

CHUẨN HÓA DỮ LIỆU

1 Bài toán

Thiết kế cơ sở dữ liệu quản lý các đơn đặt hàng (Invoice) của một công ty.

International Widgets 742 Evergreen Terrace Springfield, MO		INVOICE INVOICE NO: 125 DATE: September 13, 2002		
To: Foo, Inc. 23 Main St. Thorpleburg, TX		Customer No. 56		
QUANTITY	ITEM ID	DESCRIPTION	UNIT PRICE	AMOUNT
4	563	56" Blue Freen	3.50	\$14.00
32	851	Spline End (Xtra Large)	.25	\$8.00
5	692	3" Red Freen	12.00	\$60.00
TOTAL DUE				\$82.00

INVOICE INVOICE NO: 126 September 14, 2002	
Customer No. 2	
ITEM ID	AMOUNT
50	\$1,750.00
750	\$9,000.00
TOTAL DUE \$10,750.00	

750	692	3" Red Freen	12.00	
TOTAL DUE				\$10,750.00

2 Phân tích bài toán

Đọc trên đơn đặt hàng, chúng ta thấy có những thông tin sau:

Mã đơn đặt hàng (Invoice No: 125), Ngày đặt hàng (Date: September 13, 2002), Mã khách hàng (Customer No: 56), Tên khách hàng (Foo, Inc.), Địa chỉ của khách hàng (23 Main St. Thorpleburg, TX). Mã mặt hàng (Item ID), Mô tả mặt hàng (Description), Số lượng mặt hàng (Quantity), Giá đơn vị mặt hàng (Unit Price), Thành tiền mỗi mặt hàng (Amount= Quantity * Unit Price), Tổng giá thành (Total Due).

3 Chuẩn một

- ❖ Tạo bảng riêng cho mỗi nhóm dữ liệu liên quan
- ❖ Các cột của bảng phải có giá trị nguyên tố
- ❖ Các thuộc tính khóa cần được xác định

Từ phân tích bài toán ở trên, chúng ta có bảng quản lý các đơn đặt hàng như sau

orders
order_id (PK)
order_date
customer_id
customer_name
customer_address
item_id (PK)
item_description
item_qty
item_price
item_total_price
order_total_price

Trong đó, order_id và item_id là tổ hợp khóa chính, Tổ hợp đó xác định riêng biệt mỗi bản ghi (bộ/dòng) trong bảng. Ta thấy các cột của bảng có giá trị nguyên tố. Như vậy, ta đã xây dựng được cấu trúc bảng thỏa mãn chuẩn một.

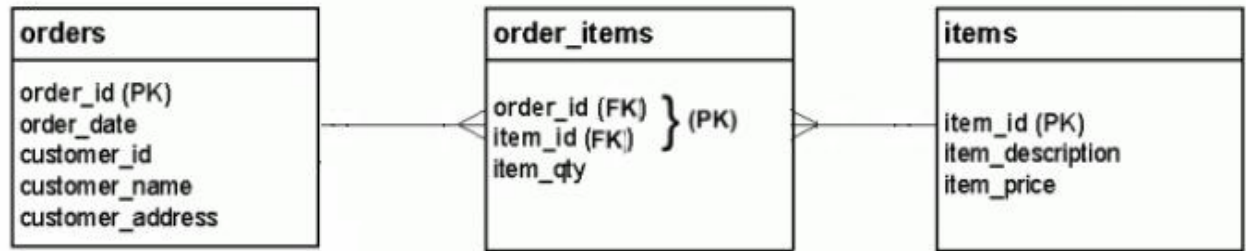
4 Chuẩn hai

- ❖ Thỏa mãn chuẩn hai
- ❖ Không có những phụ thuộc một phần
- ❖ Các bảng được liên kết thông qua khóa ngoại

Chúng ta xét tiếp bảng orders có thỏa mãn chuẩn hai. Bảng orders đã thỏa mãn chuẩn một. Tiếp đến, sự phụ thuộc một phần của các thuộc tính không phải là khóa cần được kiểm tra.

- ❖ order_date (ngày đặt hàng): ứng với mỗi order_id (mã đặt hàng), có ngày đặt hàng, tuy nhiên không nhất thiết phải có mặt hàng. Do vậy order_date chỉ phụ thuộc khóa order_id, chứ không phụ thuộc khóa item_id. → vi phạm chuẩn hai.
- ❖ customer_id (mã khách hàng): Ứng với mỗi đơn đặt hàng sẽ có một khách hàng tham gia giao dịch với công ty. Tuy nhiên việc mua hàng gì là do họ quyết định. Như vậy, customer_id phụ thuộc khóa order_id, mà không phụ thuộc khóa item_id. Các thuộc tính customer_name và customer_address lại phụ thuộc vào customer_id. Theo tính chất bắc cầu, chúng cũng giống trường hợp của customer_id. → vi phạm chuẩn hai.
- ❖ item_description (mô tả mặt hàng): dễ nhận thấy thuộc tính này phụ thuộc vào khóa item_id mà không phụ thuộc khóa order_id. Thuộc tính item_price cũng tương tự. → vi phạm chuẩn hai.
- ❖ item_qty (số lượng của một mặt hàng): Một đơn đặt hàng sẽ yêu cầu một số lượng nhất định của một mặt hàng cụ thể. Thuộc tính này phụ thuộc cả hai khóa item_id và order_id. → thỏa mãn chuẩn hai.
- ❖ Riêng trường hợp hai thuộc tính item_total_price, order_total_price: Chúng đều phụ thuộc cả hai khóa item_id và order_id. Tuy nhiên, giá trị của chúng hoàn toàn có thể tính toán dựa trên các thuộc tính khác được. Ta sẽ bỏ chúng đi để tránh dư thừa dữ liệu.

Qua phân tích trên, ta cần sửa lại cấu trúc bảng để không vi phạm điều kiện phụ thuộc một phần của chuẩn hai. Những thuộc tính chỉ phụ thuộc một khóa chúng ta sẽ tách riêng ra sang bảng khác.



5 Chuẩn ba

- ❖ Thỏa mãn chuẩn hai
- ❖ Không có phụ thuộc bắc cầu

Như đã phân tích ở trên, các bảng orders, order_items và items đều thỏa mãn chuẩn hai. Khi xét đến sự phụ thuộc bắc cầu, chúng ta thấy bảng orders vi phạm (như đã phân tích các trường customer_name, customer_address). Do đó chúng ta cần tách các thuộc tính đó sang một bảng riêng.

