

# Chuẩn hóa CSDL

# NỘI DUNG

8.1. Phụ thuộc hàm

8.5. Phủ tối thiểu

8.2. Hệ luật dẫn Armstrong

8.6. Khóa của lược đồ quan hệ

8.3. Bao đóng

8.7. Dạng chuẩn

8.4. Phụ thuộc hàm tương đương

8.8. Chuẩn hóa CSDL

## 8.1. Phụ thuộc hàm

### 8.1.1. Định nghĩa

Phụ thuộc hàm là ràng buộc giữa 2 tập thuộc tính của 01 lược đồ quan hệ

- $R(A_1, A_2, \dots, A_n)$  là lược đồ quan hệ.
- $X, Y$  là hai tập con của tập thuộc tính  $\Omega = \{A_1, A_2, \dots, A_n\}$ .

Ta nói  $Y$  phụ thuộc hàm vào  $X$ :  $X \rightarrow Y$

Với mỗi giá trị tại  $X$  trong  $R$  xác định duy nhất một giá trị của  $Y$  trong  $R$

## 8.1. Phụ thuộc hàm

Một thể hiện của R thỏa phụ thuộc hàm  $X \rightarrow Y$

### 8.1.2. Tính chất

- Nếu  $\forall t_1, t_2 \in R$

thì  $t_1.X = t_2.X \Rightarrow t_1.Y = t_2.Y$

- Nếu  $t_1.X = t_2.X \Rightarrow t_1.Y \neq t_2.Y$

thì lược đồ vi phạm phụ thuộc hàm  $X \rightarrow Y$

## 8.1. Phụ thuộc hàm

MASV	HOTEN	MONHOC	DIEMTHI
00CDTH189	Nguyễn Văn Thành	Cấu Trúc Dữ Liệu	7
00CDTH189	Nguyễn Văn Thành	Cơ Sở Dữ Liệu	9
00CDTH211	Trần Thu Hà	Kỹ Thuật Lập Trình	5
00CDTH189	Nguyễn Văn Thành	Kỹ Thuật Lập Trình	8

1. MONHOC → DIEMTHI
2. HOTEN → DIEMTHI
3. MASV → DIEMTHI

### 8.1.3. Ví dụ

A	B	C	D	E
1	2	3	4	5
1	4	3	4	5
1	2	4	4	1

1. AB → C
2. B → D
3. DE → A

## 8.1. Phụ thuộc hàm

- Nếu  $X \supseteq Y$  thì  $X \rightarrow Y$
- Với  $r$  là quan hệ bất kỳ,  $F$  là tập phụ thuộc hàm thỏa trên  $r$ , ta luôn có  $F \supseteq \{\text{các phụ thuộc hàm hiển nhiên}\}$

### 8.1.8. Phụ thuộc hàm hiển nhiên

## 8.2. Hệ luật dẫn Armstrong

### 1. Luật phản xạ (reflexive rule)

8.2.1. Ba luật

Nếu  $Y \subseteq X$  thì  $X \rightarrow Y$

### 2. Luật tăng trưởng (augmentation rule)

Nếu  $X \rightarrow Y$  thì  $XZ \rightarrow YZ$

### 3. Luật bắc cầu (Transivity Rule)

Nếu  $X \rightarrow Y$  và  $Y \rightarrow Z$  thì  $X \rightarrow Z$

## 8.2. Hệ luật dẫn Armstrong

### 1. Luật hợp (Union Rule)

### 8.2.2. Ba hệ quả

Nếu  $X \rightarrow Y$  và  $X \rightarrow Z$  thì  $X \rightarrow YZ$

### 2. Luật bắc cầu giả (Pseudotransitivity Rule)

Nếu  $X \rightarrow Y$  và  $WY \rightarrow Z$  thì  $XW \rightarrow Z$

### 3. Luật phân rã (Decomposition Rule)

Nếu  $X \rightarrow Y$  và  $Z \subset Y$  thì  $X \rightarrow Z$

## 8.3. Bao đóng

### 8.3.1. Định nghĩa

- Gọi  $F$  là tập các phụ thuộc hàm trên tập thuộc tính  $Q$
- Bao đóng của  $F$  là tất cả các phụ thuộc hàm có thể suy ra từ  $F$  dựa trên các tiên đề Armstrong

## 8.3. Bao đóng

### 8.3.2. Thuật toán tìm bao đóng

Tính liên tiếp tập các tập thuộc tính  $X^0, X^1, X^2, \dots$  theo phương pháp sau:

1. *Bước 1:*  $X^0 = X$
2. *Bước 2:* lần lượt xét các phụ thuộc hàm của F
  - Nếu  $Y \rightarrow Z$  có  $Y \subseteq X_i$  thì  $X^{i+1} = X^i \cup Z$
  - Loại phụ thuộc hàm  $Y \rightarrow Z$  khỏi F
3. *Bước 3:* Nếu ở bước 2 không tính được  $X^{i+1}$  thì  $X^i$  chính là bao đóng của X.  
Ngược lại lặp lại bước 2.

## 8.3. Bao đóng

### 8.3.3. Ví dụ

**Cho lược đồ quan hệ  $R(A,B,C,D,E,G,H)$  và tập phụ thuộc hàm**

$F = \{B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D\}$ .

**Tìm bao đóng của  $X = \{A,C\}$  trên  $F$ ?**

$$X^{(0)} = \{A, C\} \quad AC \rightarrow D$$

$$X^{(1)} = \{A, C, D\} \quad DA \rightarrow CE$$

$$X^{(2)} = \{A, C, D, E\} \quad D \rightarrow H$$

$$X^{(3)} = \{A, C, D, E, H\}$$

$$(AC)^+ = ACDEH$$

## 8.3. Bao đóng

1. Cho lược đồ quan hệ  $Q(A,B,C,D,E,G)$   
và  $F = \{A \rightarrow C; A \rightarrow EG; B \rightarrow D; G \rightarrow E\}$ . Tìm  $(AB)^+$  và  $(CGD)^+$
2. Cho lược đồ quan hệ  $U(A,B,C,D,E,G,H)$   
và  $F = \{AB \rightarrow C; D \rightarrow EG; ACD \rightarrow B; C \rightarrow A; BE \rightarrow C; CE \rightarrow AG; BC \rightarrow D; CG \rightarrow BD; G \rightarrow H\}$ .  
Tìm  $(D)^+$ ,  $(DE)^+$ ,  $(BE)^+$ ,  $(CG)^+$
3. Cho lược đồ quan hệ  $U(A,B,C,D,E,G,H)$   
và  $F = \{C \rightarrow G; BG \rightarrow CD; AEG \rightarrow BC; CG \rightarrow AE; B \rightarrow CG\}$ .  
Tìm  $(C)^+$ ,  $(B)^+$ ,  $(AEG)^+$

### 8.3.4. Bài tập

## 8.4. Phụ thuộc hàm tương đương

### 8.4.1. Định nghĩa

Hai tập phụ thuộc hàm  $F$  và  $G$  là tương đương nếu  $F^+ = G^+$

Ký hiệu:  $F \equiv G$  ( $F \sim G$ ).

