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

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

**Logo

Description automatically generated**

**BÁO CÁO ĐỒ ÁN CUỐI KÌ**

***Môn học: Kho dữ liệu và OLAP***

**Đề tài:**

**PHÂN TÍCH DỮ LIỆU KINH DOANH CÁC TRÒ CHƠI VIDEOS GAME TỪ 2016- 2018**

**Giảng viên hướng dẫn: Đỗ Thị Minh Phụng**

**Lớp: IS217.M22.HTCL**

**Sinh viên thực hiện:**

Lê Tiến Vinh MSSV: 19522521

Lê Tuấn Khanh MSSV: 19521681

***TPHCM, 2022***

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

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

**Logo

Description automatically generated**

**BÁO CÁO ĐỒ ÁN CUỐI KÌ**

***Môn học: Kho dữ liệu và OLAP***

**Đề tài:**

**PHÂN TÍCH DỮ LIỆU KINH DOANH CÁC TRÒ CHƠI VIDEOS GAME TỪ 2016- 2018**

**Giảng viên hướng dẫn: Đỗ Thị Minh Phụng**

**Lớp: IS217.M22.HTCL**

**Sinh viên thực hiện:**

Lê Tiến Vinh MSSV: 19522521

Lê Tuấn Khanh MSSV: 19521681

***TPHCM, 2022***

**MỤC LỤC**

[**LỜI CẢM ƠN** 4](#_Toc98626715)

[**NHẬN XÉT CỦA GIẢNG VIÊN** 6](#_Toc98626716)

[**CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU** 7](#_Toc98626717)

[**I.** **Phát biểu dữ liệu** 7](#_Toc98626718)

[**1.1** **Lý do chọn đề tài** 7](#_Toc98626719)

[**1.2** **Mô tả dữ liệu** 8](#_Toc98626720)

[**1.3** **Thuộc tính của dữ liệu** 8](#_Toc98626721)

[***Chú thích cụ thể :*** 9](#_Toc98626722)

[**1.4 Kho dữ liệu đã thực hiện tiền xử lí** 11](#_Toc98626723)

[**2. Xây dựng kho dữ liệu** 13](#_Toc98626724)

[**2.1 Lược đồ kho dữ liệu ( Star Schema)** 13](#_Toc98626725)

[**2.2 Chi tiết các bảng** 14](#_Toc98626726)

[**2.3 Nội dung 15 câu hỏi truy vấn** 15](#_Toc98626727)

[**CHƯƠNG 2: XÂY DỰNG KHO DỮ LIỆU ( SSIS )** 16](#_Toc98626728)

# **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 nhóm chúng 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 nhóm có những kiến thức cơ bản làm nền tảng để thực hiện đề tài này.

Đặc biệt nhóm chúng em xin gửi lời cảm ơn chân thành tới Cô Đỗ Thị Minh Phụng - giảng viên lý thuyết môn Kho dữ liệu và OLAP đã tận tình giúp đỡ, trực tiếp chỉ bảo, hướng dẫn nhóm trong suốt quá trình làm đồ án môn học. Nhờ đó, chúng 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 nhóm chúng em nghĩ đồ án này của nhóm rất khó có thể hoàn thiện được. Một lần nữa, em xin chân thành cảm ơn Cô. Ngoài ra, để đồ án được hoàn thành thì không thể nào cám ơn những người đã làm ra đó, cám ơn các bạn các thành viên trong nhóm đã chăm chỉ và chịu khó hoàn thành nhiệm vụ đúng tiến độ.

Trong thời gian một học kỳ thực hiện đề tài, nhóm tác giả đã cố gắng 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 để ứng dụng vào thực hiện đề tài “Phân tích dữ liệu kinh doanh các trò chơi video games từ năm 2016 - 2018”. Tuy nhiên, do kiến thức và kinh nghiệm còn hạn hẹp, nhóm tác giả sẽ khó tránh khỏi những thiếu sót. Chính vì vậy, nhóm rất mong nhận được những sự góp ý từ phía các Thầy nhằm hoàn thiện những kiến thức mà nhóm tác giả đã học tập và là hành trang để nhóm tác giả thực hiện tiếp các đề tài khác trong tương lai.

