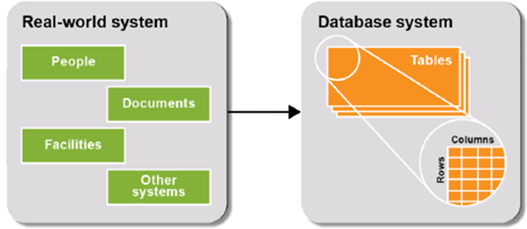
**Thiết kế cấu trúc cơ sở dữ liệu**

Thiết kế cấu trúc cơ sở dữ liệu là quá trình mô hình hóa nhằm chuyển đổi các đối tượng từ thế giới thực (Real-world system) sang các bảng trong hệ thống cơ sở dữ liệu (Database system) đáp ứng các yêu cầu lưu trữ và khai thác dữ liệu.



**Mô hình hóa việc thiết kế cấu trúc cơ sở dữ liệu**

Trong đó,

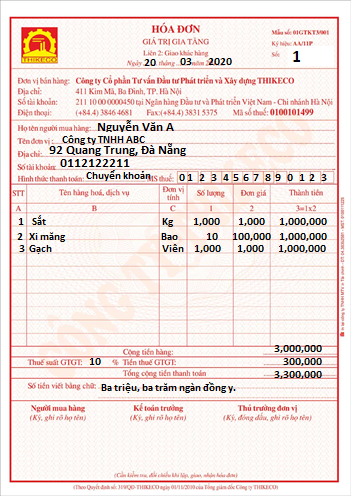
* **People (con người)**: những người tham gia vào hệ thống, bạn cần làm việc với những người này để xác định các dữ liệu cần lưu trữ, cần khai thác.
* **Documents (tài liệu)**: bạn cần khảo sát các tài liệu trong hệ thống để xác định dữ liệu.
* **Facilities(cơ sở vật chất)**: bạn cần quan tâm những cơ sở vật chất nào cần quản lý.
* **Other systems(hệ thống khác, hệ thống tương tự)**: bạn cần tìm hiểu nghiên cứu các hệ thống tương tự để thu thập thêm dữ liệu.

Để thực hiện việc thiết kế cơ sở dữ liệu chúng ta cần thực hiện các bước sau đây:

1. Xác định các thành phần dữ liệu
2. Chia nhỏ các thành phần dữ liệu ra thành các phần nhỏ nhất mà hệ thống sử dụng
3. Xác định các bảng và các cột
4. Xác định khóa chính, khóa ngoại và mối quan hệ
5. Kiểm tra cấu trúc cơ sở dữ liệu được thiết kế với qui định chuẩn hóa

**Thực hành thiết kế Cơ sở dữ liệu**

Để các bạn dễ hiểu quá trình thiết kế cơ sở dữ liệu, chúng ta sẽ thực hành thiết kế cấu trúc cơ sở dữ liệu cho hệ thống quản lý đơn hàng có mẫu hóa đơn như sau:



# **Bước 1: Xác định các thành phần dữ liệu**

Để xác định các thành phần dữ liệu chúng ta cần thực hiện các bước sau đây:

1. Phân tích hệ thống hiện tại
2. Đánh giá, xem xét các hệ thống tương tự
3. Phỏng vấn người dùng
4. Phân tích các tài liệu trong hệ thống hiện tại

Ở bài toán trên, chúng ta thu thập được dữ liệu quan trọng là mẫu hóa đơn của đơn vị đang áp dụng, chúng ta sẽ phân tích nó để định nghĩa các thành phần dữ liệu.