## 8.4. Phụ thuộc hàm tương đương

### 8.4.2. Thuật toán

*Bước 1:* Với mỗi phụ thuộc hàm  $X \rightarrow Y$  của  $F$  ta xác định xem  $X \rightarrow Y$  có là thành viên của  $G$  không.

*Bước 2:* Với mỗi phụ thuộc hàm  $X \rightarrow Y$  của  $G$  ta xác định xem  $X \rightarrow Y$  có là thành viên của  $F$  không.

Nếu cả hai bước trên đều đúng thì  $F \equiv G$

## 8.4. Phụ thuộc hàm tương đương

### 8.4.3. Ví dụ

Cho lược đồ quan hệ  $R(ABCDE)$  hai tập phụ thuộc hàm:

$$F = \{A \rightarrow BC, A \rightarrow D, C \rightarrow E\} \text{ và } G = \{A \rightarrow BCE, A \rightarrow ABD, C \rightarrow E\}$$

F có tương đương với G không?

#### Cách 1: Dùng bao đóng

$$\text{Xét } (A)_F^+ = ABCDE \supset BCE \rightarrow A \rightarrow BCE$$

$$\text{Xét } (A)_F^+ = ABCDE \supset ABD \rightarrow A \rightarrow ABD$$

$$\text{Xét } (C)_F^+ = CE \supset E \rightarrow C \rightarrow E$$

Các PTH trong G đều được suy diễn từ  $F^+$  (1)

$$(1,2) \rightarrow F \equiv G$$

$$\text{Xét } (A)_G^+ = ABCDE \supset BC \rightarrow A \rightarrow BC$$

$$\text{Xét } (A)_G^+ = ABCDE \supset D \rightarrow A \rightarrow D$$

$$\text{Xét } (C)_G^+ = CE \supset E \rightarrow C \rightarrow E$$

$$\begin{aligned} &\text{Các PTH trong } F \text{ đều được suy diễn từ } G^+ \\ &(2) \end{aligned}$$

## 8.4. Phụ thuộc hàm tương đương

### 8.4.3. Ví dụ

Cho lược đồ quan hệ  $R(ABCDE)$  hai tập phụ thuộc hàm:

$$F = \{A \rightarrow BC, A \rightarrow D, C \rightarrow E\} \text{ và } G = \{A \rightarrow BCE, A \rightarrow ABD, C \rightarrow E\}$$

$F$  có tương đương với  $G$  không?

#### Cách 2: Dùng luật Armstrong

Xét  $F$

$A \rightarrow BC \rightarrow A \rightarrow B$  (i),  $A \rightarrow C$  (ii) (luật phân rã)

(ii)  $A \rightarrow C, C \rightarrow E \rightarrow A \rightarrow E$  (iii) (luật bắc cầu)

(i, ii, iii)  $\rightarrow A \rightarrow BCE$  (iv)

(i)  $A \rightarrow B, A \rightarrow D \rightarrow A \rightarrow BD$  (v) (luật hợp)

$A \rightarrow BD \rightarrow A \rightarrow ABD$  (vi) (luật tăng trưởng)

(iv, vi) PTH trong  $G$  đều được suy diễn từ  $F$  (1)

Xét  $G$

$A \rightarrow BCE \rightarrow A \rightarrow BC$  (i),  $A \rightarrow E$  (luật phân rã)

$A \rightarrow ABD \rightarrow A \rightarrow AB, A \rightarrow D$  (ii) (luật phân rã)

(i, ii)  $\rightarrow$  Các PTH trong  $F$  đều được suy diễn từ  $G$   
(2)  $(1,2) \rightarrow F \equiv G$

## 8.4. Phụ thuộc hàm tương đương

### 8.4.4. Bài tập

Cho  $F = \{ A \rightarrow C, AC \rightarrow D, E \rightarrow AD, E \rightarrow H \}$  và  $G = \{ A \rightarrow CD, E \rightarrow AH \}$

$F$  và  $G$  có tương đương không?

## 8.5. Phụ thuộc hàm tối thiểu

### 8.5.1. Định nghĩa

Một phủ tối thiểu của tập phụ thuộc hàm F là một tập phụ thuộc hàm G. Trong đó:

- G tương đương với F (tức  $G^+ = F^+$ )
- Tất cả các phụ thuộc hàm trong G đều có dạng  $X \rightarrow A$
- Không thể làm G nhỏ hơn được nữa (nghĩa là không thể xóa đi bất kỳ PTH nào trong G, hay xóa đi bất kỳ thuộc tính nào bên phải, bên trái của mỗi phụ thuộc hàm mà G vẫn tương đương với F)

## 8.5. Phụ thuộc hàm tối thiểu

### 8.5.2. Tính chất

F là một tập phụ thuộc hàm tối thiểu nếu thỏa mãn 3 điều kiện sau:

- F là tập phụ thuộc hàm có vẽ trái không dư thừa
- F là tập phụ thuộc hàm có vẽ phải một thuộc tính
- F là tập phụ thuộc hàm không dư thừa

## 8.5. Phụ thuộc hàm tối thiểu

### 8.5.3. Thuật toán tìm phủ tối thiểu

1. Bước 1: Loại bỏ thuộc tính dư thừa ở vẽ trái của từng PTH
2. Bước 2: Tách vẽ phải mỗi PTH sao cho các vẽ phải chỉ chứa 1 thuộc tính
3. Bước 3: Loại bỏ các PTH dư thừa trong F

## 8.5. Phụ thuộc hàm tối thiểu

Cho

8.5.4. Ví dụ

$R(A, B, C, D, E, F, G, H)$

$T = \{ABH \rightarrow CK, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow AD, E \rightarrow F, BH \rightarrow E\}.$

Tìm phủ tối thiểu.

## 8.5. Phụ thuộc hàm tối thiểu

### Bước 1 – Loại bỏ thuộc tính dư thừa ở vế trái

$$T = \{ABH \rightarrow CK, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow AD, E \rightarrow F, BH \rightarrow E\}$$

#### xét $ABH \rightarrow CK$ .

- $\text{Bỏ } A; (BH)^+ = BHEFAD$  chứa  $A \rightarrow A$  dư thừa
- $\text{Bỏ } B; (AH)^+ = AHD$  không chứa  $B \rightarrow B$  không dư thừa
- $\text{Bỏ } H; (AB)^+ = ABD$  không chứa  $H \rightarrow H$  không dư thừa

$$T = \{BH \rightarrow CK, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow AD, E \rightarrow F, BH \rightarrow E\}$$

#### xét $BGH \rightarrow F$ .

- $\text{Bỏ } B; (GH)^+ = GH$  không chứa  $B \rightarrow B$  không dư thừa
- $\text{BỎ } G; (BH)^+ = BCKEFAD$  không chứa  $G \rightarrow G$  không dư thừa
- $\text{BỎ } H; (BG)^+ = BG$  không chứa  $H \rightarrow H$  không dư thừa

$$T = \{BH \rightarrow CK, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow AD, E \rightarrow F, BH \rightarrow E\}$$

