**Chuẩn hoá dữ liệu là gì?**

Chuẩn hóa là một kỹ thuật thiết kế cơ sở dữ liệu tổ chức các bảng theo cách làm giảm sự dư thừa và phụ thuộc của dữ liệu.

Nó chia các bảng lớn hơn thành các bảng nhỏ hơn và liên kết chúng bằng các mối quan hệ.

Trong hướng dẫn này, bạn sẽ học-

* Cơ sở dữ liệu biểu mẫu bình thường
* Quy tắc 1NF
* Chìa khóa là gì?
* Khóa tổng hợp là gì
* Quy tắc 2NF
* Cơ sở dữ liệu - Khóa ngoại
* Phụ thuộc chức năng bắc cầu là gì?
* Quy tắc 3NF
* Mẫu thường Boyce-Codd (BCNF)

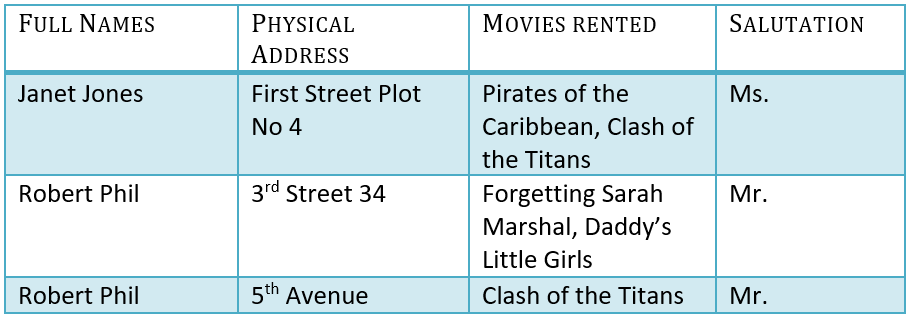
Người phát minh ra mô hình quan hệ Edgar Codd đã đề xuất lý thuyết bình thường hóa với sự ra đời của Mẫu bình thường thứ nhất, và ông tiếp tục mở rộng lý thuyết với Mẫu bình thường thứ hai và thứ ba. Sau đó, ông đã tham gia với Raymond F. Boyce để phát triển lý thuyết về hình thức bình thường Boyce-Codd.

Lý thuyết chuẩn hóa dữ liệu trong SQL vẫn đang được phát triển thêm. Ví dụ, có các cuộc thảo luận ngay cả trên Mẫu thường thứ 6 . **Tuy nhiên, trong hầu hết các ứng dụng thực tế, chuẩn hóa đạt được kết quả tốt nhất ở  Dạng thường thứ 3** . Sự phát triển của các lý thuyết Bình thường hóa được minh họa dưới đây-

[Bình thường hóa là gì?  1NF, 2NF, 3NF & BCNF với các ví dụ](https://www.guru99.com/images/NormalizationProcess(1).png)

**Ví dụ chuẩn hóa cơ sở dữ liệu -**

Giả sử một thư viện video duy trì một cơ sở dữ liệu của các bộ phim được thuê. Không có bất kỳ chuẩn hóa nào, tất cả thông tin được lưu trữ trong một bảng như dưới đây.

Bảng 1

Ở đây bạn thấy **cột Phim được thuê có nhiều giá trị** .