Khảo sát hóa đơn trên chúng ta thu được các thành phần dữ liệu sau đây:

InvoiceNo: Số hóa đơn

InvoiceDate: Ngày ghi hóa đơn

CustomerName: Tên khách hàng

CustomerCompany: Tên công ty khách hành

CustomerAddress: Địa chỉ khách hàng

CustomerAccountNo: Số tài khoản của khách hàng

PaymentMethod: Phương thức thanh toán

CustomerTaxNo: Mã số thuế của khách hàng

SequenceNo: Số thứ tự mua hành

ProductName: Tên sản phẩm

Unit: Đơn vị tính của sản phẩm

Quantity: Số lượng

UnitPrice: Đơn giá

Amount: Thành tiền

TotalAmount: Tổng tiền

VAT: Thuế giá trị gia tăng

TotalPay: Tổng tiền phải trả

ByText: Ghi bằng chữ

**Lưu ý**: Vì tiếng việt có dấu dễ gây lỗi khi code, nếu dùng không dấu thì dễ gây nhầm lẫn nên tôi khuyên các bạn nên đặt tên các thành phần dữ liệu theo tiếng Anh vừa tránh được các lỗi trên vừa giúp bạn dễ dàng tham gia các dự án quốc tế sau này.

**Loại bỏ các dữ liệu trùng ở các dạng sau:**

1. Hai thành phần dữ liệu nhưng trỏ đến một thành phần dữ liệu thực tế
2. Bỏ những thành phần tính toán được
3. Những trường không cần lưu trữ hoặc không có thực

Xem xét danh sách các thành phần ở trên chúng ta loại các trường sau:

~~Amount~~: Thành phần này được tính từ đơn giá \* số lượng

~~TotalAmount~~: Thành phần này được tính bằng tổng các mục thành tiền.

~~TotalPay~~: Thành phần này được tính bằng TotalAmount – VAT amount

~~ByText~~: Được đọc từ tổng tiền phải trả.

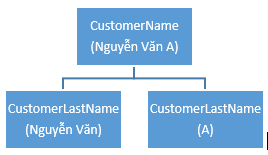
Các thành phần không cần như thông tin đơn vị bán, chữ ký người mua, chữ ký kế toán… Có nhưng thực sự bạn không cần lưu cùng đơn hàng vào cơ sở dữ liệu.

# **Bước 2: Chia nhỏ các thành phần dữ liệu thành đơn vị nhỏ nhất hữu dụng**

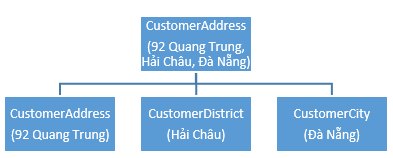
Để hiểu phần này bạn xem xét ví dụ sau:

CustomerName có giá  trị là Nguyễn Văn A, trường này có thể tách ra là Lastname (Nguyễn), Middlename (Văn) và Firstname (A). Tuy nhiên, có hệ thống thì lưu hết vào một trường là ‘Nguyễn Văn A’ như giao hàng chẳng hạn, có hệ thống chia ra là ‘Nguyễn Văn’, ‘A’ như hệ thống quản lý sinh viên, có hệ thống chia ra thành ‘Nguyễn’, ‘Văn’, ‘A’ như hệ thống quản lý bay… Do vậy, bạn cần xem xét hệ thống bạn đang xây dựng sẽ lưu như thế nào.

Trong hệ thống này do hay sắp xếp theo tên khách hàng nên chúng ta tách nó ra thành 02 phần là CustomerLastName và CustomerFirstName.



Tương tự trường CustomerAddress cung vậy, để quản lý theo tỉnh/thành phố và quận/huyện chúng ta chia nó ra thành 03 trường như sau: CustomerAddress, CustomerDistrict và CustomerCity.



Bạn hãy xem xét các trường còn lại xem có phải tách trường nào ra nữa hay không.

Sau khi tách các trường cần thiết và loại bỏ các trường không cần thiết ta có các thành phần dữ liệu như sau:

