**Chuẩn hóa cơ sở dữ liệu với các trường hợp cụ thể trong thực tế**

December 29, 2015 by [Quan Pham](http://phamquan.com/author/quanltv) [Leave a Comment](http://phamquan.com/co-so-du-lieu/thiet-ke-csdl/chuan-hoa-co-so-du-lieu-voi-cac-truong-hop-cu-the-trong-thuc-te.html#respond)

Chào các bạn, trong bài trước mình đã giới thiệu tới các bạn bắt đầu với cơ sở dữ liệu quan hệ, hôm nay chúng ta sẽ tiếp tục tìm hiểu về chuẩn hóa cơ sở dữ liệu quan hệ.

**Chuẩn hóa cơ sở dữ liệu là gì? Tại sao cần chuẩn hóa?**

**Chuẩn hóa** là kỹ thuật thiết kế cơ sở dữ liệu tổ chức các bảng theeo kiểu làm giảm bớt sự dư thừa và phạu thuộc của dữ liệu.

Là việc chia bảng lớn thành các bảng nhỏ và liên kết chúng bằng các quan hệ.

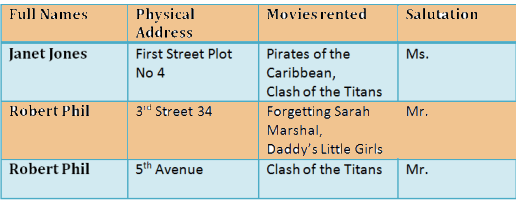
Người phát minh ra mô hìn dữ liệu quan hệ là **Edgar Codd** mục đích của lý thuyết chuẩn hóa là đưa về dang chuẩn 1 (First **Normal Form**) và anh ta tiếp tục mở rộng lý thuyết chuẩn hóa với dạng chuẩn 2 và dạng chuẩn 3 (**Second and** **Third Normal Form).** Sau đó anh ta tham gia cùng với **Raymond F. Boyce**để phát triển lý thuyết chuẩn hóa dạng **Boyce-Codd Normal Form**.

Lý thuyết chuẩn hóa vẫn đang được phát triển. Sự phát triển của lý thuyết chuẩn hóa được giải thích theo hình sau:

alt

Nào, chúng ta sẽ bắt đầu học chuẩn hóa cơ sở dữ liệu với các ví dụ cụ thể

Gỉa sử một thư viện video trữ cơ sở dữ liệu về phim. Không cần chuẩn hóa, tất cả các thông tin được lưu trữ tại 1 bảng duy nhất dưới đây.



Bảng 1

Bạn có thể thấy cột **Movies  Rented**  có nhiều giá trị.

Nào, bây giờ chúng ta sẽ chuyển sang dạng chuẩn 1.

**Các luật của dang chuẩn 1NF**

* Mỗi hàng trong bảng có một giá trị.
* Mỗi bản ghi là duy nhất.

Bảng sau là dạng chuẩn 1NF

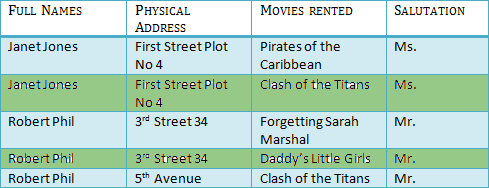
[](http://phamquan.com/wp-content/uploads/2015/12/1NF.png)

Table 1 : Dạng chuẩn 1NF.

Trước khi phân tích làm như thế nào, chúng ta cần phải hiểu một số thứ trước đã.

**Khóa là gì?**

Khóa là giá trị được sử dụng để xác định duy nhất một bản ghi trong bảng. Một khóa có thể là 1 cột hay kết nối nhiều cột với nhau.

Do đó  cột trong bảng mà không sử dụng để xác định duy nhất một bản ghi thì đó không phải là khóa.