#### xét $BH \rightarrow E$ .

- $\text{BỎ } B; (H)^+ = H$  không chứa  $B \rightarrow B$  không dư thừa
- $\text{BỎ } H; (B)^+ = B$  không chứa  $H \rightarrow H$  không dư thừa

$$T = \{BH \rightarrow CK, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow AD, E \rightarrow F, BH \rightarrow E\}$$

### 8.5.4. Ví dụ

## 8.5. Phụ thuộc hàm tối thiểu

Bước 2 – Tách vế phải còn 01 thuộc tính

8.5.4. Ví dụ

$$T = \{BH \rightarrow CK, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow AD, E \rightarrow F, BH \rightarrow E\}$$



$$T = \{BH \rightarrow C, BH \rightarrow K, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow A, F \rightarrow D, E \rightarrow F, BH \rightarrow E\}$$

## 8.5. Phụ thuộc hàm tối thiểu

### Bước 3 – Loại bỏ thuộc phụ thuộc hàm dư thừa

### 8.5.4. Ví dụ

$$T = \{BH \rightarrow C, BH \rightarrow K, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow A, F \rightarrow D, E \rightarrow F, BH \rightarrow E\}$$

Xét  $BH \rightarrow C$ ;  $(BH)^+ = BHKEFAD$  không chứa C nên  $BH \rightarrow C$  không dư thừa

Xét  $BH \rightarrow K$ ;  $(BH)^+ = BHCEFAD$  không chứa K nên  $BH \rightarrow K$  không dư thừa

Xét  $A \rightarrow D$ ;  $(A)^+ = A$  không chứa D nên  $A \rightarrow D$  không dư thừa

Xét  $C \rightarrow E$ ;  $(C)^+ = C$  không chứa E nên  $C \rightarrow E$  không dư thừa

Xét  $BGH \rightarrow F$ ;  $(BGH)^+ = BGHCKEFAD$  chứa F nên  $BGH \rightarrow F$  dư thừa

$$T = \{BH \rightarrow C, BH \rightarrow K, A \rightarrow D, C \rightarrow E, F \rightarrow A, F \rightarrow D, E \rightarrow F, BH \rightarrow E\}$$

Xét  $C \rightarrow E$ ;  $(C)^+ = C$  không chứa E nên  $C \rightarrow E$  không dư thừa

Xét  $F \rightarrow A$ ;  $(F)^+ = FD$  không chứa A nên  $F \rightarrow A$  không dư thừa

Xét  $F \rightarrow D$ ;  $(F)^+ = FAD$  chứa D nên  $F \rightarrow D$  dư thừa

$$T = \{BH \rightarrow C, BH \rightarrow K, A \rightarrow D, C \rightarrow E, F \rightarrow A, E \rightarrow F, BH \rightarrow E\}$$

Xét  $E \rightarrow F$ ;  $(E)^+ = E$  không chứa F nên  $E \rightarrow F$  không dư thừa

Xét  $BH \rightarrow E$ ;  $(BH)^+ = BHCKEFAD$  chứa E nên  $BH \rightarrow E$  dư thừa

$$T = \{BH \rightarrow C, BH \rightarrow K, A \rightarrow D, C \rightarrow E, F \rightarrow A, E \rightarrow F\}$$

## 8.5. Phụ thuộc hàm tối thiểu

Cho lược đồ quan hệ  $Q(A,B,C,D)$

và tập phụ thuộc  $F = \{AB \rightarrow CD, B \rightarrow C, C \rightarrow D\}$ .

Tìm phủ tối thiểu của  $F$ .

*8.5.5. Bài tập*

## 8.6. Khóa của lược đồ quan hệ

Định Nghĩa: Cho lược đồ quan hệ  $R(A_1, A_2, \dots, A_n)$

### 8.6.1. Định nghĩa

- $U$  là tập thuộc tính của  $R$ .
- $F$  là tập phụ thuộc hàm trên  $R$ .
- $K$  là tập con của  $U$

$K$  là một khóa của  $R$  nếu thỏa 2 điều kiện sau:

- $K^+ = U$
- Không tồn tại  $K' \subset K$  sao cho  $K'^+ = U$

## 8.6. Khóa của lược đồ quan hệ

- Bước 1: gán  $K = U$ .
- Bước 2: A là một thuộc tính của K, đặt  $K' = K - A$ .

### 8.6.2. Thuật toán tìm một khóa

- ✓ Nếu  $K'^+ = U$  thì gán  $K = K'$  thực hiện lại bước 2.
- ✓ Nếu muốn tìm các khóa khác (nếu có) của lược đồ quan hệ, ta có thể thay đổi thứ tự loại bỏ các phần tử của K.

## 8.6. Khóa của lược đồ quan hệ

### 8.6.2. Thuật toán tìm một khóa

Cho lược đồ quan hệ  $R(A,B,C,D,E)$  và tập phụ thuộc hàm  $F$  như sau:

$F=\{AB \rightarrow C, AC \rightarrow B, BC \rightarrow DE\}$  tìm khóa  $K$ ?

Giải:

B1:	$K=U$	$\Rightarrow$	$K=ABCDE$			
B2:	$(K \setminus A)^+$	$\Rightarrow$	$(BCDE)^+$	=	$BCDE$	$\neq U$
	$\Rightarrow$	$K=ABCDE$				
B3:	$(K \setminus B)^+$	$\Rightarrow$	$(ACDE)^+$	=	$ABCDE$	$= U$
	$\Rightarrow$	$K=ACDE$				
B4:	$(K \setminus C)^+$	$\Rightarrow$	$(ADE)^+$	=	$ADE$	$\neq U$
	$\Rightarrow$	$K=ACDE$				
B5:	$(K \setminus D)^+$	$\Rightarrow$	$(ACE)^+$	=	$ACEBD$	$= U$
	$\Rightarrow$	$K=ACE$				
B6:	$(K \setminus E)^+$	$\Rightarrow$	$(AC)^+$	=	$ACBDE$	$= U$
	$\Rightarrow$	$K=AC$				

## 8.6. Khóa của lược đồ quan hệ

### 8.6.3. Thuật toán tìm tất cả các khóa

- ❖ *Bước 1:* Xác định tất cả các tập con khác rỗng của  $U=\{X_1, X_2, \dots, X_{n-1}\}$
- ❖ *Bước 2:* Tìm bao đóng của các  $X_i$
- ❖ *Bước 3:* Siêu khóa là các  $X_i$  có  $X_i^+=U$

Giả sử ta đã có các siêu khóa là  $S = \{S_1, S_2, \dots, S_m\}$

- ❖ *Bước 4:* xét mọi  $S_i, S_j$  con của  $S$  ( $i \neq j$ ), nếu  $S_i \subset S_j$  thì loại  $S_j$  ( $i, j=1..n$ ), kết quả còn lại của  $S$  chính là tập tất cả các khóa cần tìm.