InvoiceNo: Số hóa đơn

InvoiceDate: Ngày ghi hóa đơn

**CustomerFirstName: Tên khách hàng**

**CustomerLastName: Họ và tên lót của khách hàng**

CustomerCompany: Tên công ty khách hành

**CustomerAddress: Địa chỉ khách hàng**

**CustomerCity: Thành phố khách hàng đang ở**

**CustomerDistrict: Quận khách hàng đang ở**

CustomerAccountNo: Số tài khoản của khách hàng

PaymentMethod: Phương thức thanh toán

CustomerTaxNo: Mã số thuế của khách hàng

SequenceNo: Số thứ tự mua hành

ProductName: Tên sản phẩm

Unit: Đơn vị tính của sản phẩm

Quantity: Số lượng

UnitPrice: Đơn giá

Amount: Thành tiền

VAT: Thuế giá trị gia tăng

# **Bước 3: Xác định các bảng và các cột cho cơ sở dữ liệu**

Thực hiện theo các bước sau:

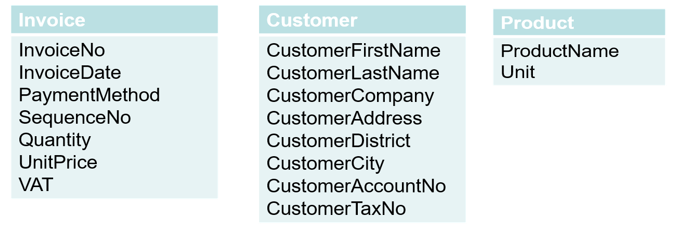
1. Nhóm các trường theo các thực thể (Entities)
2. Kiểm tra lại các trường thừa/thiếu.

Chúng ta bàn về thực thể, thực thể là người, là vật tồn tại trong hệ thống đang xem xét. Đối với thiết kế Cơ sở dữ liệu chúng ta chỉ quan tâm đến các thực thể mang thông tin. Xem xét danh sách các thành phần dữ liệu từ trên xuống chúng ta có thể liệt kê ra các thực thể thấy được như sau:

* Invoice(Hóa đơn)
* Customer (Khách hàng)
* Product (Sản phẩm)

Có thể bạn sẽ thấy khó hiểu chỗ này, bạn cần suy nghĩ kỹ để hiểu lý do vì sao chọn được 03 thực thể trên.

## Tiếp theo chúng ta nhóm các thành phần dữ liệu tương ứng vào các thực thể



## Kiểm tra các trường thừa/thiếu

Nếu có trường thừa ra, bạn cần xem xét nó có thực sự cần lưu trữ không? Nếu cần lưu trữ thì bạn cần bổ sung thực thể chứa thuộc tính này. Nếu không cần lưu trữ bạn cần loại bỏ nó đi.

Bạn cần kiểm tra từng thực thể xem có cần bổ sung thuộc tính nào không? Nếu cần thì bạn thêm vào.

# **Bước 4: Xác định khóa chính, khóa ngoại và mối quan hệ giữa các thực thể**

1. Xác định khóa chính cho các thực thể
2. Xác định quan hệ giữa các thực thể
3. Phân tách để đưa về mô hình nhị nguyên
4. Bổ sung khóa ngoại

## a. Xác định khóa chính cho các thực thể

Khóa chính của thực thể có thể xác định như sau:

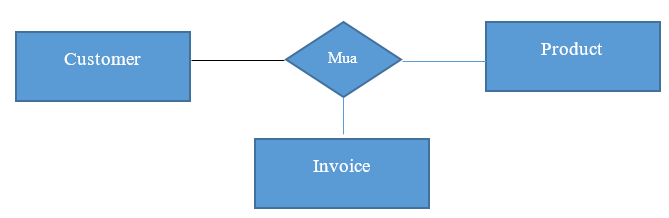
* + Chọn từ một trường có sẵn đủ điều kiện làm khóa chính như InvoiceNo chẳng hạn.
  + Nếu chưa có bạn có thể bổ sung một trường tự tăng để làm khóa chính như CustomerNo, ProductNo.

