



CƠ SỞ DỮ LIỆU

GV: ThS. Lê Thị Ngọc Thảo

Chương 6: Chuẩn hóa CSDL



- 6.1. Giới thiệu
- 6.2. Dạng chuẩn 1
- 6.3. Dạng chuẩn 2
- 6.4. Dạng chuẩn 3
- 6.5. Dạng chuẩn Boyce-Codd
- 6.6. Chuẩn hóa LĐ CSDL - phương pháp phân rã
- 6.7. Ví dụ

6.1. Giới thiệu



Xét quan hệ ĐẶT HÀNG (SốĐH, NgàyĐH, MãKH, MãHH, SốLượng)

Số ĐH	NgàyĐH	MãKH	MãHH	SốLượng
DH01	5/1/99	KH01	H01	50
DH02	13/2/99	KH05	H02	30
DH02	13/2/99	KH05	H03	40

Với tập Pth $F = \{ \text{SốĐH} \rightarrow \text{NgàyĐH, MãKH} ; \text{SốĐH, MãHH} \rightarrow \text{SốLượng} \}$

=> Có Trùng lặp thông tin

6.1. Giới thiệu



Sự trùng lặp thông tin dẫn đến:

- Tăng chi phí lưu trữ
- Tăng chi phí kiểm tra RBTV
- Thiếu nhất quán
- Vi phạm tính toàn vẹn của dữ liệu

6.1. Giới thiệu



Tổ chức lại thành 2 quan hệ như sau:

ĐẶT_HÀNG (SốĐH, NgàyĐH, MãKH)

Với $F1 = \{ \text{SốĐH} \rightarrow \text{NgàyĐH}, \text{MãKH} \}$

CHITIẾT_ĐH (SốĐH, MãHH, SốLượng)

Với $F2 = \{ \text{SốĐH}, \text{MãHH} \rightarrow \text{SốLượng} \}$

=> Không còn xảy ra tình trạng trùng lặp thông tin

6.1. Giới thiệu



❖ **Đánh giá chất lượng thiết kế của lược đồ CSDL**

- E.F.Codd đưa ra 3 dạng chuẩn (Normal Form)
- R.F.Boyce và E.F.Codd cải tiến dạng chuẩn 3 gọi là *dạng chuẩn Boyce-Codd* (BC)

❖ **Các dạng chuẩn được định nghĩa dựa trên khái niệm PTH**

6.1. Giới thiệu



❖ Mục đích của quá trình chuẩn hóa

- Biểu diễn được mọi quan hệ trong CSDL
- Tránh sai sót khi thêm, xóa, sửa dữ liệu
- Tránh phải xây dựng lại cấu trúc của các quan hệ khi cần đến các kiểu dữ liệu mới

6.2. DẠNG CHUẨN 1 (1NF)



❖ Định nghĩa:

- ❖ Một lược đồ quan hệ Q được gọi là đạt dạng chuẩn 1 nếu **mọi thuộc tính của Q đều là thuộc tính đơn**
- ❖ Một lược đồ CSDL được gọi là đạt dạng chuẩn 1 nếu **mọi lược đồ quan hệ con Q_i** của nó đều đạt dạng chuẩn 1

6.2. DẠNG CHUẨN 1 (1NF)



❖ Thuộc tính đơn:

- Giả sử có lược đồ quan hệ Q.
- Một thuộc tính A của Q gọi là thuộc tính đơn nếu nó không phải là một sự tích hợp của nhiều thuộc tính khác

❖ Ví dụ 1: CHUYÊN_MÔN (MAGV, MÔN)

MAGV	MÔN
GV01	PASC, CTDL
GV02	CSDL, PT

➔ Môn không là thuộc tính đơn

6.2. DẠNG CHUẨN 1 (1NF)



❖ Ví dụ 1 :

➔ Quan hệ CHUYÊN_MÔN không đạt dạng chuẩn 1

➔ **Khắc phục:** CHUYÊN_MÔN (MAGV, MÔN)

