**Ảnh có chứa quảng trường

Mô tả được tạo tự động** **ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

Ảnh có chứa văn bản, mẫu họa

Mô tả được tạo tự động

**BÁO CÁO ĐỒ ÁN**

**KHO DỮ LIỆU VÀ OLAP – IS217.O21**

**ĐỀ TÀI:**

**Phân tích dữ liệu về thông số trận đấu tại giải bóng rổ NBA ở Mỹ**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn**:  Ths.Đỗ Thị Minh Phụng | |
| **Sinh viên thực hiện:**  Nguyễn Trần Gia Kiệt – 21522258  Vương Thanh Linh – 21521082 | |
|  |  |

# LỜI CẢM ƠN

Trong cuộc sống của chúng ta, có lẽ ai cũng đã từng thất bại hoặc thành công, dù như thế nào thì đó cũng là kết quả nỗ lực của mỗi cá nhân cũng như tập thể. Và đằng sau đó chính là sự hỗ trợ giúp đỡ từ mọi người. Xét về mặt thành công, trong thực tế không có sự thành công nào mà không có sự giúp đỡ, nhất là trong học tập. Dân gian ta có câu “Không thầy đố mày làm nên” quả thật là đúng, học sinh không thể thành công nếu không có sự giúp đỡ, truyền đạt cũng như chỉ bảo tận tình của người Thầy. Hôm nay, để có thể hoàn thành được đồ án môn học này, nhóm chúng em rất biết ơn những thầy cô đã hỗ trợ tận tình, đã cung cấp cho chúng em nhiều kiến thức cũng như kinh nghiệm của người đi trước.

Lời đầu tiên, nhóm Nhóm 18 chúng em xin gửi 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 chúng em có những kiến thức cơ bản làm nền để thực hiện đề tài này.

Đặc biệt nhóm chúng em gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới Thạc sĩ Đỗ Thị Minh Phụng đã trực tiếp hướng dẫn tận tình, sửa chữa và đóng góp nhiều ý kiến quý báo giúp chúng em hoàn thành tốt báo cáo môn học của mình. Ngoài ra, chúng em cũng gửi lời cảm ơn đến tập thể lớp *IS217.O21* vì thời gian qua đã đồng hành cùng nhau. Cùng nhau học tập, tranh luận một cách sôi nổi để xây dựng bài học một cách tốt nhất. Cảm ơn các bạn đã cùng thảo luận, đánh giá và đóng góp ý kiến, cùng học hỏi nghiên cứu để thực hiện đồ án một cách tốt nhất có thể.

Cuối cùng, 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 chúng 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, nhóm chúng 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, nhóm chúng em không tránh khỏi những thiếu sót. Chính vì vậy, nhóm chúng em rất mong nhận được những sự đóng góp từ phía thầy cô nhằm hoàn thiện những kiến thức mà nhóm chúng em đã học tập và là hành trang để nhóm chúng em thực hiện tiếp các đề tài khác trong tương lai.

Một lần nữa, chúng em xin được gửi lời cảm ơn chân thành nhất của mình đến thầy, cô, chúng em chúc thầy, cô thật dồi dào sức khoẻ và thành công trên con đường giảng dạy của mình!

Thành phố Hồ Chí Minh, năm 2024

**Nhóm sinh viên thực hiện**

# NHẬN XÉT CỦA GIẢNG VIÊN

*……., ngày……...tháng……năm 2024*

**Người nhận xét**

*(Ký tên và ghi rõ họ tên****)***

# MỤC LỤC

[LỜI CẢM ƠN 1](#_Toc162716320)

[NHẬN XÉT CỦA GIẢNG VIÊN 3](#_Toc162716321)

[MỤC LỤC 5](#_Toc162716322)

[CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ KHO DỮ LIỆU 7](#_Toc162716323)

[1.1 Giới thiệu về dữ liệu 7](#_Toc162716324)

[1.1.1 Lý do chọn đề tài 7](#_Toc162716325)

[1.1.2 Giới thiệu dataset 7](#_Toc162716326)

[1.2 Mô tả các thuộc tính 8](#_Toc162716327)

[1.3 Thiết kế kho dữ liệu 11](#_Toc162716328)