Lúc này chúng ta có các thực thể như sau:

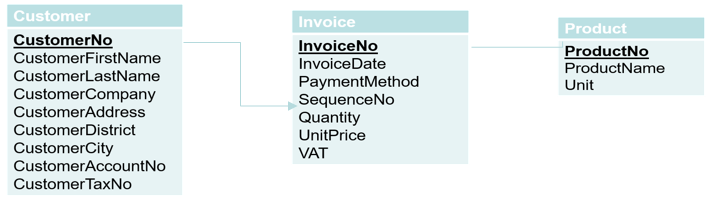


## b. Xác định mối quan hệ giữa các bảng

Xem xét các thực thể bạn có để xác định các định mối quan hệ của chúng, chúng ta có các thực thể Customer, Product và Invoice thì mối quan hệ của chúng chỉ có thể là Customer mua Product và sinh ra Invoice để ghi nhận thông tin.

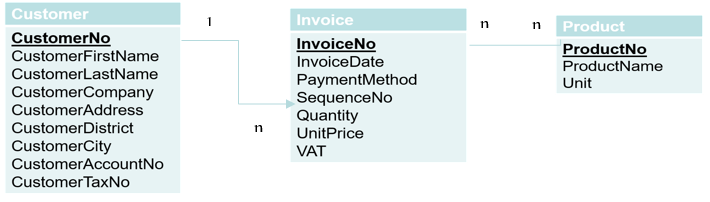


Mối quan hệ trong thế giới thực của các thực thể. Chuyển sang mô hình thực thể nó  được biểu diễn như sau:



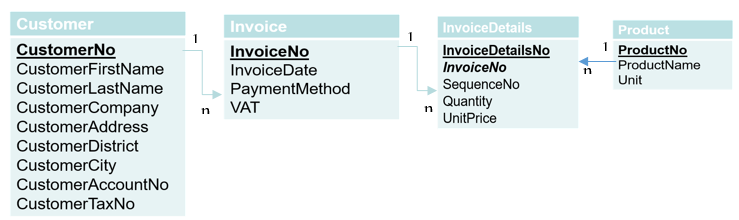
Xác định loại quan hệ giữa các thực thể như sau:

1. Quan hệ giữa Customer và Invoice, chúng ta thấy mỗi khách hàng có thể mua nhiều đơn hàng, nhưng mỗi đơn hàng chỉ bán cho 1 khách hàng. Do vậy quan hệ này là 1-n.
2. Tương tự quan hệ giữa Invoice và Product, mỗi hóa đơn có thể mua nhiều sản phẩm, mỗi sản phẩm có thể bán cho nhiều hóa đơn nên quan hệ này là quan hệ n-n.



## c. Phân tách các quan hệ để đưa về mô hình nhị nguyên

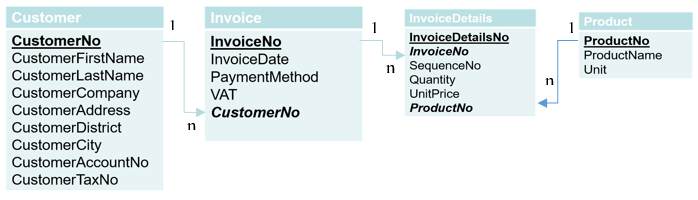
Theo mô hình cơ sở dữ liệu quan hệ nếu bạn để tồn tại mối quan hệ n-n nó sẽ gây ra dư thừa dữ liệu. Do vậy, bạn cần tách quan hệ ra thành các quan hệ 1-n bằng cách thêm vào bảng dữ liệu mới. Trong ví dụ trên chúng ta thêm vào bảng InvoiceDetails để tách nó ra thành 02 quan hệ 1-n như sau:



Ở đây bạn sẽ khó hiểu là tách trường nào cho bảng InvoiceDetails, dựa vào đâu. Việc này sẽ được giải thích rõ ở phần [**chuẩn hóa cơ sở dữ liệu**](https://iviettech.vn/blog/thiet-ke-co-so-du-lieu/10592-chuan-hoa-co-so-du-lieu.html)

## d. Bổ sung khóa ngoại cho các mối quan hệ

Khi đã xác định xong các mối quan hệ, bạn cần đặt các khóa ngoại vào các bảng bên n trong  quan hệ 1-n  để tạo liên kết giữa chúng. Lúc này chúng ta có cấu trúc cơ sở dữ liệu như sau:



# **Bước 5: Chuẩn hóa cơ sở dữ liệu**

Bước này giúp bạn xem lại cơ sở dữ liệu vừa thiết kế có đáp ứng được qui định của cơ sở dữ liệu quan hệ hay không. Tuy nhiên, cần phải [***Chuẩn hóa cơ sở dữ liệu***](https://iviettech.vn/blog/thiet-ke-co-so-du-lieu/10592-chuan-hoa-co-so-du-lieu.html) nữa.