## 8.6. Khóa của lược đồ quan hệ

### 8.6.3. Thuật toán tìm tất cả các khóa

$$Q(C, S, Z); F = \{ f_1 : CS \rightarrow Z; f_2 : Z \rightarrow C \}$$

$X_i$	$X_i^+$	Super key	Key
C	C		
S	S		
CS	CSZ	CS	CS
Z	ZC		
CZ	CZ		
SZ	SZC	SZ	SZ
CSZ	CSZ	CSZ	

**Answer** {C, S} và {S, Z}

## 8.6. Khóa của lược đồ quan hệ

### 8.6.4. Thuật toán cải tiến tìm tất cả các khóa

*Bước 1:* Tạo tập thuộc tính nguồn **TN** và Tập thuộc tính trung gian **TG**

**TN:** bao gồm các thuộc tính chỉ xuất hiện ở vé trái, không xuất hiện ở vé phải của F và các thuộc tính không xuất hiện ở cả vé trái và vé phải của F

**TG:** Chứa các thuộc tính xuất hiện ở cả vé trái và vé phải

*Bước 2:*

Nếu  $TG = \emptyset$  thì lược đồ quan hệ chỉ có một khóa  $K = TN$ ; kết thúc

Ngược lại Qua bước 3

*Bước 3:* tìm tất cả các tập con  $X_i$  của tập trung gian **TG**

*Bước 4:* tìm các siêu khóa  $S_i$  bằng cách

$$\forall X_i; \text{ if } (TN \cup X_i)^+ = U \text{ then } S_i = TN \cup X_i$$

*Bước 5:* tìm khóa bằng cách loại bỏ các siêu khóa không tối thiểu.

$$\forall S_i, S_j \in S; \text{ if } S_i \subset S_j \text{ then Loại } S_j \text{ ra khỏi tập siêu khóa } S$$

S còn lại chính là tập khóa cần tìm.

## 8.6. Khóa của lược đồ quan hệ

### 8.6.4. Thuật toán cải tiến tìm tất cả các khóa

Cho lược đồ quan hệ  $Q = \{ABCD\}$  và tập phụ thuộc hàm

$F = \{AB \rightarrow C, C \rightarrow A, D \rightarrow A\}$ .

Tìm tất cả các khóa thuộc tập cơ sở dữ liệu trên.

$$TN = \{B; D\}$$

$$TG = \{A; C\}$$

$x_i$	$TN \cup X_i$	$(TN \cup X_i)^+$	Siêu khóa	Khóa
$\emptyset$	BD	$ABCD = Q$	BD	BD
A	ABD	ABCD	ABD	
C	BCD	ABCD	BCD	
AC	ABCD	ABCD	ABCD	

## 8.6. Khóa của lược đồ quan hệ

### 8.6.4. Thuật toán cải tiến tìm tất cả các khóa

**Cho lược đồ quan hệ  $R(C, S, Z)$  và tập phụ thuộc hàm  $F = \{CS \rightarrow Z; Z \rightarrow C\}$ . Tìm tất cả các khóa thuộc tập cơ sở dữ liệu trên.**

$$TN = \{S\}$$

$$TG = \{C; Z\}$$

$x_i$	$TN \cup X_i$	$(TN \cup X_i)^+$	Siêu khóa	Khóa
$\emptyset$	S	S		
C	CS	CSZ	CS	CS
Z	SZ	CSZ	SZ	SZ
CZ	CSZ	CSZ	CSZ	

## 8.6. Khóa của lược đồ quan hệ

Cho lược đồ quan hệ  $R(A,B,C,D,E,G,H,I)$  và tập phụ thuộc hàm **8.6.5. Bài tập**

$$F = \{ AC \rightarrow B; BI \rightarrow AC; ABC \rightarrow D; H \rightarrow I; ACE \rightarrow BCG; CG \rightarrow AE \}$$

Tìm K?

# Chuẩn hóa (Normalization)

- Định nghĩa chuẩn hóa
- Các dạng chuẩn hóa

- Chuẩn hóa là kỹ thuật dùng để tạo ra một tập các quan hệ có các đặc điểm mong muốn dựa vào các yêu cầu về dữ liệu của 1 xí nghiệp
- Chuẩn hóa 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:** loại bỏ các bất thường của 1 quan hệ để có được các quan hệ có cấu trúc tốt hơn, nhỏ hơn
- **Quan hệ có cấu trúc tốt (well-structured relation):** là quan hệ có sự dư thừa dữ liệu là tối thiểu và cho phép người dùng thêm, sửa, xóa mà không gây ra mâu thuẫn dữ liệu
- Quan hệ được chuẩn hóa là quan hệ trong đó mỗi miền của một thuộc tính chỉ chứa những giá trị nguyên tố. Do đó mỗi giá trị trong quan hệ cũng là nguyên tố. Quan hệ có chứa các miền trị là không nguyên tố gọi là quan hệ không chuẩn hóa.
- Một quan hệ được chuẩn hóa có thể được tách thành nhiều quan hệ chuẩn hóa khác và không làm mất thông tin.

## Ví dụ :

MANHACC	MATHANG	
	MAMH	SOLUONG
1	100	1
	200	2
	300	1
2	100	4
	200	2
3	400	5
	500	1

Quan hệ không chuẩn hóa

MANHACC	MAMH	SOLUONG
1	100	1
	200	2
	300	1
2	100	4
	200	2
3	400	5
	500	1

Quan hệ chuẩn hóa<sup>39</sup>

- Quá trình chuẩn hóa được thực hiện qua nhiều bước.  
Mỗi bước tương ứng một dạng chuẩn
- Các dạng chuẩn:
  - Dạng chuẩn 1(1NF – first normal form)
  - Dạng chuẩn 2(2NF- second normal form)
  - Dạng chuẩn 3(3NF – third normal form)
  - Dạng chuẩn BCNF – Boyce Codd
  - Dạng chuẩn 4NF

# Bảng chưa chuẩn hóa

- Bảng không ở dạng chuẩn 1 (hay chưa chuẩn hóa) nếu nó chứa một hoặc nhiều nhóm lặp lại hoặc các giá trị phức hợp
- Nhóm lặp lại (Repeating group):** một nhóm nhiều hàng có thể có cùng chung một thuộc tính

Bảng chưa chuẩn hóa

MASV	HOVATEN	KHOA	TENMONHOC	DIEMTHI
99023	NGUYENTHITHU	CONG NGHE THONG TIN	KY THUAT LAP TRINH TOAN ROI RAC CO SO DU LIEU	6 8 4
99030	LE VAN THANH	DIEN TU	VI XULY	4

Bảng đã chuẩn hóa ở dạng chuẩn 1

MASV	HOVATEN	KHOA	TENMONHOC	DIEMTHI
99023	NGUYENTHITHU	CONG NGHE THONG TIN	KY THUAT LAP TRINH	6
99023	NGUYENTHITHU	CONG NGHE THONG TIN	TOAN ROI RAC	8
99023	NGUYENTHITHU	CONG NGHE THONG TIN	CO SO DU LIEU	4
99030	LE VAN THANH	DIEN TU	VI XULY	4

