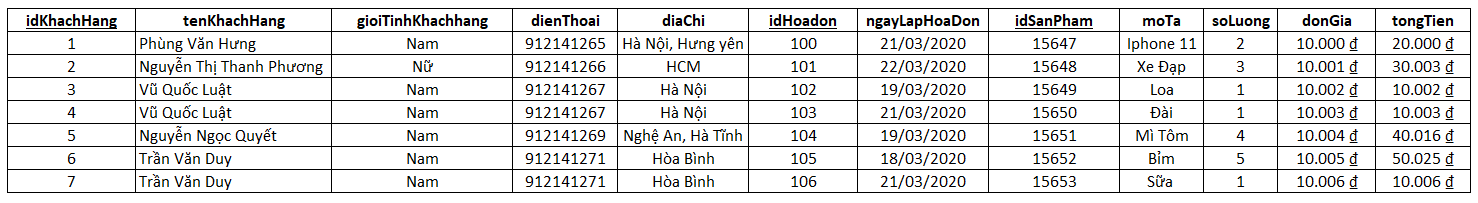
* 1. **Chuẩn hóa cơ sở dữ liệu: (MỨC LOGIC)**

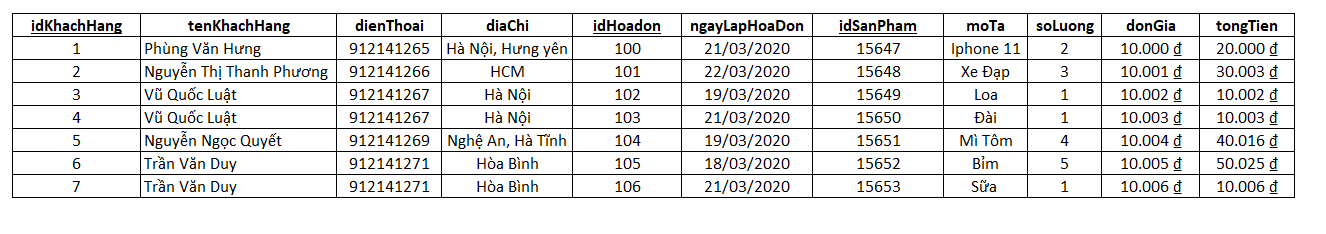
1. Giai đoạn này là 1 cách tiếp cận từ dưới lên (bottom-up approach) để thiết kế CSDL, bắt đầu từ các mối liên hệ giữa các thuộc tính. Mục đích chuẩn hóa là loại bỏ các bất thường của 1 quan hệ và để có được các quan hệ có cấu trúc tốt hơn, nhỏ hơn. **Giảm thiểu sự dư thừa dữ liệu ở mức thấp nhất và cho phép người dùng thêm, sửa, xóa mà không gây ra mâu thuẫn dữ liệu.** Các dạng chuẩn được sắp xếp theo thứ tự từ thấp đến cao. Để chuẩn hóa 2NF thì cơ sở dữ liệu của bạn phải đạt chuẩn 1NF, tương tự nếu đạt chuẩn 3NF thì phải đạt chuẩn 1NF và 2NF. Và chuẩn Boyce-Codd sẽ bao gồm 3 loại chuẩn 1NF, 2NF và 3NF. Mục đích: **Giảm sự dư thừa của dữ liệu - đảm bảo tính nhất quán của dữ liệu - hỗ trợ dễ dàng cập nhật, truy xuất dữ liệu.**



1. **Các dạng chuẩn hóa:**
   * **Dạng chuẩn 1 (1NF – first normal form)**

Một bảng được gọi là ở dạng **1NF** nếu miền giá trị của một thuộc tính chỉ chứa giá trị nguyên tố đơn (không phân chia được) và giá trị của mỗi thuộc tính cũng là một giá trị đơn lấy từ miền giá trị của nó. Để bảng đạt chuẩn hóa dạng **1NF:**

* Các thuộc tính của bảng phải là nguyên tố không phải là thuộc tính đa trị tức là sự không thể phân chia một thuộc tính thành các phần nhỏ hơn.
* Giá trị của các thuộc tính trên bảng phải là đơn trị - không chứa nhóm lặp
* Không có một thuộc tính nào có giá trị có thể tính toán được từ một thuộc tính khác.
* Xác định được thuộc tính khóa chính.



Bảng dữ liệu hóa đơn trên vi phạm điều kiện chuẩn 1NF vì: **Địa chỉ** chứa các giá trị trùng lặp, hơn thế nữa, giá trị **diaChi** trong từng hàng không phải là đơn trị. Thuộc tính **tongtien** có thể tính được bằng cách nhân số lượng\*đơn giá không nhất thiết đưa vào bảng, gây ra dư thừa dữ liệu (Thuộc tính dẫn xuất). Vì vậy ta phải tách bảng thành bảng **KhachHang – HoaDon – DiaChi.**

**KhachHang**(idKhachHang, tenKhachHang, dienthoai).

**HoaDon**(idHoaDon, **ngayLapHoaDon**, idSanpham, **moTa**, **soLuong**, **donGia**).

**DiaChi**(idKhachHang, diachi)

* + **Dạng chuẩn 2 (2NF – second normal form)**
* Phải thỏa mãn chuẩn **1NF**
* Phụ thuộc hàm đầy đủ vào khóa chính
* Với các quan hệ có tính khóa đơn thì không phải xét – chỉ kiểm tra lược đồ có chưa phụ thuộc hàm bộ phận
* Nói ngắn gọn hơn: **Các trường thuộc tính không phải khóa chính**, **phải phụ thuộc hoàn toàn vào khóa chính**. Không được phép thụ thuộc vào 1 phần của khóa chính.