**Cơ sở dữ liệu biểu mẫu bình thường**

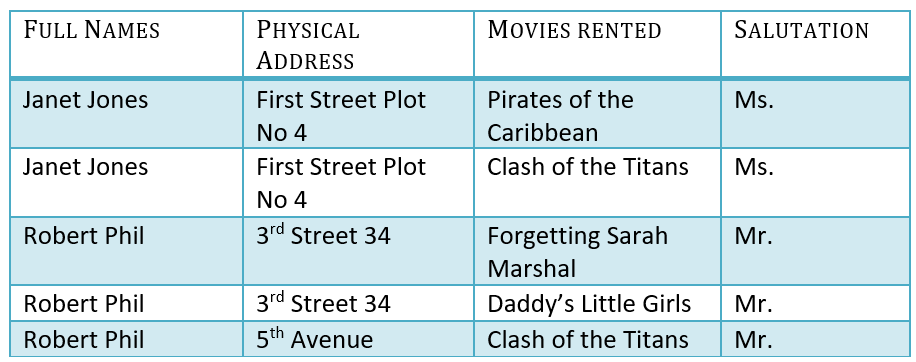
Bây giờ hãy chuyển sang dạng thứ 1

**Quy tắc 1NF (Mẫu thường đầu tiên)**

* Mỗi ô bảng nên chứa một giá trị duy nhất.
* Mỗi hồ sơ cần phải là duy nhất.

Bảng trên trong 1NF-

**Ví dụ 1NF**

Bảng 1: Trong mẫu 1NF

Trước khi chúng ta tiến hành, hãy hiểu một vài điều -

**Chìa khóa là gì?**

Một khóa là một giá trị được sử dụng để xác định một bản ghi trong một bảng duy nhất. Một KEY có thể là một cột đơn hoặc kết hợp nhiều cột

Lưu ý: Các cột trong bảng KHÔNG được sử dụng để xác định một bản ghi duy nhất được gọi là các cột không khóa.

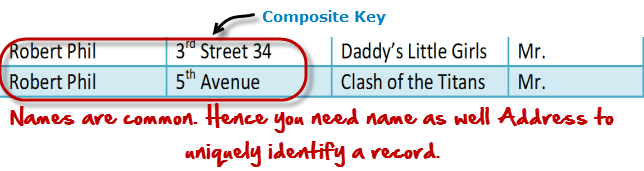
Khóa chính là gì?

|  |  |
| --- | --- |
| [Bình thường hóa là gì?  1NF, 2NF, 3NF & BCNF với các ví dụ](https://www.guru99.com/images/PrimaryKey.png) | Chính là một giá trị cột duy nhất được sử dụng để xác định một bản ghi cơ sở dữ liệu duy nhất.  Nó có các thuộc tính sau   * Khóa chính không thể là NULL * Giá trị khóa chính phải là duy nhất * Các giá trị khóa chính hiếm khi được thay đổi * Khóa chính phải được cung cấp một giá trị khi một bản ghi mới được chèn. |

**Khóa tổng hợp là gì?**

Khóa tổng hợp là khóa chính gồm nhiều cột được sử dụng để xác định một bản ghi duy nhất

Trong cơ sở dữ liệu của chúng tôi, chúng tôi có hai người có cùng tên Robert Phil, nhưng họ sống ở những nơi khác nhau.



Do đó, chúng tôi yêu cầu cả Họ và Tên để xác định một bản ghi duy nhất. Đó là một khóa tổng hợp.

Hãy chuyển sang dạng 2NF bình thường thứ hai

**Quy tắc 2NF (Mẫu thường thứ hai)**

* Quy tắc 1- Trở thành 1NF
* Quy tắc 2- Khóa chính cột đơn

Rõ ràng là chúng ta không thể tiến lên để tạo cơ sở dữ liệu đơn giản của mình ở dạng Chuẩn hóa thứ 2 trừ khi chúng ta phân vùng bảng ở trên.

[](https://www.guru99.com/images/Table2.png)

Bảng 1

[](https://www.guru99.com/images/Table1.png)

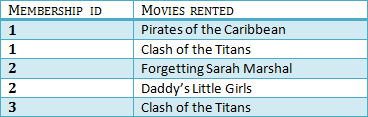
Bảng 2

Chúng tôi đã chia bảng 1NF của chúng tôi thành hai bảng viz. Bảng 1 và Bảng2. Bảng 1 chứa thông tin thành viên. Bảng 2 chứa thông tin về các bộ phim được thuê.