## Ví dụ :

Table name: RPT\_FORMAT      Database name: Ch05\_ConstructCo

	PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
► 15	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8	
		101	John G. News	Database Designer	\$105.00	19.4	
		105	Alice K. Johnson *	Database Designer	\$105.00	35.7	
		106	William Smithfield	Programmer	\$35.75	12.6	
		102	David H. Senior	Systems Analyst	\$96.75	23.8	
18	Amber Wave	114	Annelise Jones	Applications Designer	\$48.10	24.6	
		118	James J. Frommer	General Support	\$18.36	45.3	
		104	Anne K. Ramoras *	Systems Analyst	\$96.75	32.4	
		112	Darlene M. Smithson	DSS Analyst	\$45.95	44.0	
22	Rolling Tide	105	Alice K. Johnson	Database Designer	\$105.00	64.7	
		104	Anne K. Ramoras	Systems Analyst	\$96.75	48.4	
		113	Delbert K. Joenbrood *	Applications Designer	\$48.10	23.6	
		111	Geoff B. Wabash	Clerical Support	\$26.87	22.0	
		106	William Smithfield	Programmer	\$35.75	12.8	
25	Starflight	107	Maria D. Alonzo	Programmer	\$35.75	24.6	
		115	Travis B. Bawangi	Systems Analyst	\$96.75	45.8	
		101	John G. News *	Database Designer	\$105.00	56.3	
		114	Annelise Jones	Applications Designer	\$48.10	33.1	
		108	Ralph B. Washington	Systems Analyst	\$96.75	23.6	
		118	James J. Frommer	General Support	\$18.36	30.5	
		112	Darlene M. Smithson	DSS Analyst	\$45.95	41.4	

Repeating group

# Dạng chuẩn 1 (1NF – first normal form)

- Bảng ở dạng chuẩn 1 nếu
  - Có khóa chính
  - Không có nhóm lặp lại
- Bảng ở 1NF nếu mọi thuộc tính của R đều chứa các giá trị nguyên tố (không có thuộc tính đa trị)

**MONHOC(MaMH, TenMH)**

MaMH	TenMH
T1	Toán
AV	Anh văn

# Biến đổi về dạng chuẩn 1

- Quá trình chuẩn hóa gồm 3 bước:
  - Loại bỏ các nhóm lặp lại
  - Xác định khóa chính của bảng
  - Xác định tất cả các phụ thuộc (dependencies) trong bảng
- Lược đồ phụ thuộc (dependency diagram): để giúp mô tả tất cả các phụ thuộc trong bảng

# Ví dụ quan hệ có thuộc tính đa trị (multivalued attributes)

## Quan hệ Employee\_Course

Emp_ID	Name	Dept_Name	Salary	Course_Title	Date_Completed
100	M.Simpson	Marketing	48000	SPSS	6/19/2001
				Surveys	12/12/2002
140	A.Beeton	Accounting	52000	Tax Acc	12/8/2003
110	C.Lureco	Info System	43000	SPSS C++	1/12/2003 2/6/2004
190	L.Davis	Finance	55000		
150	S.Martin	Marketing	42000	SPSS	6/16/2002
				Java	5/7/2004

# Ví dụ quan hệ có thuộc tính đa trị (multivalued attributes)

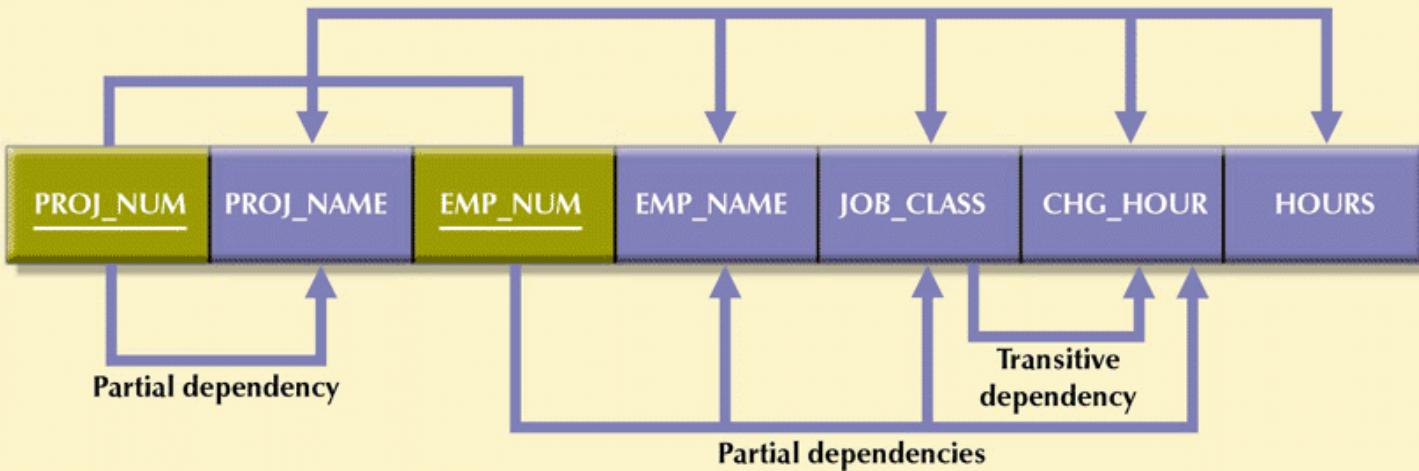
Emp_ID	Name	Dept_Name	Salary	Course_Title	Date_Completed
100	M.Simpson	Marketing	48000	SPSS	6/19/2001
100	M.Simpson	Marketing	48000	Surveys	12/12/2002
140	A.Beeton	Accounting	52000	Tax Acc	12/8/2003
110	C.Lureco	Info System	43000	SPSS	1/12/2003
110	C.Lureco	Info System	43000	C++	2/6/2004
190	L.Davis	Finance	55000		
150	S.Martin	Marketing	42000	SPSS	6/16/2002
150	S.Martin	Marketing	42000	Java	5/7/2004

→ Dạng chuẩn 1

→ Khóa là EmpID + CourseTitle

# Ví dụ quan hệ có thuộc tính đa trị (multivalued attributes)

Xác định tất cả các phụ thuộc (dependencies) trong bảng



→ Dạng chuẩn 1

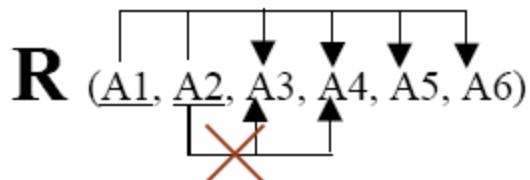
→ Khóa là EmpID + EMP\_NUM

# Dạng chuẩn 1 (1NF – Normal First Form)

- Nhận xét:
  - Dạng chuẩn 1 vẫn có thể có các bất thường khi cập nhật
- Ví dụ: trong lược đồ Employee\_Course, sẽ có các bất thường sau:
  - Thêm 1 nhân viên mới chưa tham gia khóa học nào → vi phạm quy luật bảo toàn thực thể
  - Thay đổi tên phòng → phải thay đổi hàng loạt thông tin này cho tất cả các nhân viên của phòng đó
  - Xóa 1 course mà chỉ có 1 nhân viên học, thông tin course sẽ bị xóa theo

# Phụ thuộc hàm đầy đủ (Full functional dependency)

- $X \rightarrow A$  là phụ thuộc hàm đầy đủ nếu không tồn tại  $Y \subset X$  để cho  $Y \rightarrow A$
- Sơ đồ mô tả



→ Vi phạm chuẩn 2

– Ví dụ 2: Cho  $Q(ABC)$  và  $F = \{ A \rightarrow B; A \rightarrow C; AB \rightarrow C \}$

–  $A \rightarrow B$ :  $A \rightarrow C$  là các phụ thuộc hàm đầy đủ.