<u>MAGV</u>	<u>MÔN</u>
GV01	PASC
GV01	CTDL
GV02	CSDL
GV02	PTTKHT

6.3. DẠNG CHUẨN 2 (2NF)



❖ Định nghĩa:

- ❖ Một lược đồ quan hệ Q gọi là đạt dạng chuẩn 2 nếu:
 - Q đạt dạng chuẩn 1
 - Mọi thuộc tính không khóa của Q đều **phụ thuộc đầy đủ** vào các khóa của Q
- ❖ Một lược đồ CSDL được gọi là đạt dạng chuẩn 2 nếu mọi lược đồ quan hệ con Q_i của nó đều ở dạng chuẩn 2

6.3. DẠNG CHUẨN 2 (2NF)



❖ Phụ thuộc đầy đủ:

- Giả sử có 1 lược đồ quan hệ Q và tập PTH F.
- Thuộc tính A được gọi là **phụ thuộc đầy đủ** vào 1 tập thuộc tính X nếu:
 - $A \in X_F^+$
 - $X \rightarrow A$ là phụ thuộc hàm nguyên tố
(không tồn tại $X' \subsetneq X$, mà $X' \rightarrow A$)

6.3. DẠNG CHUẨN 2 (2NF)



❖ Ví dụ:

- ĐẶT_HÀNG (SốĐH, MãHH, NgàyĐH, MãKH, SL)
 - $F = \{ \text{SốĐH} \rightarrow \text{NgàyĐH}, \text{MãKH}; \text{SốĐH}, \text{MãHH} \rightarrow \text{SL} \}$
- ➔ Không đạt dạng chuẩn 2

❖ Khắc phục: Tách thành 2 quan hệ:

- ĐẶT_HÀNG (SốĐH, NgàyĐH, MãKH)
Với $F1 = \{ \text{SốĐH} \rightarrow \text{NgàyĐH}, \text{MãKH} \}$
- CHITIẾT_ĐH (SốĐH, MãHH, SL)
Với $F2 = \{ \text{SốĐH}, \text{MãHH} \rightarrow \text{SL} \}$

6.3. DẠNG CHUẨN 2 (2NF)



❖ Nhận xét

- Nếu lược đồ quan hệ Q chỉ có 1 khóa K và K chỉ có 1 thuộc tính thì Q đạt dạng chuẩn 2
- Một lược đồ quan hệ Q ở dạng chuẩn 2 vẫn có thể chứa đựng sự trùng lặp thông tin.

6.4. DẠNG CHUẨN 3 (3NF)



❖ Định nghĩa:

❖ Một lược đồ quan hệ Q đạt dạng chuẩn 3 nếu:

- Q ở dạng chuẩn 2
- Mọi thuộc tính không khóa của Q đều không **phụ thuộc bắc cầu** vào một khóa nào của Q

❖ Một lược đồ CSDL được gọi là đạt dạng chuẩn 3 nếu mọi lược đồ quan hệ con Q_i của nó đều đạt dạng chuẩn 3

6.4. DẠNG CHUẨN 3 (3NF)



❖ Phụ thuộc bắc cầu:

- Thuộc tính $A \in Q^+$ được gọi là phụ thuộc bắc cầu vào tập thuộc tính X nếu $\exists Y \in Q^+$:

1) $X \rightarrow Y \in F^+$ và $Y \rightarrow A \in F^+$

2) $Y \rightarrow X \notin F^+$

3) $A \notin (X \cup Y)$

- Khi đó $X \rightarrow A$ được gọi là PTH bắc cầu

6.4. DẠNG CHUẨN 3 (3NF)



❖ Ví dụ:

- **GIẢNG_DẠY**(MãLớp, MãSốGV, TênGV, Địa chỉ)
- $F = \{MãLớp \rightarrow MãSốGV; MãSốGV \rightarrow TênGV, Địa chỉ \}$