[1.3.1 Sơ đồ hình sao minh hoạ 11](#_Toc162716329)

[1.3.2 Các bảng chiều 11](#_Toc162716330)

[1.3.2.1 DimGame 11](#_Toc162716331)

[1.3.2.2 DimTime 12](#_Toc162716332)

[1.3.2.3 DimGameType 12](#_Toc162716333)

[1.3.2.4 DimTeam 12](#_Toc162716334)

[1.3.3 Bảng sự kiện 13](#_Toc162716335)

[1.3.3.1 FactGameStats 13](#_Toc162716336)

[1.4 Câu truy vấn 13](#_Toc162716337)

[CHƯƠNG 2: QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS) 15](#_Toc162716338)

[2.1 Chuẩn bị cơ sở dữ liệu 15](#_Toc162716339)

[2.2 Tạo mới dự án SSIS 20](#_Toc162716340)

[2.3 Mô hình SSIS 29](#_Toc162716341)

[2.4 Sequences 29](#_Toc162716342)

[2.4.1 Sequence Reset Database 29](#_Toc162716343)

[2.4.2 Sequence Load Dimension Tables 34](#_Toc162716344)

[2.4.2.1 Load DimTeam 35](#_Toc162716345)

[2.4.2.2 Load DimGameType 39](#_Toc162716346)

[2.4.2.3 Load DimTime 43](#_Toc162716347)

[2.4.2.4 Load DimGame 52](#_Toc162716348)

[2.4.3 Sequence Create Foreign Key 55](#_Toc162716349)

[2.4.4 Sequence Load Fact Table 60](#_Toc162716350)

[2.5 Chạy dự án và kiểm tra 79](#_Toc162716351)

[2.5.1 Chạy trên Visual Studio 79](#_Toc162716352)

[2.5.2 Kiểm tra dữ liệu trên Microsoft SQL Server Management Studio 80](#_Toc162716353)

[2.5.2.1 Bảng DimGame 80](#_Toc162716354)

[2.5.2.2 Bảng DimGameType 80](#_Toc162716355)

[2.5.2.3 Bảng DimTeam 80](#_Toc162716356)

[2.5.2.4 Bảng DimTime 81](#_Toc162716357)

[2.5.2.5 Bảng FactGameStats 81](#_Toc162716358)

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

## Giới thiệu về dữ liệu

### Lý do chọn đề tài

* **Sự phổ biến của NBA**: NBA là một trong những giải đấu bóng rổ nổi tiếng và phổ biến nhất trên toàn thế giới. Với sự hâm mộ rộng lớn, việc phân tích dữ liệu từ giải đấu này sẽ mang lại thông tin hữu ích và thú vị cho một đối tượng người đọc rất lớn.
* **Dữ liệu phong phú**: NBA cung cấp một lượng lớn dữ liệu từ mỗi trận đấu, bao gồm điểm số, thống kê cá nhân của cầu thủ, tỷ lệ thành công, và nhiều thông số khác. Điều này tạo điều kiện cho việc thực hiện phân tích sâu sắc để hiểu rõ hơn về các yếu tố ảnh hưởng đến kết quả của trận đấu.
* **Phân tích dữ liệu thể thao**: Bóng rổ là một trong những môn thể thao mà dữ liệu có thể được sử dụng rất hiệu quả để phân tích và dự đoán kết quả. Việc nghiên cứu về các mô hình dữ liệu trong bóng rổ có thể cung cấp thông tin giá trị cho các nhà quản lý đội bóng, người chơi, và cả những người cá cược.

### Giới thiệu dataset

* Tên dataset: NBA Traditional Boxscores 1997-2023
* Lần cập nhập gần nhất: 02 tháng 3 năm 2024
* Dataset bao gồm 29 cột thuộc tính và 65574 dòng dữ liệu chứa các điểm số của cầu thủ và đội từ mùa giải NBA 1996-97 đến 2022-23 được lấy trên web từ NBA.com.
* Dataset được một dataset expert có tên Szymon Jóźwiak cung cấp dựa vào dữ liệu trên NBA.com.
* Nguồn tải dataset: <https://www.kaggle.com/datasets/szymonjwiak/nba-traditional>