Xin chân thành cảm ơn Cô !

Nhóm sinh viên thực hiện

# **NHẬN XÉT CỦA GIẢNG VIÊN**

…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

# **CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ DỮ LIỆU**

1. **Phát biểu dữ liệu**
   1. **Lý do chọn đề tài**

Bên cạnh sự phát triển về công nghệ vào thời đại 4.0 như hiện nay, các trò chơi video games ngày càng được giới trẻ cũng như thanh niên trên toàn thế giới yêu thích như một công cụ giải trí vào những thời gian rảnh, hoặc sau những giờ làm việc căng thẳng. Cũng chính vì lí do đó, thị trường dần xuất hiện đa dạng các loại sản phẩm giúp con người có thể nâng cao chất lượng trải nghiệm với trò chơi. Với lượng sản phẩm đa dạng như vậy, nhiều thử thách được đặt ra với các doanh nghiệp hoạt động trong lĩnh vực buôn bán và thương mại, đặc biệt nhất có thể kể tới là việc phải cân nhắc, chọn lọc và đưa ra được những sản phẩm phù hợp với nhu cầu của khách hàng.

A picture containing graphical user interface

Description automatically generated

Do số lượng người chơi và số lượng nhà phát triển ngày càng tăng, việc phân tích dữ liệu là vô cùng cần thiết để có thể có cái nhìn tổng quát về chất lượng cũng như sự phù hợp của trò chơi với từng lứa tuổi. Bên cạnh đó, việc phân tích dữ liệu về các trò chơi video games có sự tham gia của các nhà phê bình cũng như của chính những khách hàng sử dụng trò chơi cũng giúp những nhà phát triển có được cái nhìn, có được số điểm khách quan về trò chơi của mình, qua đó có thể sửa chữa, cải thiện, làm mới và phát triển thế giới trò chơi video games hơn.

* 1. **Mô tả dữ liệu**
* Kho dữ liệu Videos Game Sales with Rating là một kho dữ liệu thu nhập thông tin kinh doanh của các tựa trò chơi video games với rating và các số liệu khác trong khoảng thời gian từ 2016 – 2018.
* Thông qua kho dữ liệu người dùng có thể biết được thông tin về tựa game, nền tảng phát triển của game, năm mà game được phát hành, nhà phát hành tựa game, nhà phát triển game, rating của games ở các quốc gia hàng đầu, và nhiều thông tin khác,…
* Kho dữ liệu gồm 16708 dòng và 16 thuộc tính.
* Link gốc: <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>.
  1. **Thuộc tính của dữ liệu**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên Thuộc Tính** | **Kiểu** | **Ý Nghĩa Thuộc Tính** | **Gía trị** |
| 1 | Video\_Name | string | Tên của trò chơi | Khoảng hơn 60 tựa game. |
| 2 | Platform | string | Nền tảng mà trò chơi sử dụng | 3DS, DC, DS, GBA, GC, PS, PC, PS2, PS3, PS4, PSP, PSV, Wii, WiiU, X360, XB, X1. |
| 3 | Year\_of\_Realease | date | Năm mà trò chơi được phát hành | Trong khoảng 1985 tới 2016 |
| 4 | Gener | string | Hay còn gọi là Category, thể loại của trò chơi. | Action, Adventure, Fighting, Misc, Platform, Puzzel, Racing, Role-Playing, Shooter, Simulation, Sports, Stagery. |
| 5 | Publisher | string | Nhà phát hành | Khoảng hơn 20 nhà phát hành. |
| 6 | NA\_sales | int | Số liệu doanh thu ở khu vực NA.( Bắc Mĩ) | Trong khoảng 0 đến 41.36 |
| 7 | EU\_sales | int | Số liệu doanh thu ở khu vực EU. ( Châu Âu) | Trong khoảng 0 đến 28.96 |
| 8 | JP\_sales | int | Số liệu doanh thu ở Nhật bản | Trong khoảng 0 đến 6.5 |
| 9 | Other sales | int | Các số liệu doanh thu khác | Trong khoảng 0 đến 10.57 |
| 10 | Global sales | int | Số liệu doanh thu toàn cần | Trong khoảng 0 đến 82.53 |
| 11 | Critic\_Score | int | Điểm phê bình của trò chơi | Trong khoảng 13 đến 98 |
| 12 | Critic\_Count | int | Tổng sổ phê bình của trò chơi | Trong khoảng 3 đến 113 |
| 13 | User\_Score | int | Điểm của người dùng cho trò chơi | Trong khoảng 0.5 đến 9.6 |
| 14 | User\_Count | int | Tổng số người dùng tham gia khảo sát | Trong khoảng 4 đến 10665 |
| 15 | Developer | string | Nhà phát triển | Khoảng hơn 20 nhà phát triển trò chơi. |
| 16 | Rating | decimal | Đánh giá trò chơi theo độ tuổi của người chơi( từ trẻ em cho đến người lớn, theo các thang mức E,T, E10+,…) | AO, E, E10+, K-A, M, RP, T. |