➔ **Không đạt dạng chuẩn 3**

❖ Khắc phục: Tách thành 2 quan hệ:

- **GIẢNG_DẠY**(MãLớp, MãSốGV)
 $F1 = \{MãLớp \rightarrow MãSốGV\}$
- **GIÁO_VIÊN**(MãSốGV, TênGV, Địa chỉ)
 $F2 = \{MãSốGV \rightarrow TênGV, Địa chỉ \}$

6.4. DẠNG CHUẨN 3 (3NF)



❖ Nhận xét:

- ❖ Chính phụ thuộc hàm bậc cầu là nguyên nhân dẫn đến tình trạng trùng lặp thông tin
- ❖ Dạng chuẩn 3 là **tiêu chuẩn tối thiểu** trong thiết kế cơ sở dữ liệu

6.5. DẠNG CHUẨN BOYCE-CODD



❖ Định nghĩa:

- ❖ Một lược đồ quan hệ Q được gọi là đạt dạng chuẩn Boyce-Codd (BC) nếu **mọi phụ thuộc hàm không hiển nhiên của F đều có vế trái chứa khóa**.
- ❖ Một lược đồ CSDL được gọi là ở dạng chuẩn BC nếu mọi lược đồ quan hệ con Q_i của nó đều đạt dạng chuẩn BC.

6.5. DẠNG CHUẨN BOYCE-CODD



❖ Nhận xét:

- ❖ Nếu 1 lược đồ quan hệ Q đạt dạng chuẩn BC thì **cũng đạt dạng chuẩn 3**.
- ❖ Trong 1 lược đồ quan hệ Q đạt dạng chuẩn BC, việc kiểm tra phụ thuộc hàm chủ yếu là **kiểm tra khóa nội**.

6.5. DẠNG CHUẨN BOYCE-CODD



❖ Ví dụ:

- **ĐẶT_HÀNG** (SốĐH, NgàyĐH, MãKH)

$$F1 = \{ \text{SốĐH} \rightarrow \text{NgàyĐH}, \text{MãKH} \}$$

- **CHITIẾT_ĐH** (SốĐH, MãHH, SốLượng)

$$F2 = \{ \text{SốĐH}, \text{MãHH} \rightarrow \text{SốLượng} \}$$

➔ 2 quan hệ đều đạt dạng chuẩn Boyce-Codd.

6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



❖ Quá trình chuẩn hóa 1 lược đồ CSDL:

❖ Nhằm mục đích nâng cao chất lượng thiết kế

❖ Đưa các lược đồ quan hệ con từ dạng chuẩn thấp lên dạng chuẩn cao hơn mà tối thiểu phải là dạng chuẩn 3.

❖ **Phương pháp phân rã** là 1 phương pháp dùng để chuẩn hóa 1 lược đồ CSDL

6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



❖ Sự bảo toàn thông tin

- Việc chuẩn hóa 1 lược đồ quan hệ hay 1 lược đồ CSDL phải bảo đảm yêu cầu: ***bảo toàn thông tin***
- Phép phân rã Q thành $Q_1, Q_2, \dots$ được gọi là bảo toàn thông tin nếu:

$$\forall T_Q: T_Q = T_Q [Q_1] \bowtie T_Q [Q_2] \dots \bowtie \dots$$

6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



❖ Định lý Delobel

- Cho lược đồ quan hệ $Q(XYZ)$ và tập PTH F
- Nếu $X \rightarrow Y \in F^+$ thì phép phân rã Q thành 2 lược đồ quan hệ con: $Q_1(XY)$ và $Q_2(XZ)$ là bảo toàn thông tin