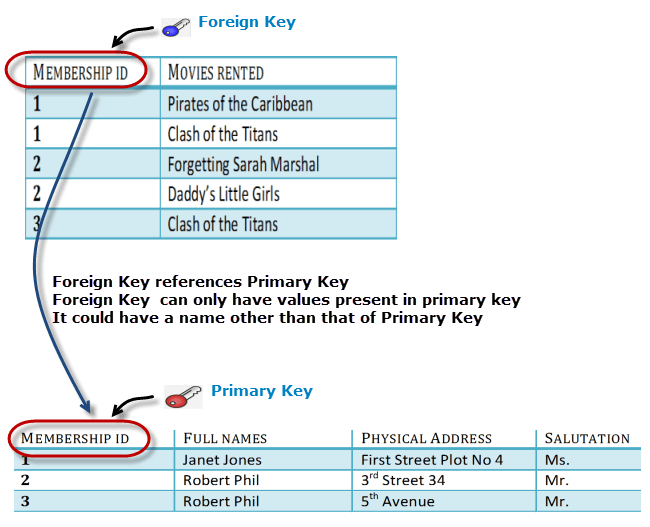
Chúng tôi đã giới thiệu một cột mới có tên là Memberhip\_id, đây là khóa chính cho bảng 1. Các bản ghi có thể được xác định duy nhất trong Bảng 1 bằng cách sử dụng id thành viên

**Cơ sở dữ liệu - Khóa ngoại**

Trong Bảng 2, Memberhip\_ID là Khóa ngoài

[](https://www.guru99.com/images/foreign_key_table.png)

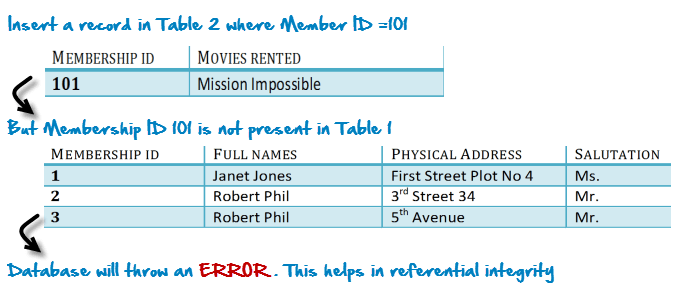
|  |  |
| --- | --- |
| [Bình thường hóa là gì?  1NF, 2NF, 3NF & BCNF với các ví dụ](https://www.guru99.com/images/ForeignKey.png) | Khóa ngoài tham chiếu khóa chính của Bảng khác! Nó giúp kết nối các Bảng của bạn   * Khóa ngoại có thể có một tên khác với khóa chính của nó * Nó đảm bảo các hàng trong một bảng có các hàng tương ứng trong một bảng khác * Không giống như khóa Chính, chúng không phải là duy nhất. Thông thường họ không * Khóa ngoại có thể là null ngay cả khi khóa chính không thể |

[](https://www.guru99.com/images/ForeignKeyRelationWithPrimary.png)

**Tại sao bạn cần một khóa ngoại?**

Giả sử, một người mới chèn một bản ghi vào Bảng B, chẳng hạn như

Bạn sẽ chỉ có thể chèn các giá trị vào khóa ngoại tồn tại trong khóa duy nhất trong bảng cha. Điều này giúp trong tính toàn vẹn tham chiếu.

[](https://www.guru99.com/images/WhyDataBaseIsImportant.png)

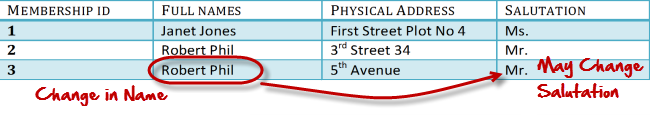
Vấn đề trên có thể được khắc phục bằng cách khai báo id thành viên từ Bảng 2 là khóa ngoại của id thành viên từ Bảng1

Bây giờ, nếu ai đó cố gắng chèn một giá trị trong trường id thành viên không tồn tại trong bảng cha, một lỗi sẽ được hiển thị!

**Phụ thuộc chức năng bắc cầu là gì?**

Một phụ thuộc chức năng bắc cầu là khi thay đổi cột không khóa, có thể khiến bất kỳ cột không khóa nào khác thay đổi

Xem xét bảng 1. Thay đổi cột không khóa Tên đầy đủ có thể thay đổi Tính toán.