***Chú thích cụ thể:***

Khi các bài đánh giá về một bộ phim nhất định được tích lũy, điểm Rotten Tomatoes đo lường tỷ lệ phần trăm tích cực hơn tiêu cực và ấn định đánh giá tổng thể mới hoặc thấp cho bộ phim. Điểm trên 60 phần trăm được coi là tốt, và điểm từ 59 phần trăm trở xuống là xấu.

ESRB sử dụng 7 xếp hạng khác nhau cho các game. 4 xếp hạng trong số này rất phổ biến, trong khi 2 tùy chọn khác khá hiếm gặp và loại xếp hạng cuối cùng chỉ mang tính "giữ chỗ".

* **Early Childhood (EC)** là xếp hạng thấp nhất. Nó biểu thị các game dành cho khán giả nhí. Do đó, những tựa game này không có nội dung phản cảm và có thể không mấy thú vị đối với khán giả nói chung, vì chúng dành cho trẻ nhỏ. Xếp hạng này không phổ biến lắm. Các game ví dụ bao gồm Dora the Explorer: Dance to the Rescue và Bubble Guppies.
* **Everyone (E)** là xếp hạng cơ sở. Các game với xếp hạng này có nội dung “phù hợp với mọi lứa tuổi”. Trước năm 1998, xếp hạng này được gọi là Kids to Adults (KA). Các game được xếp hạng E bao gồm Mario Kart 8 Deluxe và Rocket League.
* **Everyone 10+ (E10+)** biểu thị các game phù hợp cho trẻ em từ 10 tuổi trở lên. So với một game được xếp hạng E, những tựa game này có thể chứa một số nội dung mang tính khêu gợi, hài hước hoặc bạo lực nhiều hơn. Đáng chú ý, đây là xếp hạng duy nhất mà ESRB đã thêm kể từ khi thành lập. Một số game có đánh giá này là Super Smash Bros. Ultimate và Kingdom Hearts III.
* **Teen (T)** là cấp độ tiếp theo. Đánh giá này phù hợp cho người chơi từ 13 tuổi trở lên. Các tiêu đề có thể có nội dung khơi gợi tính dục, ngôn ngữ mạnh hơn và có những cảnh liên quan đến máu me. Bạn sẽ tìm thấy xếp hạng Teen trên các game như Apex Legends và Fortnite.
* **Mature (M)** là xếp hạng “thông thường” ở mức cao nhất. Các game xếp hạng M được coi là phù hợp cho những người từ 17 tuổi trở lên. So với các tựa game Teen, chúng có thể chứa những cảnh bạo lực dữ dội hơn, nội dung tình dục mạnh hơn, thậm chí cả hình ảnh khoả thân và ngôn ngữ khá mạnh. Một số “kho” game không bán các game được xếp hạng M cho trẻ vị thành niên, nhưng đây không phải là một tiêu chuẩn hợp pháp. Các tựa game ví dụ được xếp hạng M bao gồm Red Dead Redemption II và Assassin’s Creed Odyssey.
* **Adults Only (AO)** là xếp hạng 18+ của ESRB. Mức xếp hạng này được áp dụng cho các game có nội dung tình dục đồ họa hoặc những game cho phép đánh bạc bằng tiền thật. Tuy nhiên, nó thực sự là một đánh giá khập khiễng. Không có nhà sản xuất console lớn nào cho phép các game AO trên hệ thống của họ và rất ít nhà bán lẻ bán những game AO trong “kho” game của mình.