## Mô tả các thuộc tính

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Ý nghĩa | Kiểu dữ liệu |
|  | Gameid | NBA.com Game ID. | Long |
|  | Date | Ngày diễn ra trận đầu. | Date |
|  | Type | Loại của trận đấu. | Varchar |
|  | Teamid | NBA.com Team ID. | Long |
|  | Team | Tên đội bóng. | Varchar |
|  | Home | Đội bóng sân nhà. | Varchar |
|  | Away | Đội bóng sân khách. | Varchar |
|  | Min | Số phút thi đấu của một cầu thủ hoặc một đội. | Int |
|  | PTS | Số điểm đội bóng ghi được trong một trận đấu. | Int |
|  | FGM | Số lượng cú ném thành công mà cầu thủ hoặc đội làm được trong “field”. Bao gồm ném 2 điểm và ném 3 điểm. | Int |
|  | FGA | Số lần đội bóng hoặc cầu thủ cố gắng đưa bóng vào rổ từ bên trong các vùng “field” trong một trận đấu. | Int |
|  | FG% | Tỉ lệ ném thành công từ “field”. | Double |
|  | 3PM | Số bàn thắng 3 điểm mà cầu thủ hoặc đội ghi được trong một trận đấu. | Int |
|  | 3PA | Tổng số lần cầu thủ hoặc đội thực hiện cú ném 3 điểm, kể cả thành công hoặc không thành công. | Int |
|  | 3P% | Tỉ lệ thành công của các cú ném 3 điểm mà cầu thủ đội đó thực hiện | Double |
|  | FTM | Số quả ném phạt mà cầu thủ hoặc đội đó đã thực hiện và thành công. | Int |
|  | FTA | Số quả ném phạt mà cầu thủ hoặc đội đó đã thực hiện. | Int |
|  | FT% | Tỉ lệ thành công của các cú ném phạt được đội hoặc cầu thủ thực hiện. | Double |
|  | OREB | Số lần cầu thủ hoặc đội giành lại được bóng khi đội mình thực hiện cú ném không thành công từ vị trí tấn công. | Int |
|  | DREB | Số lần cầu thủ hoặc đội giành lại được bóng khi đội đối thủ thực hiện cú ném không thành công từ vị trí tấn công. | Int |
|  | REB | Tổng số lần Rebound cầu thủ hoặc đội bóng thực hiện được | Int |
|  | AST | Số lần kiến tạo của cầu thủ thuộc đội bóng đó thực hiện được | Int |
|  | TOV | Số lần đội mất quyền kiểm soát bóng vào tay đội đối thủ. | Int |
|  | STL | Số lần cầu thủ hoặc đội cướp được bóng từ đối thủ mà không gây ra lỗi. | Int |
|  | BLK | Số lần cầu thủ hoặc đội chặn cú ném từ đội đối thủ, ngăn bóng vào rổ | Int |
|  | PF | Số lỗi cá nhân mà cầu thủ hoặc đội đó vi phạm. | Int |
|  | +/- | “plus-minus” tổng số điểm đội mình ghi được trừ cho số điểm đội đối phương ghi được trong thời gian 2 đội thi đấu. | Int |
|  | Win | Đội chiến thắng (đội thắng sẽ là 1 còn đội thua là 0) | Int |
|  | season | Năm mùa giải diễn ra | Int |

## Thiết kế kho dữ liệu

### Sơ đồ hình sao minh hoạ

A screenshot of a computer

Description automatically generated

### Các bảng chiều

#### DimGame

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu** | **Ràng buộc** | **Ý nghĩa** |
|  | gameKey | int | Primary |  |
|  | gameId | float |  |  |
|  | 1stTeam | nvarchar(255) |  | Đội chủ nhà |
|  | 2ndTeam | nvarchar(255) |  | Đội khách |
|  | 1stTeamPoint | Int |  | Điểm đội chủ nhà |
|  | 2ndTeamPoint | Int |  | Điểm đội khách |
|  | duration | Int |  |  |
|  | season | Int |  |  |

#### DimTime

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu** | **Ràng buộc** | **Ý nghĩa của thuộc tính** |
|  | timekey | int | Primary Key |  |
|  | gameId | float | Foreign Key |  |
|  | day | int |  |  |
|  | month | int |  |  |
|  | year | int |  |  |