[](https://www.guru99.com/images/transitive_functional_dependencies.png)

Hãy chuyển sang 3NF

**Quy tắc 3NF (Mẫu thường thứ ba)**

* Quy tắc 1- Trở thành 2NF
* Quy tắc 2- Không có phụ thuộc chức năng bắc cầu

Để di chuyển bảng 2NF của chúng tôi thành 3NF, một lần nữa chúng ta cần chia lại bảng của mình.

**Ví dụ 3NF**

[](https://www.guru99.com/images/2NFTable1.png)

BẢNG 1

[](https://www.guru99.com/images/2NFTable2.png)

Bảng 2

[](https://www.guru99.com/images/2NFTable3.png)

Bảng 3

Chúng tôi đã một lần nữa phân chia các bảng của chúng tôi và tạo ra một bảng mới lưu trữ Salutations.

Không có phụ thuộc chức năng bắc cầu và do đó bảng của chúng tôi nằm trong 3NF

Trong Bảng 3 ID tính toán là khóa chính và trong Bảng 1 ID tính toán là ngoại lệ đối với khóa chính trong Bảng 3

Bây giờ ví dụ nhỏ của chúng tôi đang ở mức không thể phân tách hơn nữa để đạt được các hình thức chuẩn hóa cao hơn. Trong thực tế, nó đã ở dạng bình thường hóa cao hơn. Những nỗ lực riêng biệt để chuyển sang các cấp độ chuẩn hóa dữ liệu tiếp theo thường là cần thiết trong các cơ sở dữ liệu phức tạp. Tuy nhiên, chúng tôi sẽ thảo luận về các mức độ chuẩn hóa tiếp theo một cách ngắn gọn trong phần sau.

**Mẫu thường Boyce-Codd (BCNF)**

Ngay cả khi cơ sở dữ liệu ở dạng bình thường thứ 3 , vẫn sẽ có kết quả bất thường nếu nó có nhiều hơn một Khóa **ứng viên** .

Đôi khi BCNF cũng được gọi là **3,5 Mẫu thường.**

**Quy tắc 4NF (Mẫu thường thứ tư)**

Nếu không có cá thể bảng cơ sở dữ liệu nào chứa hai hoặc nhiều dữ liệu độc lập và đa trị mô tả thực thể có liên quan, thì đó là ở dạng thông thường thứ 4 .

**Quy tắc 5NF (Mẫu thường thứ năm)**

Một bảng ở dạng bình thường thứ 5 chỉ khi nó ở 4NF và nó không thể được phân tách thành bất kỳ số lượng bảng nhỏ hơn nào mà không mất dữ liệu.

**Đề xuất 6NF (Mẫu thường thứ sáu)**

Mẫu thường thứ 6 không được chuẩn hóa, tuy nhiên, nó đang được các chuyên gia cơ sở dữ liệu thảo luận một thời gian. Hy vọng rằng, chúng ta sẽ có một định nghĩa rõ ràng và được chuẩn hóa cho Mẫu thường thứ 6 trong tương lai gần ...

Đó là tất cả để bình thường hóa !!!

**Tóm lược**

* Thiết kế cơ sở dữ liệu là rất quan trọng để thực hiện thành công một hệ thống quản lý cơ sở dữ liệu đáp ứng các yêu cầu dữ liệu của hệ thống doanh nghiệp.
* Bình thường hóa giúp hệ thống cơ sở dữ liệu sản phẩm có hiệu quả chi phí và có các mô hình bảo mật tốt hơn.
* phụ thuộc hàm là một thành phần rất quan trọng của việc xử lý dữ bình thường hóa
* Hầu hết các hệ cơ sở dữ liệu được chuẩn hóa cơ sở dữ liệu lên đến các hình thức bình thường thứ ba.
* Khóa chính xác định duy nhất là bản ghi trong Bảng và không thể rỗng
* Khóa ngoại giúp kết nối bảng và tham chiếu khóa chính