Chính vì vậy, chỉ có một số ít các game nhận được mức xếp hạng này. Hầu hết các game AO nhận được đánh giá này do nội dung nặng về tình dục. Các nhà xuất bản sẽ thay đổi game của mình để tránh bị xếp hạng ở mức này, vì về cơ bản, đó giống như một bản án tử hình. Các game có xếp hạng AO bao gồm Seduce Me và Ef: A Fairy Tale of the Two.

* **Rating Pending (RP)** chỉ mang tính “giữ chỗ”. Nó xuất hiện cùng với quảng cáo cho các game chưa được đánh giá.

A screenshot of a computer

Description automatically generated with medium confidence

## **1.4 Kho dữ liệu đã thực hiện tiền xử lí**

* Sau khi dữ liệu được xử lí, sẽ bao gồm 16708 dòng và 16 cột.

Table

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên Thuộc Tính** | **Kiểu** | **Ý Nghĩa Thuộc Tính** | **Gía trị** |
| 1 | Video\_Name | string | Tên của trò chơi | Khoảng hơn 60 tựa game. |
| 2 | Platform | string | Nền tảng mà trò chơi sử dụng | 3DS, DC, DS, GBA, GC, PS, PC, PS2, PS3, PS4, PSP, PSV, Wii, WiiU, X360, XB, X1. |
| 3 | Year\_of\_Realease | date | Năm mà trò chơi được phát hành | Trong khoảng 1985 tới 2016 |
| 4 | Gener | string | Hay còn gọi là Category, thể loại của trò chơi. | Action, Adventure, Fighting, Misc, Platform, Puzzel, Racing, Role-Playing, Shooter, Simulation, Sports, Stagery. |
| 5 | Publisher | string | Nhà phát hành | Khoảng hơn 20 nhà phát hành. |
| 6 | NA\_sales | int | Số liệu doanh thu ở khu vực NA.( Bắc Mĩ) | Trong khoảng 0 đến 41.36 |
| 7 | EU\_sales | int | Số liệu doanh thu ở khu vực EU. ( Châu Âu) | Trong khoảng 0 đến 28.96 |
| 8 | JP\_sales | int | Số liệu doanh thu ở Nhật bản | Trong khoảng 0 đến 6.5 |
| 9 | Other sales | int | Các số liệu doanh thu khác | Trong khoảng 0 đến 10.57 |
| 10 | Global sales | int | Số liệu doanh thu toàn cần | Trong khoảng 0 đến 82.53 |
| 11 | Critic\_Score | int | Điểm phê bình của trò chơi | Trong khoảng 13 đến 98 |
| 12 | Critic\_Count | int | Tổng sổ phê bình của trò chơi | Trong khoảng 3 đến 113 |
| 13 | User\_Score | int | Điểm của người dùng cho trò chơi | Trong khoảng 0.5 đến 9.6 |
| 14 | User\_Count | int | Tổng số người dùng tham gia khảo sát | Trong khoảng 4 đến 10665 |
| 15 | Developer | string | Nhà phát triển | Khoảng hơn 20 nhà phát triển trò chơi. |
| 16 | Rating | decimal | Đánh giá trò chơi theo độ tuổi của người chơi( từ trẻ em cho đến người lớn, theo các thang mức E,T, E10+,…) | AO, E, E10+, K-A, M, RP, T. |