#### DimGameType

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu** | **Ràng buộc** | **Ý nghĩa của thuộc tính** |
|  | gameTypeKey | int | Primary Key |  |
|  | typeName | nvarchar(255) |  | Tên loại trận đấu |

#### DimTeam

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu** | **Ràng buộc** | **Ý nghĩa** |
|  | teamKey | int | Primary key |  |
|  | teamId | float |  |  |
|  | teamName | nvarchar(255) |  |  |

### Bảng sự kiện

#### 1.3.3.1 FactGameStats

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu** | **Ràng buộc** | **Ý nghĩa** |
|  | statsId | int | Primary Key | Thứ tự thống kê |
|  | gameKey | int | Foreign Key |  |
|  | gameTypeKey | int | Foreign Key |  |
|  | teamKey | int | Foreign Key |  |
|  | timeKey | int | Foreign Key |  |
|  | FGM | int |  |  |
|  | FGA | int |  |  |
|  | 3PM | int |  |  |
|  | 3PA | int |  |  |
|  | FTM | int |  |  |
|  | FTA | int |  |  |
|  | OREB | int |  |  |
|  | BREB | int |  |  |
|  | REB | int |  |  |
|  | TOV | int |  |  |
|  | STL | int |  |  |

## Câu truy vấn

1. Tổng số trận đấu của mỗi đội bóng?
2. Tổng số điểm ghi được bởi mỗi đội trong mùa giải?
3. Tổng số lần cản phá của mỗi đội?
4. Số lần cản phá cao nhất của mỗi đội trong một trận đấu?
5. Trung bình số lần cắt bóng của mỗi đội?
6. Tổng số lần cướp bóng của mỗi đội?
7. Trung bình số lần phát bóng thành công từ ngoài vòng 3 điểm của mỗi đội?
8. Trung bình số lần phát bóng thành công từ trong vòng 3 điểm của mỗi đội?

# 

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

## 2.1 Chuẩn bị cơ sở dữ liệu

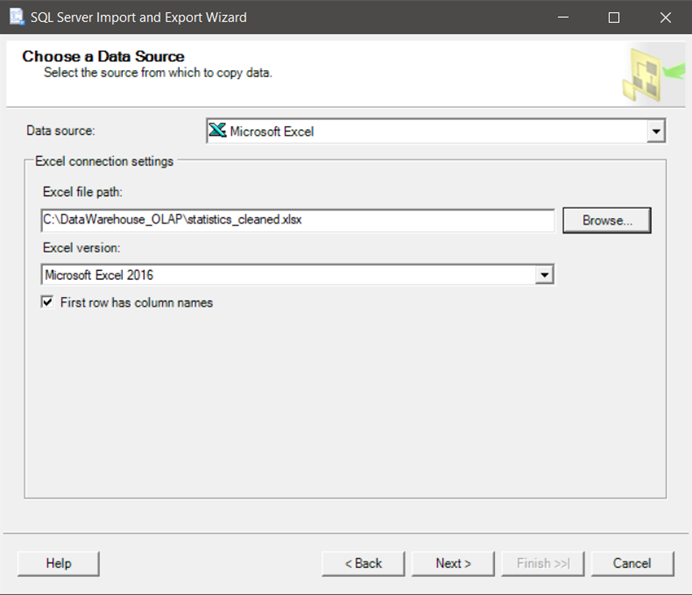
**- Bước 1:** Mở phần mềm **Microsoft SQL Server Management Studio** thực hiện tạo 2 database là **NBAStats** và **NBAStatsDW** – tương ứng với dữ liệu gốc và kho dữ liệu

A close-up of a logo

Description automatically generated

- **Bước 2:** Đổ dữ liệu thu thập được vào cơ sở dữ liệu **NBAStats**

+ **Bước 2.1:** Sử dụng phần mềm **SQL Server Import and Export Wizard** để thực hiện đổ dữ liệu từ file Excel vào cơ sở dữ liệu. Chọn **Data Source** là **Microsoft Excel**, sau đó chọn **Next**