**Phân tích: moTa**, **donGia** phụ thuộc vào **idSanPham** chứ không phụ thuộc vào cả 2 khóa còn lại như **idKhachHang**,**idHoaDon.** Thuộc tính **tenKhachHang** và **dienthoai** chỉ phụ thuộc vào khóa **idKhachHang**, thuoc tinh **ngayLapHoaDon** phụ thuộc vào **idKhachHang và idHoaDon.** Thuộc tính **soLuong** phụ thuộc vào **idHoaDon và idSanpham.**

**Sau khi phân tích để đạt chẩun 2NF:**

* + Tách thuộc tính (idSanpham, moTa, donGia) thành một bảng SanPham.
  + Tách thuộc tính (idHoaDon, idKhachHang, ngayLapHoaDon) thành một bảng HoaDon.
  + Tách thuộc tính (idHoaDon, idSanpham, soLuong) thành một bảng trung gian giữa HoaDon và SanPham và đặt tên là HoaDonSanPham

**KhachHang**(idKhachHang, tenKhachHang, dienthoai).

**DiaChi**(idKhachHang, diachi)

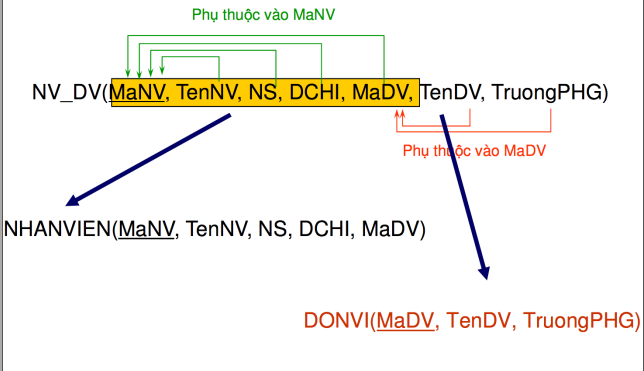
**SanPham**(idSanpham, moTa, donGia).

**HoaDon**(idHoaDon, idKhachHang, ngayLapHoaDon).

**HoaDonSanPham**(idHoaDon, idSanpham, soLuong)

Ví dụ bổ sung

* + **Dạng chuẩn 3 (3NF – third normal form)**
* Phải thỏa mãn chuẩn **2NF**
* Các trường thuộc tính không phải là khóa chính, **phải phụ thuộc trực tiếp vào khóa chính**. **Không được phép phụ thuộc bắc cầu thông qua thuộc tính khác**. Nếu có phải tách nhóm các thuộc tính đó thành quan hệ mới.
* Ví dụ: NhanVien (id, taikhoan,**maTPthanhpho**,**tenthanhpho**.)...



* + **Dạng chuẩn BCNF (Boyce-Codd normal form):** Dạng chuẩn Boyce–Codd (hoặc BCNF hoặc 3.5NF) là một dạng chuẩn được sử dụng trong chuẩn hóa dữ liệu. Nó là phiên bản mạnh hơn một chút của dạng chuẩn 3 (3NF). BCNF được phát triển vào năm 1975 bởi Raymond F. Boyce và Edgar F. Codd để giải quyết một số loại dị thường không được xử lý bởi 3NF như được định nghĩa ban đầu.
    - Phải đạt chuẩn **3NF**
    - Không có thuộc tính khóa nào phụ thuộc vào thuộc tính không khóa
      * Bước 1: Loại bỏ các thuộc tính khóa phụ thuộc hàm vào thuộc tính không khóa ra khỏi quan hệ.
      * Bước 2: Tách thuộc tính vừa loại bỏ thành một quan hệ riêng có khoá chính là thuộc tính không khóa gây ra phụ thuộc.
    - Giả sử bảng có các thuộc tính (A, B, C, D, E, **F**) với AB là khóa chính **AB** => C **AB** => D **AB** => E **F** => **B** Ta thấy bảng trên vi phạm chuẩn Boyce-Codd vì B là thuộc tính khóa nhưng lại phụ thuộc vào F là thuộc tính không khóa. Từ đó, ta sẽ tách bảng ra để theo chuẩn như sau: Bảng 1: (A, F, C, D, E) Bảng 2: (F, B). (employee\_id, first\_name, last\_name, title. Trong bảng này, cột emloyee\_id giúp xác định first\_name và last\_name. Tương tự, cặp (first\_name, last\_name) cũng sẽ xác định được employee\_id).
  + **Dạng chuẩn 4NF**
    - Phải đạt chuẩn **3NF – SV có thể tìm hiểu thêm.**
  1. **ERD**