**Làm thế nào để biết được cơ sở dữ liệu được thiết kế đúng?**

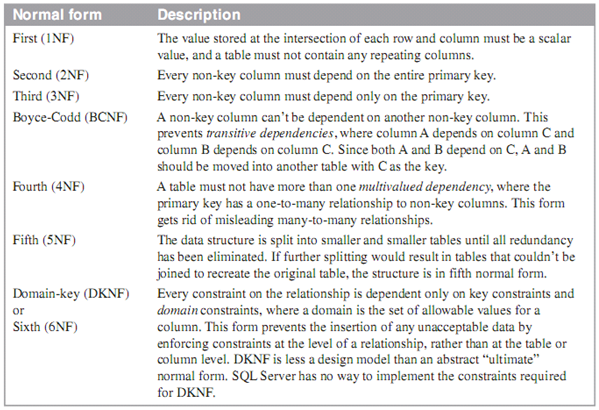
Bạn chỉ có thể kết luận cơ sở dữ liệu được thiết kế đúng khi các chức năng của hệ thống cài đặt thành công trên nó. Do vậy, bạn có thể chạy thử (dry run) các chức năng trên cấu trúc cơ sở dữ liệu để kiểm tra thiết kế của nó có phù hợp hay không.

**Chuẩn hóa cơ sở dữ liệu**

Chuẩn hóa cơ sở dữ liệu là áp dụng các qui tắc cho thiết kế của cơ sở dữ liệu nhằm giúp giảm dư thừa dữ liệu trong quá trình thiết kế Cơ sở dữ liệu. Đây là bước quan trọng để giúp bạn kiểm tra tính đúng đắn của quá trình thiết kế cơ sở dữ liệu.

**Các mức chuẩn hóa**

Về lý thuyết thì có 7 mức chuẩn hóa khác nhau:



Tạm thời mình để đầy đủ các chuẩn và mô tả bằng tiếng Anh bạn nào quan tâm thì chịu khó đọc. Tuy nhiên, đối với thiết kế Cơ sở dữ liệu thì bạn chỉ cần áp dụng đến mức 3 là được. Do vậy, chúng ta sẽ chỉ tìm hiểu kỹ 03 mức chuẩn hóa đầu tiên.

***Chuẩn 1 (1NF – First Normal Form)***: Giá trị được lưu trữ trong các ô phải là các giá trị đơn (scalar value) và trong bảng không có cột nào lặp lại.

***Chuẩn 2 (2NF – Second Normal Form)***: Mọi trường không phải là khóa phải phụ thuộc vào khóa chính.

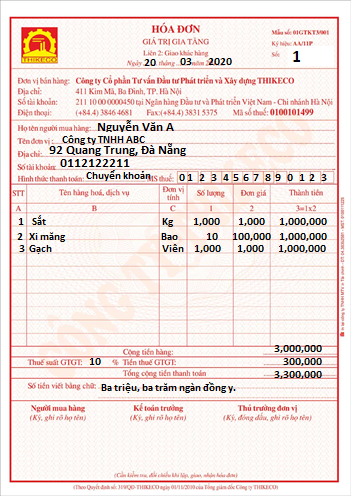
***Chuẩn 3 (3NF – Third Normal Form):***Mọi trường không phải là khóa chỉ phụ thuộc vào khóa chính mà thôi.