6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



Begin

$$F^* = F \setminus \{ f \in F^+ / VT(f) \cup VP(f) \in Q^+ \}$$

IF ($F^* \neq \emptyset$) Then



**Phương pháp
phân rã:**

Begin

B1. Chọn 1 $f_0: X \rightarrow Y \in F$

B2. Tạo các lược đồ quan hệ con Q_1 và Q_2 :

$$Q_1 = X \cup Y$$

$$F_1 = \{ f \in F^+ / VT(f) \cup VP(f) \subseteq Q_1^+ \}$$

$$Q_2 = Q^+ \setminus Y$$

$$F_2 = \{ f \in F^+ / VT(f) \cup VP(f) \subseteq Q_2^+ \}$$

B3. Phân rã đệ quy Q_1 và Q_2

End

End

6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



❖ Ví dụ:

- $Q(ABCDEG)$;
- $F = \{ AE \rightarrow C, CG \rightarrow A, BD \rightarrow G, GA \rightarrow E \}$

❖ Nhận xét:

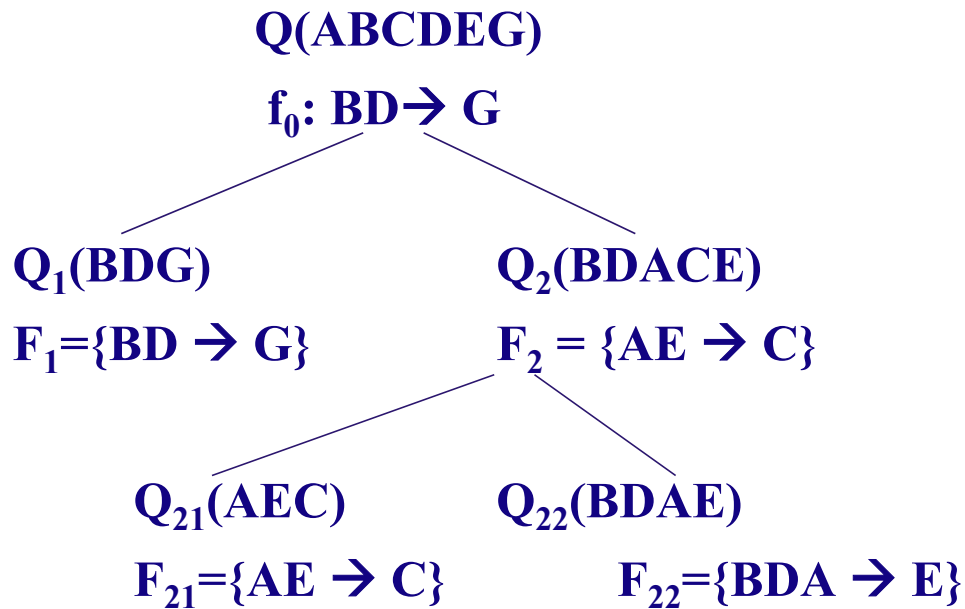
- Khóa của Q là $\{BDA\}$
- $BD \rightarrow G \rightarrow$ **không đạt dạng chuẩn 2**

➔ Sử dụng phương pháp phân rã

6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



Ví dụ:



6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



❖ Nhận xét:

- Thuật toán phân rã như trên là bảo toàn thông tin (định lý Delobel)
- Các lược đồ quan hệ con cuối cùng (nút lá trong cây phân rã) đều đạt ít nhất là **dạng chuẩn 3**
- Thuật toán phân rã có thể tạo ra các lược đồ quan hệ con không có nhiều ngữ nghĩa trong thực tế

6.6. CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ



❖ Nhận xét:

- Chất lượng của CSDL kết quả có phụ thuộc vào việc chọn pth f_0 ở từng bước phân rã.
- Thông thường pth được chọn là pth gây ra chất lượng xấu của lược đồ quan hệ. (pth không đầy đủ, pth bắc cầu).