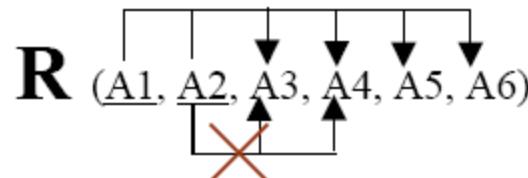
–  $AB \rightarrow C$  không là phụ thuộc hàm đầy đủ vì có  $A \rightarrow C$ .

*Chú ý rằng, một phụ thuộc hàm mà vẽ trái chỉ có một thuộc tính là phụ thuộc hàm đầy đủ*

# Phụ thuộc hàm đầy đủ

## (Full functional dependency)

- $X \rightarrow A$  là **phụ thuộc hàm đầy đủ** nếu không tồn tại  $Y \subset X$  để cho  $Y \rightarrow A$
- Sơ đồ mô tả



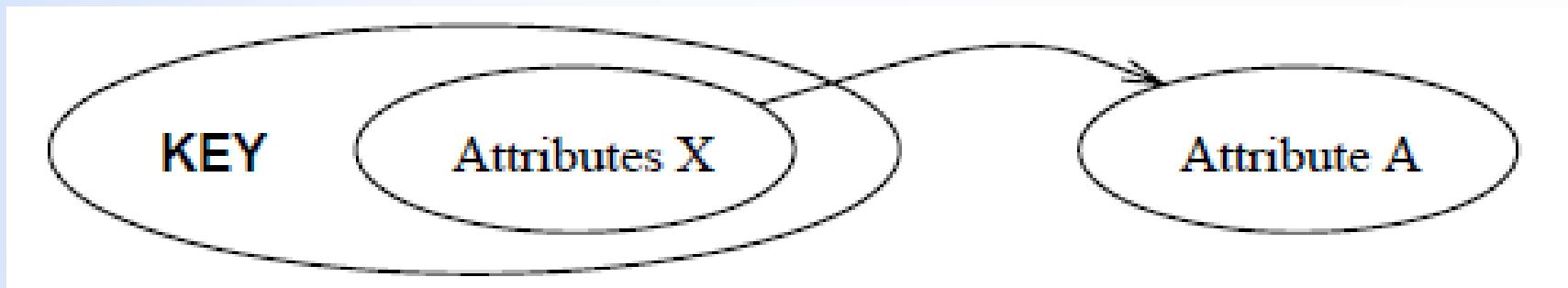
- Ví dụ 3: quan hệ I
  - Khóa là Emp\_ID, Course
  - $\text{Emp\_ID}, \text{Course} \rightarrow \text{Grade}$  là phụ thuộc hàm đầy đủ
  - $\text{Emp\_ID} \rightarrow \text{Name}, \text{Dept\_Name}$  là phụ thuộc hàm đầy đủ
  - $\text{Emp\_ID}, \text{Course} \rightarrow \text{Name}, \text{Dept\_Name}$  là phụ thuộc hàm không đầy đủ

$\text{Emp\_ID} \rightarrow \text{Name}, \text{Dept\_Name}$

$\text{Emp\_ID} \subset \{\text{Emp\_ID}, \text{Course}\}$

# Phụ thuộc hàm đầy đủ (Full functional dependency)

- Phụ thuộc hàm riêng phần (partial FD)  
 $X \rightarrow A$ , tồn tại  $Y \subset X$  sao cho  $Y \rightarrow A$



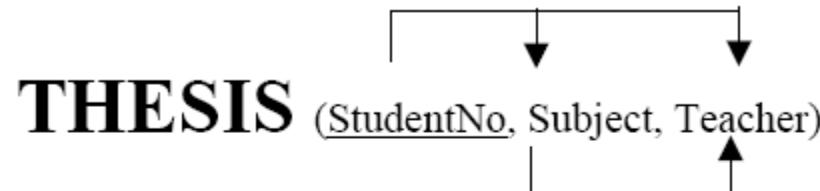
## Ví dụ 1:

*customer-name, loan-number → customer-name*

*customer-name → customer-name*

# Dạng chuẩn 2 (2NF – second Normal Form)

- Lược đồ quan hệ R ở dạng 2NF đối với tập phụ thuộc hàm F nếu:
  - R ở dạng chuẩn 1
  - Mọi thuộc tính không khóa đều phụ thuộc đầy đủ vào mọi khóa của R
- Nếu quan hệ R chỉ có các khóa đơn thì đương nhiên quan hệ này ở dạng chuẩn 2