**Tiến trình chuẩn hóa**

1. Tiến trình để đưa bảng dữ liệu về chuẩn 1:
   * Chia các thành phần dữ liệu thành đơn vị nhỏ nhất hữu dụng
   * Loại bỏ các trường lặp lại, các trường tính toán trong bảng chúng ta có chuẩn 1
2. Tiến trình để đưa bảng dữ liệu về chuẩn 2:
   * Từ chuẩn 1, tách các trường không phụ thuộc vào khóa chính ra bảng riêng ta sẽ được chuẩn 2.
3. Tiến trình để đưa bảng dữ liệu về chuẩn 3:
   * Từ chuẩn 2, tách các trường không phụ thuộc hoàn toàn vào khóa chính (có nghĩa là có phụ thuộc thêm ít nhất một trường khác nữa ngoài khóa chính) ra bảng khác chúng ta sẽ được chuẩn 3.

**Thực hành việc chuẩn hóa**

Để hiểu rõ về các chuẩn và chuẩn hóa chúng ta thực hành sử dụng chuẩn hóa để xây dựng Cơ sở dữ liệu cho hệ thống quản lý đơn hàng



**Mẫu đơn hàng**

Sau khi phân tích dữ liệu ta được các trường như sau:

**Bảng Invoice (Hóa đơn)**

InvoiceNo: Số hóa đơn

InvoiceDate: Ngày ghi hóa đơn

CustomerName: Tên khách hàng

CustomerCompany: Tên công ty khách hành

CustomerAddress: Địa chỉ khách hàng

CustomerAccountNo: Số tài khoản của khách hàng

PaymentMethod: Phương thức thanh toán

CustomerTaxNo: Mã số thuế của khách hàng

SequenceNo: Số thứ tự mua hành

ProductName: Tên sản phẩm

Unit: Đơn vị tính của sản phẩm

Quantity: Số lượng

UnitPrice: Đơn giá

Amount: Thành tiền

TotalAmount: Tổng tiền

VAT: Thuế giá trị gia tăng

TotalPay: Tổng tiền phải trả

ByText: Ghi bằng chữ

# **1. Đưa bảng dữ liệu về chuẩn 1:**

1. Chia các thành phần dữ liệu thành đơn vị nhỏ nhất hữu dụng:
   * Chia trường CustomerName thành CustomerFirstName và CustomerLastName
   * Chia trường CustomerAddress thành CustomerAddress, CustomerDistrict và CustomerCity
2. Loại bỏ các trường lặp lại và tính toán
   * Loại bỏ các trường: Amount, TotalAmount, TotalPay, ByText

***Lúc này ta được bảng dữ liệu theo chuẩn 1 như sau:***

**Bảng Invoice (Hóa đơn)**

InvoiceNo: Số hóa đơn

InvoiceDate: Ngày ghi hóa đơn

CustomerFirstName: Tên khách hàng

CustomerLastName: Họ lót khách hàng

CustomerCompany: Tên công ty khách hành

CustomerAddress: Địa chỉ khách hàng

CustomerDistrict: Quận nơi khách hàng đang ở

CustomerCity: Thành phố nơi khách hàng đang ở

CustomerAccountNo: Số tài khoản của khách hàng

PaymentMethod: Phương thức thanh toán

CustomerTaxNo: Mã số thuế của khách hàng

SequenceNo: Số thứ tự mua hành

ProductName: Tên sản phẩm

Unit: Đơn vị tính của sản phẩm

Quantity: Số lượng

UnitPrice: Đơn giá

VAT: Thuế giá trị gia tăng

# **2. Đưa bảng dữ liệu về chuẩn 2:**

*Từ chuẩn 1, xem xét các trường không phụ thuộc vào khóa chính để tách ra thành bảng riêng.*

1. **Xem xét từ trên xuống chúng ta sẽ thấy các trường:**   
   CustomerFirstName, CustomerLastName, CustomerCompany, CustomerAddress, CustomerDistrict, CustomerCity, CustomerAccountNo, CustomerTaxNo phụ thuộc vào khách hàng chứ không phụ thuộc vào hóa đơn nên tách nhóm này ra thành bảng **Khách hàng (Customer)**, phát sinh mã khách hàng làm khóa chính.
2. **Các trường:**

ProductName, Unit phụ thuộc vào sản phẩm chứ không phải hóa đơn nên tách ra thành bảng Sản phẩm (Product) và phát sinh khóa chính cho nó.

