHILDREN, [Measures].[Rating Dist Total] > 5000) ON 1

FROM [SSIS]

WHERE [DIM PUBLISHDATE].[Publish Year].&[2000];

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--5. Hiển thị những quyển sách có cùng nhà xuất bản ra chúng trong năm 2008

SELECT [Measures].[Rating Dist Total] ON 0,

NON EMPTY [DIM PUBLISHER].[ID\_PUBLISHER], [DIM BOOK].[NAME].MEMBERS ON 1

FROM [SSIS]

WHERE [DIM PUBLISHDATE].[Publish Year].&[2008];Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

--Truy vấn Pivot:

--6. Cho biết quyển sách có bao nhiêu trang trong các quyển sách có lượt rating 5 > 100 trong năm 2000:

SELECT [Measures].[Pages Number]ON 0,

FILTER ([DIM BOOK].[Name].CHILDREN, [Measures].[Rating Dist5] > 100) ON 1

FROM [SSIS]

WHERE [DIM PUBLISHDATE].[Publish Year].&[2000];

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--7. Hiển thị tổng số lượng rating 5 sao của các quyển sách có ngôn ngữ là

-- tiếng Pháp (fre) theo thứ tự giảm dần:

SELECT [Measures].[Rating Dist5] ON 0,

NON EMPTY ORDER ([DIM BOOK].[Name].CHILDREN, [Measures].[Rating Dist5], DESC) ON 1

FROM [SSIS]

WHERE [DIM LANGUAGE].[Language].&[fre];

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--8. Hiển thị tổng số lượng rating 3 sao của các quyển sách có ngôn ngữ là

-- tiếng Đức (gre) theo thứ tự tăng dần:

SELECT [Measures].[Rating Dist5] ON 0,

NON EMPTY ORDER ([DIM BOOK].[Name].CHILDREN, [Measures].[Rating Dist5], ASC) ON 1

FROM [SSIS]

WHERE [DIM LANGUAGE].[Language].&[ger];

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--9. Hiển thị tổng số lượng đánh giá, tổng số lượt review dạng văn bản của

-- từng nhà văn theo từng ngôn ngữ

SELECT {[Measures].[Rating Dist Total],[Measures].[Counts Of Review]} ON 0,

NON EMPTY [DIM AUTHOR].[Authors].CHILDREN \* [DIM LANGUAGE].[Language].MEMBERS ON 1

FROM [SSIS];

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

--Slice and Dice kết hợp Drill-Down:

--10.Hiển thị top 10 cuốn sách có lượng đánh giá cao nhất, trong đó ở mỗi cuốn

-- sách, chỉ lấy thông tin cuốn sách với đánh giá cao nhất:

SELECT [Measures].[Rating Dist Total] ON 0,

TOPCOUNT ([DIM BOOK].[Name].CHILDREN, 10, [Measures].[Rating Dist Total]) ON 1

FROM [SSIS];

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--11. Cho biết top 5 quyển sách có lượng rating 5 sao cao nhất, trong các công

-- việc cố tổng rating > 10000:

SELECT [Measures].[Rating Dist5] ON 0,

TOPCOUNT ([DIM BOOK].[Name].CHILDREN, 5, [Measures].[Rating Dist5]) ON 1

FROM

(

SELECT [Measures].[Rating Dist5] ON 0,

FILTER ([DIM AUTHOR].[Authors].CHILDREN, [Measures].[Rating Dist Total] > 10000) ON 1

FROM [SSIS]

);

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

--12. Hiển thị các lượt rating lần lượt 5 sao, 4 sao,...,1 sao của những quyển

-- sách của nhà văn Adam Brate:

SELECT {[Measures].[Rating Dist5],[Measures].[Rating Dist4],[Measures].[Rating Dist3],[Measures].[Rating Dist2],[Measures].[Rating Dist1]} ON 0,

NON EMPTY [DIM BOOK].[Name].CHILDREN ON 1

FROM [SSIS]

WHERE [DIM AUTHOR].[Authors].&[Adam Brate];

Ảnh có chứa văn bản, hàng, Phông chữ, ảnh chụp màn hình

Mô tả được tạo tự động

--13. Hiển thị top 1 quyển sách trong danh sách số lượng trang sách giảm dần

-- của nhà văn Agatha Christie

SELECT [Measures].[Pages Number] ON 0,

TOPCOUNT ([DIM BOOK].[Name].CHILDREN, 1, [Measures].[Rating Dist5]) ON 1

FROM

(

SELECT [Measures].[Pages Number] ON 0,

ORDER ([DIM BOOK].[Name].CHILDREN, [Measures].[Pages Number], DESC) ON 1

FROM [SSIS]

WHERE [DIM AUTHOR].[Authors].&[Agatha Christie]

);

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

# BẢNG PHÂN CÔNG CÔNG VIỆC TRONG NHÓM

|  |  |  |
| --- | --- | --- |
| **Công việc thực hiện** | Nguyễn Cẩm Tú  20520837 | Nguyễn Thị Thu Thủy  20520797 |
| Viết báo cáo | X | X |
| Thực hiện SSIS project. | X |  |
| Thực hiện SSAS project và 10 câu truy vấn. |  | X |
| Thực hiện SSRS project. |  | X |
| Thực hiện MDX với 10 câu truy vấn. | X |  |
| Thực hiện trực quan hóa 10 câu truy vấn MDX trên PowerBI. | X |  |
| Thực hiện trực quan hóa 10 câu truy vấn MDX trên Excel. |  | X |
| Thực hiện Data Mining với 2 thuật toán: Decision Tree và Random Forest. |  | X |
| Quay demo SSIS, SSAS, SSRS. |  | X |
| Quay demo MDX, PowerBI, Excel, Data Mining. | X |  |
| **Mức độ hoàn thành** | 100% | 100% |