# **2. Xây dựng kho dữ liệu**

## **2.1 Lược đồ kho dữ liệu ( Star Schema)**

Graphical user interface

Description automatically generated with medium confidence

## **2.2 Chi tiết các bảng**

* **Bảng FACT:**

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ràng buộc |
| Date\_ID | Date | Khóa chính, khóa ngoại |
| Video\_Name | String | Khóa chính, khóa ngoại |
| NA\_Sales | Int |  |
| EU\_Sales | Int |  |
| JP\_Sales | Int |  |
| Other\_Sales | Int |  |
| Global\_Sales | Int |  |
| Critic\_Score | Int |  |
| Critic\_Count | Int |  |
| User\_Score | Int |  |
| User\_Count | Int |  |

* **Bảng Dim\_Date :**

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ràng buộc |
| Date\_ID | Date | Khóa chính, khóa ngoại |
| Year\_of\_Realease | Date |  |

* **Bảng Dim\_VideoGames:**

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ràng buộc |
| Video\_Name | String | Khóa chính, khóa ngoại |
| Gener | String |  |
| Publisher | String |  |
| Developer | String |  |
| Rating | decimal |  |
| Platform | String |  |

## **2.3 Nội dung 15 câu hỏi truy vấn**

**15 câu truy vấn :**

1. Tìm ra trò chơi thuộc nền tảng … có đánh giá là M ( mature), được phát hành năm 20xx và đạt mức doanh thu bán ra lớn hơn 5. tại châu âu (EU)
2. Tìm ra tên trò chơi, năm phát hành, nhà sản xuất, nhà phát triển của những tựa game có đánh giá là E10.
3. Tìm nền tảng có trò chơi với đánh giá thuộc E, được phát hành năm 2007 và có số điểm đánh giá phê bình lớn hơn 7, bé hơn 8 và số người đánh giá phê bình >100 người.
4. Tìm ra thông tin top 10 trò chơi có doanh thu lớn nhất và nhỏ nhất trên toàn thể giới.
5. Liệt kê những trò chơi có đánh giá phê bình lớn hơn 50 với số người đánh giá trong khoảng 70-80 người.
6. Đưa ra thông tin tựa game, năm phát hành, nhà xuất bản, nhà phát triển, tổng doanh thu toàn cầu của từng trò chơi thuộc thể loại SPORTS.
7. Đưa ra top 3 trò chơi có mức thu nhập thấp nhất tại Châu Âu.
8. Cho biết thông tin của nhà phát hành nào có tựa game thuộc đánh giá T và sở hữu điểm đánh giá phê bình từ người sử dụng lớn hơn 8.5.
9. Thống kê doanh thu ở toàn thế giới và phần trăm đánh giá từng loại nền tảng trò chơi sắp xếp theo thứ tự phần trăm đánh giá giảm dần.
10. Cho biết thông tin nhà phát triển có ít sản phẩm nhất và có nhiều sản phẩm nhất.
11. Liệt kê những trò chơi thuộc top 5 doanh thu cao nhất.
12. Với mỗi nền tảng, đưa ra top 3 trò chơi có doanh thu cao nhất, tính trên toàn thế giới.
13. Đưa ra doanh thu của những trò chơi được phát hành trong khoảng thời gian từ 1999- 20008 với doanh thu bán ra tại Nhật Bản lớn hơn 1.42.
14. Liệt kê top 10 những trò chơi được phát hành trong năm 2016 có doanh thu cao nhất tại các châu lục còn lại ( tức là trừ Nhật Bản, Châu Âu, và Bắc Mĩ).
15. Liệt kê top 3 nền tảng có những sản phẩm được phát hành năm 2014 có đánh giá thuộc loại T, có đánh giá từ người dùng phải lớn hơn hoặc bằng 7 và tổng doanh thu trên toàn thế giới của nền tảng đó với tổng các trò chơi phải lớn hơn 700.

# **CHƯƠNG 2: XÂY DỰNG KHO DỮ LIỆU ( SSIS )**