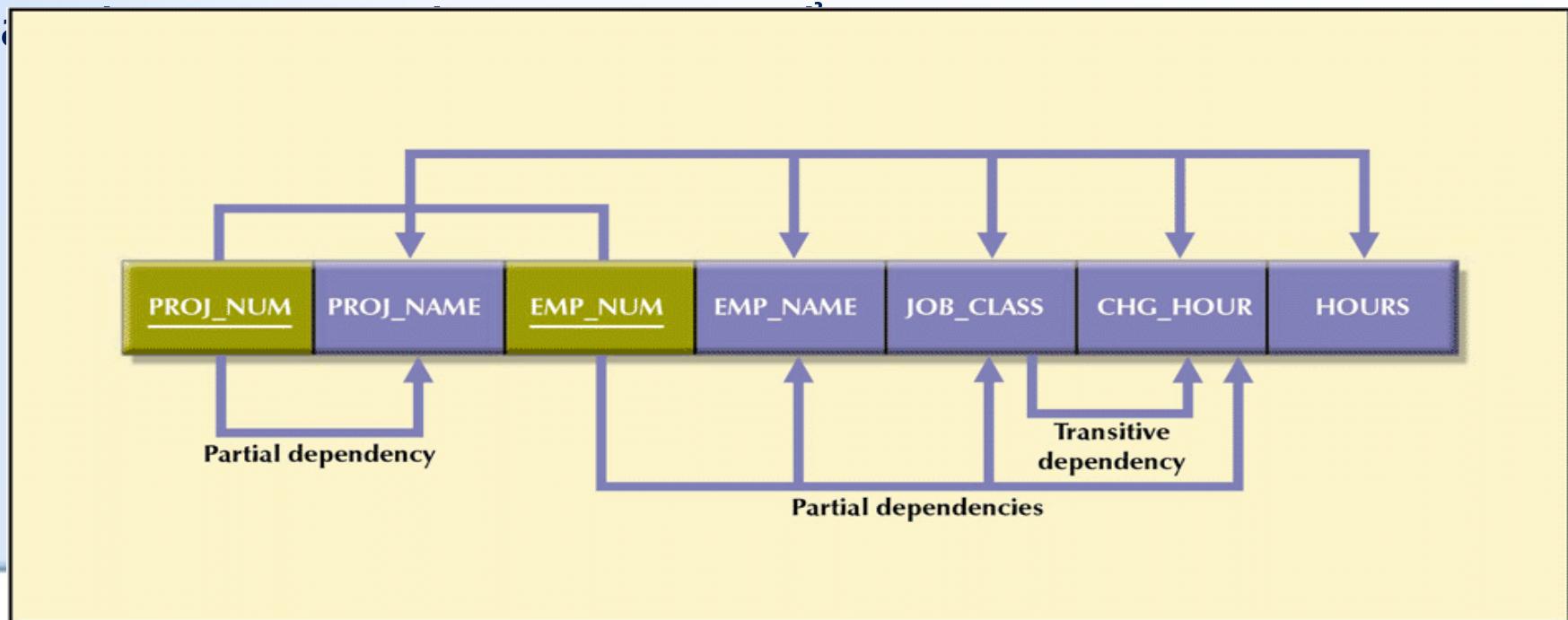


<u>StudentNo</u>	Subject	Teacher
SV01	1	Nguyễn Văn Hiệu
SV02	2	Ngô Lan Phương
SV03	1	Nguyễn Văn Hiệu
SV04	1	Nguyễn Văn Hiệu

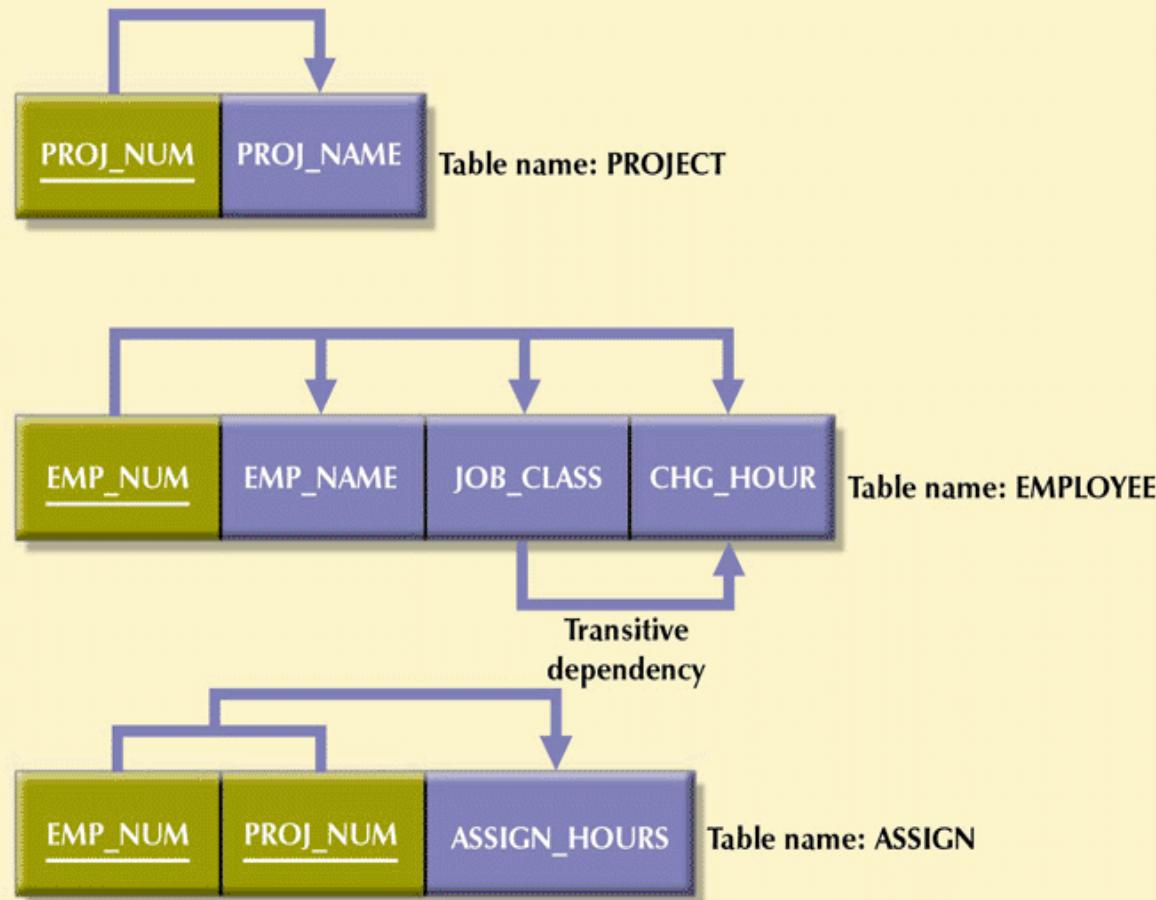
# Biến đổi thành 2NF

- Loại bỏ các phụ thuộc hàm riêng phần và tạo thêm các quan hệ mới tương ứng với các phụ thuộc hàm riêng phần

Qua



# Biến đổi thành 2NF



# Dạng chuẩn 2

- Quan hệ ở 2NF vẫn có thể có các bất thường khi cập nhật
- Ví dụ: xét quan hệ EMPLOYEE đã ở chuẩn 2NF
  - Khi thêm 1 loại công việc mới mà công việc này chưa có nhân viên nào làm sẽ vi phạm ràng buộc khoá chính
  - Khi sửa đổi lương giờ (CHR\_HOUR) của 1 loại công việc mà có nhiều nhân viên đang cùng làm
  - Khi xoá 1 nhân viên đang làm công việc mà chỉ có nhân viên đó làm thì sẽ làm mất luôn thông tin về công việc đó

## Thuật toán kiểm tra dạng chuẩn 2

- Vào: lược đồ quan hệ Q, tập phụ thuộc hàm F
- Ra: khẳng định Q đạt chuẩn 2 hay không đạt chuẩn 2.
- **Bước 1:** Tìm tất cả khóa của Q
- **Bước 2:** Với mỗi khóa K, tìm bao đóng của tất cả tập con thật sự S của K.
- **Bước 3:** Nếu có bao đóng S+ chứa thuộc tính không khóa thì Q không đạt chuẩn 2. Ngược lại thì Q đạt chuẩn 2

# Dạng chuẩn 2

- Ví dụ 1: Cho lược đồ quan hệ  $Q(A,B,C,D)$  và tập phụ thuộc hàm  $F=\{AB\rightarrow C; B\rightarrow D; BC\rightarrow A\}$ . Hỏi  $Q$  có đạt chuẩn 2 không?
-

# Dạng chuẩn 2

- Quan hệ sau đạt chuẩn 2.