**Khóa chính là gì?**

|  |  |
| --- | --- |
| http://cdn.guru99.com/images/PrimaryKey.png | Khóa chính là một cột giá trị sử dụng để xác định duy nhất bản ghi trong cơ sở dữ liệu quan hệ.  Nó có các thuộc tính sau:   * Khóa chính không thể NULL * Gitrị phải là duy nhất. * Không thể thay đổi giá trị. * Phải có giá trị khi bản ghi được chèn thêm. |

**Khóa phức là gì?**

Khoa phức là khóa chính gồm nhiều cột kết nối vói nhau để xác định một bản ghi duy nhất.

Trong cơ sở dữ liệu của chúng ta, Chúng ta có 2 người cúng tên Robert Phil nhưng họ sống tại địa chỉ khác nhau.

[](http://phamquan.com/wp-content/uploads/2016/01/khoaphuc.png)

Do đó cần phải dùng cả tên và địa chỉ để xác định bản ghi. Đó là khóa phức. Rất đơn giản đúng không??

**Dạng chuẩn 2NF**

2NF Rules

* Rule 1- là dạng chuẩn 1NF
* Rule 2- Một cột là khóa chính.

Chúng ta không thể chuyển thẳng sang dạng chuẩn 2NF mà phải chia bảng ra như sau:

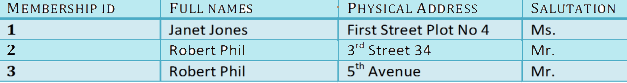
[](http://phamquan.com/wp-content/uploads/2015/12/Table2.png)

Table 1

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

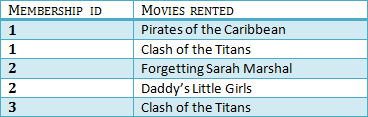
Bảng 2

Chúng ta chia bảng ở dạng chuẩn 1NF thành 2 bảng. Bảng 1 chứa thông tin thành viên. Bản 2 chứa thông tin phim.

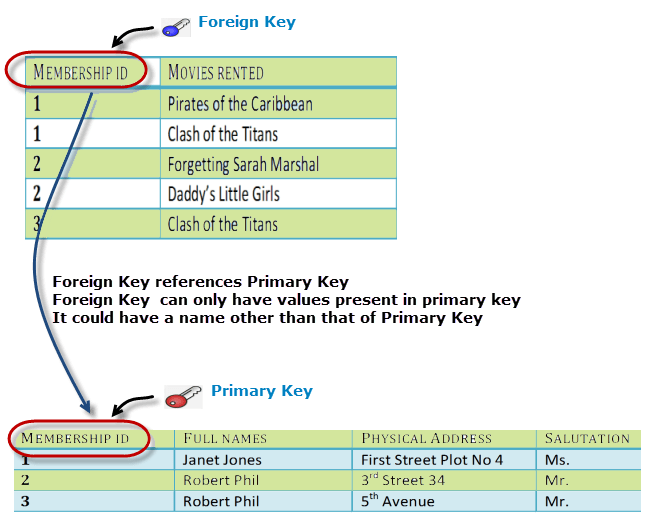
Chúng ta se thảo luật về cột mới có tên Membership\_id tại sao là khóa chính cho bảng 1. Bản ghi sẽ được xác định duy nhất trong bảng 1 sử dụng Membership ID

**Khóa ngoại (foreignkey) là gì?**

Trong bảng 2, Membership\_ID là khóa ngoại (foreign key).



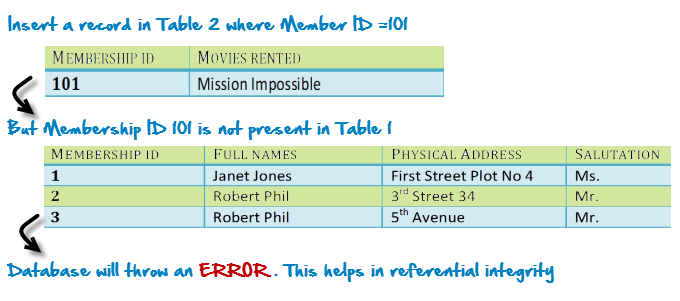
|  |  |
| --- | --- |
| alt | Khóa ngoại tham chiếu tới bảng khác. Nó giúp liên kết các bảng của bạn.   * Khóa ngoại có thể có tên khác nhau từ kháo chính của nó. * Nó chắc chắn rằng hàng trong bảng phù hợp với hàng khác trong bảng khác * Khác với khóa chính, khóa ngoại không có tính duy nhất trong bảng * Khóa ngoại có thể NULL nhưng khóa chính thì không thể. |

[](http://phamquan.com/wp-content/uploads/2016/01/khoangoai.png)

Tại sao bạn cần khóa ngoại ?