+ **Bước 2.2:** Chọn cơ sở dữ liệu đích để đổ vào, ở đây cần chỉnh **Integrated Security** thành **True, Data Source** là **localhost** và **Initial Catalog** là tên cơ dữ liệu, ở đây là **NBAStats**, sau đó chọn **Next**

A screenshot of a computer

Description automatically generated

+ **Bước 2.3:** Ở bước này, tick chọn vào **Copy data from one or more tables or views** để ánh xạ các cột trong file Excel thành các cột trong bảng dữ liệu, sau đó chọn **Next**

A screenshot of a computer screen

Description automatically generated

+ **Bước 2.4:** Chỉnh sửa lại tên cho bảng, chọn **Edit Mappings** để lựa chọn thuộc tính ánh xạ và **Preview** để xem trước dữ liệu được đổ vào, sau cho đó chọn **Next**

A screenshot of a computer

Description automatically generated

+ **Bước 2.5:** Xác nhận đổ dữ liệu và thực hiện quá trình đổ dữ liệu

A screenshot of a computer

Description automatically generated

- **Bước 3:** Tạo các bảng chiều và bảng sự kiện

+ **Bước 3.1:** Tạo bảng **DimGame**

**A screenshot of a computer code

Description automatically generated**

+ **Bước 3.2:** Tạo bảng **DimGameType**

**A close-up of a number

Description automatically generated**

+ **Bước 3.3:** Tạo bảng **DimTime**

**A computer screen shot of a computer code

Description automatically generated**

+ **Bước 3.4:** Tạo bảng **DimTeam**

**A close-up of a number

Description automatically generated**

+ **Bước 3.5:** Tạo bảng **FactGameStats**

**A screenshot of a computer program

Description automatically generated**

## 2.2 Tạo mới dự án SSIS

- **Bước 1:** Mở **Visual Studio**, chọn **Create new project**

A screenshot of a computer

Description automatically generated

**- Bước 2:** Tìm từ khóa integration, sau đó chọn template **Integration Services Project**, sau đó chọn **Next**

A screenshot of a computer

Description automatically generated

- **Bước 3:** Đặt tên và chọn location cho dự án, sau đó chọn **Create**

A screenshot of a computer

Description automatically generated

- **Bước 4**: Thực hiện kết nối với các cơ sở dữ liệu đã tạo

+ **Bước 4.1:** Nhấn chuột phải vào mục **Connection Managers**, chọn **New Connection Manager**

A screenshot of a computer

Description automatically generated

**+ Bước 4.2:** Cửa sổ **Add SSIS Connection Manager** hiện lên, lựa chọn **ADO.NET**, sau đó chọn **Add**

**A screenshot of a computer

Description automatically generated**

**+ Bước 4.3:** Ở cửa sổ **Configue ADO.NET Connection Manager**, chọn **New** để thêm 1 connection mới

A screenshot of a computer

Description automatically generated

+ **Bước 4.4:** Ở cửa sổ **Connection Manager**, chọn **Server name** là **localhost**, nhập database là **NBAStats**

**A computer screen shot of a computer

Description automatically generated**

**+ Bước 4.5:** Chọn **Test Connection** để kiểm tra kết nối, kết nối thành công sẽ được kết quả như sau

A screenshot of a computer

Description automatically generated

**+ Bước 4.6:** Sau khi kiểm tra kết nối, trở lại cửa sổ **Configue ADO.NET Connection Manager** và được kết quả như sau

A screenshot of a computer

Description automatically generated

+ **Bước 4.7:** Thực hiện tương tự với database **NBAStatsDW**

## 2.3 Mô hình SSIS

**A screenshot of a computer

Description automatically generated**

- Mô hình bao gồm các **Sequence** để chứa các **Task** và các đường kết nối để sắp xếp thứ tự thực hiện của các **Sequence**

## 2.4 Sequences

### 2.4.1 Sequence Reset Database

- Sequence này chứa 1 **Execute SQL Task**, dùng để chạy các câu lệnh SQL để reset database chứa kho dữ liệu, ở đây là database **NBAStatsDW**. Sau khi reset database thì tạo lại các bảng chiều.