6.7. VÍ DỤ



Ví dụ 1: Xét quan hệ

SODA	TENDA	TRGDA	MSNV	TENNV	PHG	HSL	TGIAN
100	TM Dien Tu	789	123	TUAN	KD	65	10
			990	NAM	KT	45	6
			234	NGA	KT	35	6
			542	MINH	BH	30	12
110	DTao Tu Xa	820	432	HONG	KD	50	5
			689	LE	KT	35	12
			712	DUNG	BH	30	8
120	Cap Quang	980	834	HOA	HT	80	4
			380	HOA	KT	35	11
			553	THAI	BH	30	12
			123	TUAN	KD	65	7
130	Trac Nghiem	550	340	CANH	KD	65	7

➔ Không đạt dạng chuẩn 1

6.7. VÍ DỤ



Cách 1: Chuyển quan hệ về dạng

<u>SODA</u>	TENDA	TRGDA	<u>MSNV</u>	TENNV	PHG	HSL	TGIAN
100	TM Dien Tu	789	123	TUAN	KD	65	10
100	TM Dien Tu	789	990	NAM	KT	45	6
100	TM Dien Tu	789	234	NGA	KT	35	6
100	TM Dien Tu	789	542	MINH	BH	30	12
110	Dtao tu xa	820	432	HONG	KD	50	5
110	Dtao tu xa	820	689	LE	KT	35	12
110	Dtao tu xa	820	712	DUNG	BH	30	8
120	Cap Quang	980	834	HOA	HT	80	4
120	Cap Quang	980	380	HOA	KT	35	11
120	Cap Quang	980	553	THAI	BH	30	12
120	Cap Quang	980	123	TUAN	KD	65	7
130	Trac Nghiem	550	340	CANH	KD	65	7

→ Dạng chuẩn 1

6.7. VÍ DỤ



Cách 2: Phân rã thành 2 quan hệ DEAN(SODA, TENDA, TRGDA)

<u>SODA</u>	TENDA	TRGDA
100	TM Dien Tu	789
110	DTao Tu Xa	820
120	Cap Quang	980
130	Trac Nghiem	550

→ Dạng chuẩn 1

6.7. VÍ DỤ

DEAN_NHANVIEN

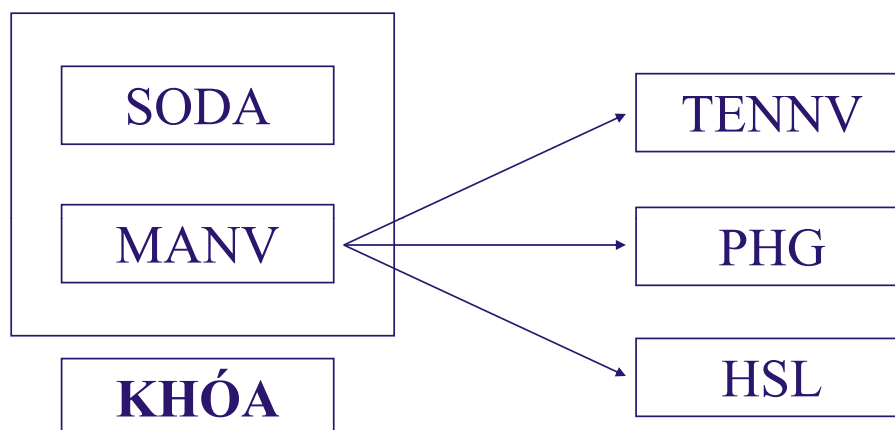
<u>SODA</u>	<u>MSNV</u>	TENNV	PHG	HSL	TGIAN
100	123	TUAN	KD	65	10
100	990	NAM	KT	45	6
100	234	NGA	KT	35	6
100	542	MINH	BH	30	12
110	432	HONG	KD	50	5
110	689	LE	KT	35	12
110	712	DUNG	BH	30	8
120	834	HOA	HT	80	4
120	380	HOA	KT	35	11
120	553	THAI	BH	30	12
120	123	TUAN	KD	65	7
130	340	CANH	KD	65	7