Giả sử một thằng ngốc nào đó (chắc nó chừa mình ra) chén một bản ghi vào bảng B như sau

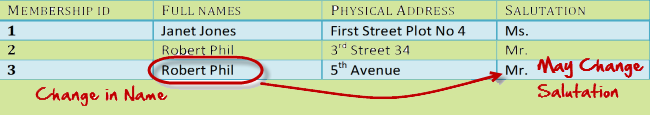
Bạn chỉ có thể chèn giá trị vào khóa ngoại của bạn mà nó tồn tại trong khóa duy nhất trong bảng cha. Nó sẽ giúp cho sự đồng nhất về dữ liệu.

[](http://phamquan.com/wp-content/uploads/2016/01/khoangoai2.png)

**Sự phụ thuộc hàm là gì?**

Sự phụ thuộc hàm là khi chúng ta thay đổi trường (cột) không phải khóa có thể ảnh hưởng tới cột khác.

Việc thay đổi tên có thể dẫn tới thay đổi giới tính?

[](http://phamquan.com/wp-content/uploads/2016/01/suphuthuocham.png)

Nào, chúng ta sẽ bắt đầu chuyển qua dạng chuẩn 3NF. Trước hết bạn cần biết được các luật hay quy ước của nó trước khi tiến hành chuyển đổi.

**Luật của dạng chuẩn 3NF**

* Luật 1- là dạng chuẩn 2NF
* Luật 2- không có sự phụ thuộc hàm.

Để chuyển dạng chuẩn 2NF sang 3NF chúng ta cân phải chia bảng ra.



Bảng 1.



Bảng 2.



Bảng 3

Chúng ta phải chia bảng của chúng ta ra và tạo bảng mới để lưu trữ Salutations.

Do không có một sự phụ thuộc nào nên bảng của chúng ta thuộc dạng chuẩn 3NF.

Trong bảng 3 Salutation ID là khóa chính (primary key) và trong bảng 1,  Salutation ID là khóa ngoại tham chiếu tới khóa chính trong bảng 3.

Ví dụ này có thể không phân tách để đạt tới cấp độ chuẩn hóa cao. Trong thực tế, nó luôn ở dạng chuẩn ở mức độ cao. Sự nỗ lực chia ra để đạt tới cấp độ cao hơn của sự chuẩn hóa là cần thiết trong các cơ sở dữ liệu phức tạp.  Tuy nhiên chúng ta cần thảo luận và mức độ chuẩ tiếp theo một cách nhanh chóng như sau.

Dạng chuẩn Boyce-Codd hay Boyce-Codd Normal Form (BCNF)

Mặc dù khi dữ liệu ở dạng chuẩn 3 , nó vẫn có kết quả khác thướng hơn một khóa Candidate.

Một số BCNF chuyển đến như dang chuẩn **3.5 Normal Form.**

**Dạng chuẩn 4**

Nếu không có bảng nào bao gồm 2 hoặc nhiều hơn độc lập và đa giá trị thì nó là dạng chuẩn 4.

**Dạng chuẩn 5**

Một bảng là dạng chuẩn 5 nếu nó là dạng chuẩn 4 và không thể tách thành cách bảng nhỏ hơn mà không bị đánh mất dữ liệu.

**Dạng chuẩn 6**

Không phải là dạng chuẩn tuy nhiên nó vẫn thỉnh thoảng được thảo luận bởi các chuyên gia về dữ liệu. Hi vọng chúng ta sẽ có dạng chuẩn 6 trong tương lai gần