***Lúc này chúng ta có được 03 bảng dữ liệu như sau:***

**Customer**

**CustomerNo**

CustomerFirstName

CustomerLastName

CustomerCompany

CustomerAddress

CustomerDistric

CustomerCity

CustomerAccountNo

CustomerTaxNo

**Product**

**ProductNo**

ProductName

Unit

**Invoice**

**InvoiceNo**

InvoiceDate

PaymentMethod

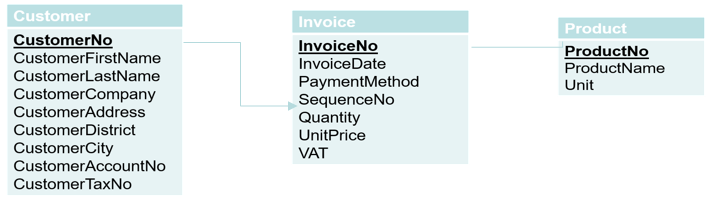
SequenceNo

Quantity

UnitPrice

VAT

Xác định mối quan hệ và đặt khóa ngoại cho trường CustomerNo và ProductNo chúng ta được sơ đồ thực thể cho mức 2 như sau:



# **3. Đưa các bảng dữ liệu về chuẩn 3:**

*Xem xét các trường từ các bảng không phụ thuộc hoàn toàn vào khóa chính ra bảng khác để về chuẩn 3.*

1. Các trường ở bảng Customer và Product đã phụ thuộc hoàn toàn vào khóa chính nên các bảng này đã ở chuẩn 3
2. Các trường SequenceNo, Quantity, UnitPrice của bảng Invoice bên cạnh việc phụ thuộc vào hóa đơn nó còn phải phụ thuộc vào ProductNo (mặt hàng nào) nữa mới biết được giá trị của nó. Nên các trường này không phụ thuộc hoàn toàn vào khóa chính. Tách các trường này ra bảng mới và đặt tên thành OrderDetails và phát sinh khóa chính cho nó.

Lúc này bảng Invoice được chia thành Invoice và InvoiceDetails như sau:

**Invoice**

**InvoiceNo**

InvoiceDate

PaymentMethod

VAT

***CustomerNo***

**InvoiceDetails**

**InvoiceDetailsNo**

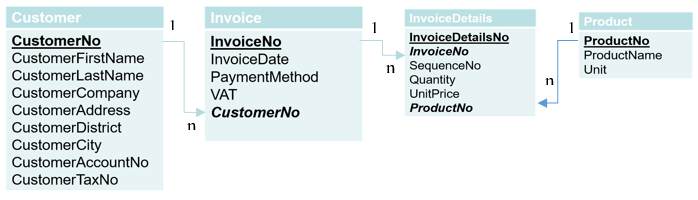
SequenceNo

***ProductID***

Quantity

UnitPrice

Lúc này tất cả các bảng đã ở chuẩn mức 3. Thiết lập mối quan hệ và khóa ngoại ta được sơ đồ quan hệ thực thể như sau:



**Lưu ý**

Trong bài này để bạn nắm rõ về các mức chuẩn hóa tôi đã áp dụng chuẩn hóa từ Mức 1 -> Mức 2 -> Mức 3. Trong thực tế, khi thiết kế cơ sở dữ liệu bạn làm qua 4 bước đầu tiên trong quá trình thiết kế cơ sở dữ liệu như ở bài Thiết kế cơ sở dữ liệu để có sơ đồ quan hệ thực thể. Sau đó, áp dụng các qui tắc chuẩn hóa này để kiểm tra các bảng và đưa các bảng về chuẩn 3 thì sẽ dễ hiểu và dễ thực hiện hơn.

**Kết luận**

Chuẩn hóa Cơ sở dữ liệu để đạt chuẩn 3 là công việc bắt buộc trong thiết kế cơ  sở dữ liệu quan hệ.