→ Đạt dạng chuẩn 1

6.7. VÍ DỤ

Quan hệ DEAN_NHANVIEN

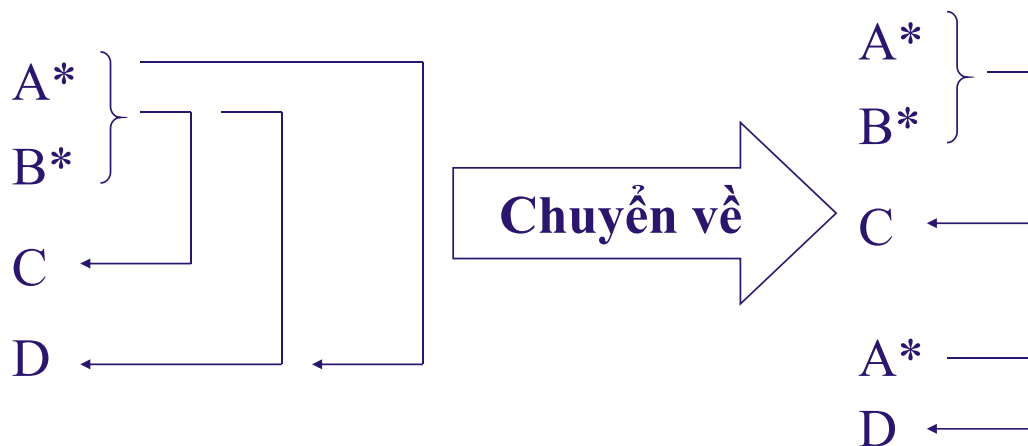
không đạt dạng chuẩn 2 vì phụ thuộc không đầy đủ



6.7. VÍ DỤ



Chuyển về dạng chuẩn 2



6.7. VÍ DỤ



PHANCONG

<u>SODA</u>	<u>MSNV</u>	<u>TGIAN</u>
100	123	10
100	990	6
100	234	6
100	542	12
110	432	5
110	689	12
110	712	8
120	834	4
120	380	11
120	553	12
120	123	7
130	340	7

NHANVIEN

<u>MSNV</u>	<u>TENNV</u>	<u>PHG</u>	<u>HSL</u>
123	TUAN	KD	65
990	NAM	KT	45
234	NGA	KT	35
542	MINH	BH	30
432	HONG	KD	50
689	LE	KT	35
712	DUNG	BH	30
834	HOA	HT	80
380	HOA	KT	35
553	THAI	BH	30
340	CANH	KD	65

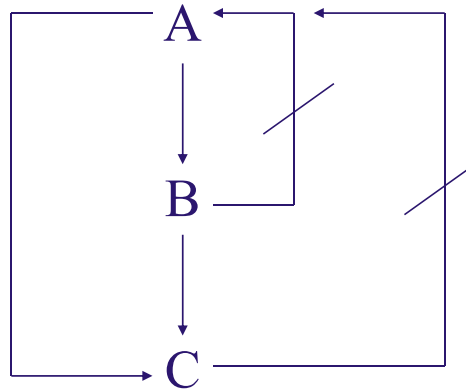
→ Đạt dạng chuẩn 2

6.7. VÍ DỤ



Giả sử có pth PHG $\rightarrow$ HSL

$\rightarrow$ Quan hệ NHÂNVIEN **không đạt dạng chuẩn 3**
vì tồn tại phụ thuộc hàm bậc cầu