- **Bước 1:** Tại **SSIS ToolBox**, chọn **Execute SQL Task**

**A screenshot of a computer menu

Description automatically generated**

- **Bước 2:** Nhấn chuột phải chọn **Edit**

**A screenshot of a computer

Description automatically generated**

- **Bước 3:** Cửa sổ **Execute SQL Task Editor** hiện lên, sửa **Name** thành **Reset Database,** chọn **ConnectionType** là **ADO.NET** và **Connection** là **localhost.NBAStatsDW**

A screenshot of a computer

Description automatically generated

- **Bước 4:** Chọn vào dòng **SQLStatement** để dán các câu lệnh SQL để reset database

A screenshot of a computer

Description automatically generated

- **Bước 5:** Copy đoạn mã như hình và dán vào, đoạn mã bao gồm việc xóa database sau đó tạo lại cũng như thực hiện tạo lại các bảng chiều. Trong hình là đoạn mã thực hiện xóa database và bảng **DimGame** (các bảng chiều khác thực hiện tương tự)

A screenshot of a computer code

Description automatically generated

### 2.4.2 Sequence Load Dimension Tables

- Sequence dùng để load các bảng chiều cho kho dữ liệu từ dữ liệu gốc

#### 2.4.2.1 Load DimTeam

**A screenshot of a computer flowchart

Description automatically generated**

- **DimTeam** là bảng chiều chứa dữ liệu về các đội bóng tại giải đấu NBA

- **Bước 1:** Tạo **ADO.NET Source** để lấy dữ liệu gốc ở bảng **gameStatistics** trong database **NBAStats**

A screenshot of a computer

Description automatically generated

**- Bước 2:** Tạo **Aggregate** nhằm gom nhóm theo 2 thuộc tính là **teamId** và **teamName** để loại bỏ dữ liệu trùng

A screenshot of a computer

Description automatically generated

**- Bước 3:** Tạo **ADO.NET Destination** để lưu dữ liệu vừa được gom nhóm và lưu vào bảng **DimTeam** trong cơ sở dữ liệu **NBAStatsDW**

A screenshot of a computer

Description automatically generated

#### 2.4.2.2 Load DimGameType

**A diagram of a data processing process

Description automatically generated**

- **DimGameType** là bảng chiều chứa dữ liệu về loại trận đấu

- **Bước 1:** Tạo **ADO.NET Source** để lấy dữ liệu gốc ở bảng **gameStatistics** trong database **NBAStats**

A screenshot of a computer

Description automatically generated

**- Bước 2:** Tạo **Aggregate** nhằm gom nhóm theo 2 thuộc tính là **type** để loại bỏ dữ liệu trùng

A screenshot of a computer

Description automatically generated

**- Bước 3:** Tạo **ADO.NET Destination** để lưu dữ liệu vừa được gom nhóm và lưu vào bảng **DimGameType** trong cơ sở dữ liệu **NBAStatsDW**

A screenshot of a computer

Description automatically generated

#### 2.4.2.3 Load DimTime

**A diagram of data flow

Description automatically generated**

- **DimTime** là bảng chiều chứa dữ liệu về ngày, tháng, năm thi đấu của 1 trận đấu nào đó

- **Bước 1:** Tạo **ADO.NET Source** để lấy dữ liệu gốc ở bảng **gameStatistics** trong database **NBAStats**

A screenshot of a computer

Description automatically generated

- **Bước 2:** Tạo thêm **ADO.NET Source** tương tự **Bước 1**

A screenshot of a computer

Description automatically generated

- **Bước 3:** Đối với nguồn dữ liệu đầu tiên, sử dụng **Derived Column** thực hiện tách ngày tháng năm tại thuộc tính **date**, ở đây cần lựa chọn thuộc tính **gameId** để phục vụ cho việc kết bảng ở cuối

A screenshot of a computer

Description automatically generated

- **Bước 4:** Ở nguồn dữ liệu còn lại, sử dụng **Aggregate** để thực hiện gom nhóm theo thuộc tính **gameId**

A screenshot of a computer

Description automatically generated

- **Bước 5:** Ở phần dữ liệu dùng để tách ngày tháng năm, tạo 1 **Sort** để thực hiện sắp xếp theo thứ tự tăng dần theo thuộc tính **gameId** (bởi vì sau bước này sẽ là phép kết bảng yêu cầu dữ liệu cần được sắp xếp theo thứ tự cụ thể), tick chọn **Remove rows with duplicate sort values** để xóa các giá trị thuộc tính bị lặp khi sắp xếp