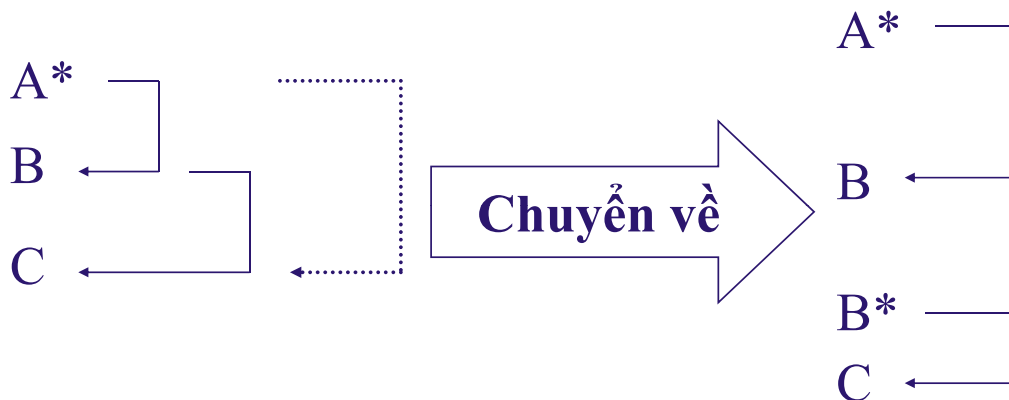


Điều kiện để xác định phụ thuộc hàm bậc cầu

6.7. VÍ DỤ



Chuyển về dạng chuẩn 3



6.7. VÍ DỤ



NHANVIEN

<u>MSNV</u>	<u>TENNV</u>	<u>PHG</u>
123	TUAN	KD
990	NAM	KT
234	NGA	KT
542	MINH	BH
432	HONG	KD
689	LE	KT
712	DUNG	BH
834	HOA	HT
380	HOA	KT
553	THAI	BH
340	CANH	KD

PHONG

<u>PHG</u>	<u>HSL</u>
KD	65
KT	45
KT	35
BH	30
KD	50
KT	35
BH	30
HT	80
KT	35
BH	30
KD	65

→ **Đạt dạng chuẩn 3**

6.7. VÍ DỤ



Ví dụ 2: Xét quan hệ SANXUAT

<u>MS_NHAMAY</u>	<u>TEN_NHAMAY</u>	<u>MS_SANPHAM</u>	<u>SOLUONG</u>
M101	TIEN DAT	K1234	1200
M101	TIEN DAT	J0055	600
M101	TIEN DAT	Y3333	350
M222	SONY VN	K1234	800

- ❖ Mỗi nhà máy có Mã số và Tên duy nhất, 1 nhà máy có thể SX nhiều SP, nhiều nhà máy có thể SX cùng loại SP → **Có 2 Khóa:**
 - (MS_NHAMAY, MS_SANPHAM)
 - (TEN_NHAMAY, MS_SANPHAM)

→ **Đạt dạng chuẩn 3**

6.7. VÍ DỤ



Xét quan hệ: SANXUAT

MS_NHAMAY	TEN_NHAMAY	MS_SANPHAM	SOLUONG
M101	TIEN DAT	K1234	1200
M101	TIEN DAT	J0055	600
M101	TIEN DAT	Y3333	350
M222	SONY VN	K1234	800

❖ Quan hệ SANXUAT:

- Có thể gây dữ liệu xấu khi cập nhật.
- Ví dụ như khi 1 nhà máy đổi tên có thể làm cho **dữ liệu thiếu nhất quán.**

6.7. VÍ DỤ



Cách 1: Tách thành 2 quan hệ

NHAMAY

<u>MS_NHAMAY</u>	TEN_NHAMAY
M101	TIEN DAT
M222	SONY VN

➔ **Đạt dạng chuẩn BC**

NHAMAY_SANPHAM

<u>MS_NHAMAY</u>	<u>MS_SANPHAM</u>	SOLUONG
M101	K1234	1200
M101	J0055	600
M101	Y3333	350
M222	K1234	800

Các cấp chuẩn hóa quan hệ



Dạng chưa được chuẩn hoá

Loại bỏ các thuộc tính không nguyên tố

Dạng chuẩn thứ nhất - 1NF

Loại bỏ các phụ thuộc không đầy đủ

Dạng chuẩn thứ hai - 2 NF

Loại bỏ các phụ thuộc bắc cầu

Dạng chuẩn thứ ba - 3 NF