A screenshot of a computer

Description automatically generated

- **Bước 6:** Thực hiện tương tự **Bước 5** đối với nguồn dữ liệu còn lại

A screenshot of a computer

Description automatically generated

- **Bước 7:** Tạo **Merge Join** để thực hiện kết 2 nguồn dữ liệu sau khi đã sắp xếp, lựa chọn các thuộc tính cần có để đổ vào bảng **DimTime**

A screenshot of a computer

Description automatically generated

- **Bước 8:** Sử dụng đối tượng **ADO.NET Destination** để lưu dữ liệu vừa được kết và lưu vào bảng **DimTime** trong cơ sở dữ liệu **NBAStatsDW**

A screenshot of a computer

Description automatically generated

#### 2.4.2.4 Load DimGame

A screenshot of a computer

Description automatically generated

- **DimGame** là bảng chiều chứa dữ liệu về trận đấu, gồm tên đội và điểm của 2 đội, thời gian và thuộc mùa bóng nào

- **Bước 1:** Tạo **ADO.NET Source** để lấy dữ liệu gốc ở bảng **gameStatistics** trong database **NBAStats**

A screenshot of a computer

Description automatically generated

- **Bước 2:** Tạo **Conditional Split** để tách các dòng dữ liệu với điều kiện cho trước, bởi vì 2 dòng liên tiếp trong dữ liệu gốc là cùng thuộc về 1 trận đấu và 2 dòng này có thể là dữ liệu của **home** thì dòng còn lại sẽ của **away** và ngược lại. Để thực hiện tách, sử dụng thuộc tính **team**

A screenshot of a computer

Description automatically generated

- **Bước 3:** Tạo **Sort** ở cả 2 nhánh điều kiện, thực hiện sắp xếp thứ tự tăng dần theo thuộc tính **gameId**, tick chọn **Remove rows with duplicate sort values** để xóa các thuộc tính bị lặp khi sắp xếp

A screenshot of a computer

Description automatically generated

- **Bước 4:** Tạo **Merge Join** để thực hiện kết 2 nguồn dữ liệu sau khi đã sắp xếp, lựa chọn các thuộc tính cần có để đổ vào bảng **DimGame**

- **Bước 5:** Sử dụng **ADO.NET Destination** để lưu dữ liệu vừa được kết và lưu vào bảng **DimGame** trong cơ sở dữ liệu **NBAStatsDW**

### 2.4.3 Sequence Create Foreign Key

- Sequence này chứa 1 **Execute SQL Task**, dùng để chạy các câu lệnh SQL để tạo các khóa ngoại từ bảng sự kiện đến các bảng chiều

- **Bước 1:** Tại **SSIS ToolBox**, chọn **Execute SQL Task**

**A screenshot of a computer menu

Description automatically generated**

- **Bước 2:** Nhấn chuột phải chọn **Edit**

**A screenshot of a computer

Description automatically generated**

- **Bước 3:** Cửa sổ **Execute SQL Task Editor** hiện lên, sửa **Name** thành **Create Foreign Key,** chọn **ConnectionType** là **ADO.NET** và **Connection** là **localhost.NBAStatsDW**

A screenshot of a computer

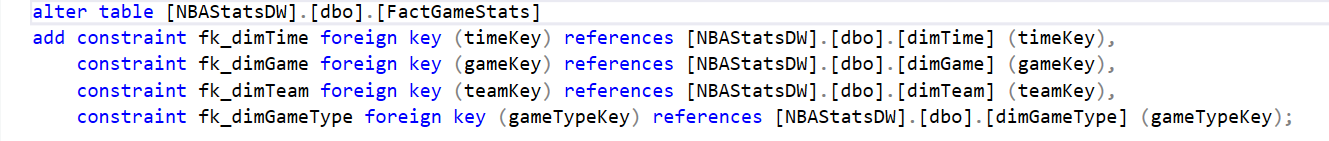
Description automatically generated

- **Bước 4:** Chọn vào dòng **SQLStatement** để dán các câu lệnh SQL để reset database

A screenshot of a computer

Description automatically generated

- **Bước 5:** Copy đoạn mã như hình và dán vào, đoạn mã thực hiện tạo khóa ngoại cho bảng sự kiện tới các bảng chiều



### 2.4.4 Sequence Load Fact Table

A screenshot of a computer

Description automatically generated

- **FactGameStats** là bảng sự kiện chứa các thông số thống kê liên quan đến trận đấu

- **Bước 1:** Tạo các **ADO.NET Source** để lấy dữ liệu từ dữ liệu gốc và các bảng chiều

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

- **Bước 2:** Tạo các **Sort** để sắp xếp các nguồn dữ liệu để sử dụng cho kết dữ liệu

+ **Bước 2.1:** Sắp xếp dữ liệu gốc theo các thuộc tính **gameId, type, teamId**

A screenshot of a computer

Description automatically generated

+ **Bước 2.2:** Sắp xếp bảng **DimGameType** theo thuộc tính **typeName**

A screenshot of a computer

Description automatically generated

+ **Bước 2.3:** Sắp xếp bảng **DimGame** theo thuộc tính **gameId**

A screenshot of a computer

Description automatically generated

+ **Bước 2.4:** Sắp xếp bảng **DimTeam** theo thuộc tính **teamId**

A screenshot of a computer

Description automatically generated

+ **Bước 2.5:** Sắp xếp bảng **DimTime** theo thuộc tính **gameId**

A screenshot of a computer

Description automatically generated

- **Bước 3:** Tạo **Merge Join** để kết dữ liệu gốc và bảng **DimGameType** thông qua thuộc tính **type**, sau đó thực hiện chọn các thuộc tính để phù hợp với bảng sự kiện

A screenshot of a computer

Description automatically generated

- **Bước 4:** Tạo **Sort** để sắp xếp dữ liệu vừa kết ở **Bước 3** theo thuộc tính **gameId**

A screenshot of a computer

Description automatically generated

- **Bước 5:** Tạo **Merge Join** để kết dữ liệu vừa sắp xếp ở **Bước 4** với bảng **DimGame** thông qua thuộc tính **gameId**

A screenshot of a computer

Description automatically generated

- **Bước 6:** Tạo **Sort** để sắp xếp dữ liệu vừa kết ở **Bước 5** theo thuộc tính **teamId**

A screenshot of a computer

Description automatically generated

- **Bước 7:** Tạo **Merge Join** để kết dữ liệu vừa sắp xếp ở **Bước 6** với bảng **DimTeam** thông qua thuộc tính **teamId**

A screenshot of a computer

Description automatically generated

- **Bước 8:** Tạo **Sort** để sắp xếp dữ liệu vừa kết ở **Bước 7** theo thuộc tính **gameId**

A screenshot of a computer

Description automatically generated

- **Bước 9:** Tạo **Merge Join** để kết dữ liệu vừa sắp xếp ở **Bước 8** với bảng **DimTime** thông qua thuộc tính **gameId**

A screenshot of a computer

Description automatically generated

- **Bước 10:** Tạo **ADO.NET Destination** để lưu dữ liệu vừa kết vào bảng sự kiện, chọn **Mappings** để map các thuộc tính sao cho phù hợp

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## 2.5 Chạy dự án và kiểm tra

### 2.5.1 Chạy trên Visual Studio

**A screenshot of a computer

Description automatically generated**

### 2.5.2 Kiểm tra dữ liệu trên Microsoft SQL Server Management Studio

#### 2.5.2.1 Bảng DimGame

A screenshot of a computer

Description automatically generated

#### 2.5.2.2 Bảng DimGameType

A computer screen shot of a computer

Description automatically generated

#### 2.5.2.3 Bảng DimTeam

A screenshot of a computer

Description automatically generated

#### 2.5.2.4 Bảng DimTime

A screenshot of a computer

Description automatically generated

#### 2.5.2.5 Bảng FactGameStats

A screenshot of a computer

Description automatically generated

Link tài liệu: <https://drive.google.com/drive/folders/1wSs9B0sCg5RHBhZ3SrNwfSDCymvcqy9p>