$Q(G,M,V,N,H,P) \quad F = \{G \rightarrow M; G \rightarrow N; G \rightarrow H; G \rightarrow P; M \rightarrow V; NHP \rightarrow M\}$

- Giải:

$TN = \{G\} \quad TG = \{M, N, H, P\}$

# Dạng chuẩn 2

- Hệ quả:
  - Nếu Q đạt chuẩn 1 và tập thuộc tính không khóa của Q bằng rỗng thì Q đạt chuẩn 2
  - Nếu tất cả khóa của quan hệ chỉ gồm một thuộc tính thì quan hệ đó ít nhất đạt chuẩn 2.
- Ví dụ 4: Q(A,B,C,D,E,H) F={A → E; C → D; E → DH}
  - Giải:

# Phụ thuộc bắc cầu (Transitive dependency)

- Q là lược đồ quan hệ, X,Y là hai tập con của Q+, A là một thuộc tính. Nói rằng A phụ thuộc bắc cầu vào X nếu cả ba điều sau thỏa:  $X \rightarrow A$  được gọi là phụ thuộc bắc cầu nếu tồn tại Y để cho

$$X \rightarrow Y, Y \rightarrow A,$$

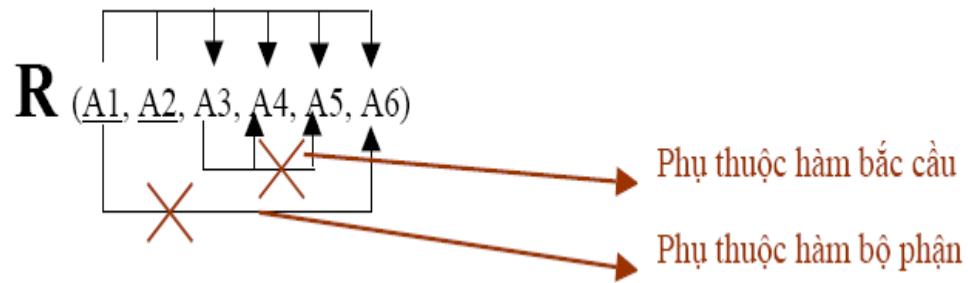
$$Y \rightarrow X /$$

Và  $A \notin XY$

- Nguyên nhân gây ra các bất thường khi cập nhật bảng 2NF là do có các thuộc tính không khóa phụ thuộc bắc cầu vào khóa của quan hệ

# Dạng chuẩn 3 (3NF – third normal form)

- Định nghĩa 1: Lược đồ quan hệ R ở 3NF đối với tập phụ thuộc hàm F nếu:
  - R ở dạng 2NF
  - Mọi thuộc tính không khóa đều không phụ thuộc bắc cầu vào khóa chính của R
- Định nghĩa 2: Lược đồ quan hệ R ở 3NF đối với tập phụ thuộc hàm F nếu R ở dạng chuẩn 1 và mọi phụ thuộc hàm  $X \rightarrow A$  với  $A \notin X$  thì X là 1 siêu khoá của R hoặc A là 1 thuộc tính khoá
- Biểu diễn bằng sơ đồ



# Dạng chuẩn 3

- Quan hệ ở 3NF vẫn có thể có các bất thường khi cập nhật
  - Ví dụ: xét lược đồ quan hệ EMPLOYEE\_TEACHER(EmpId, Course, Teacher)  
Có 2 phụ thuộc hàm:
    - EmpId, Course → Teacher
    - Teacher → Course
- ➔ Thuộc dạng 3NF, bất thường xảy ra teacher thay đổi môn dạy

- Hệ quả
  - Hệ quả 1: Nếu Q đạt chuẩn 3 thì Q đạt chuẩn 2
  - Hệ quả 2: Nếu Q không có thuộc tính không khóa thì Q đạt chuẩn 3.
- Định lý
  - Q là lược đồ quan hệ
  - F là tập các phụ thuộc hàm có vẽ phải một thuộc tính.
  - Q đạt chuẩn 3 nếu và chỉ nếu mọi phụ thuộc hàm  $X \rightarrow A \in F$  với  $A \notin X$  đều có X là siêu khóa hay A là thuộc tính khóa

## **Thuật toán kiểm tra dạng chuẩn 3**

- Vào: lược đồ quan hệ Q, tập phụ thuộc hàm F
- Ra: khẳng định Q đạt chuẩn 3 hay không đạt chuẩn 3.
- **Bước 1:** Tìm tất cả khóa của Q
- **Bước 2:** Từ F tạo tập phụ thuộc hàm tương đương  $F_{1tt}$  có vẽ phải một thuộc tính.
- **Bước 3:** Nếu mọi phụ thuộc hàm  $X \rightarrow A \in F_{1tt}$  với  $A \notin X$  đều có X là siêu khóa hoặc A là thuộc tính khoá thì Q đạt chuẩn 3 ngược lại Q không đạt chuẩn 3

# Dạng chuẩn 3

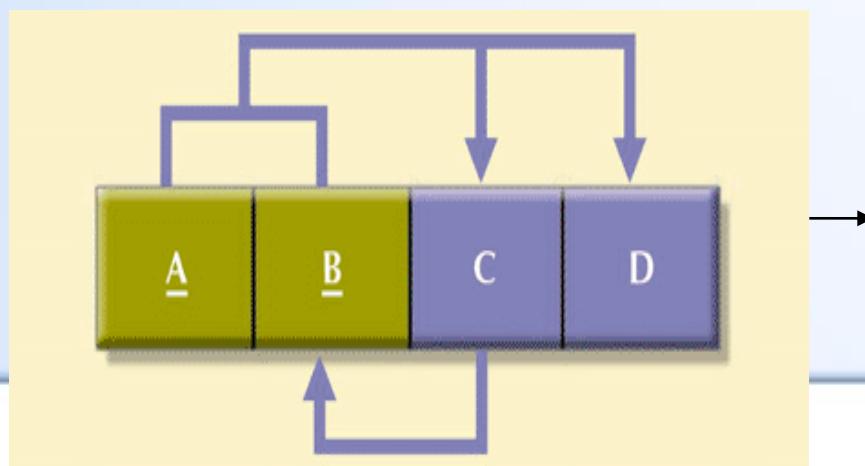
- Ví dụ 5: Cho lược đồ quan hệ  $Q(A,B,C,D)$   $F=\{AB \rightarrow C; D \rightarrow B; C \rightarrow ABD\}$ . Hỏi  $Q$  có đạt chuẩn 3 không?
- Giải:  $TN = \emptyset$   $TG = \{ABCD\}$

$X_i$	$(TN \cup X_i)$	$(TN \cup X_i)^+$	Siêu khóa	khóa
$\emptyset$	$\emptyset$	$\emptyset$		
A	A	A		
B	B	B		
AB	AB	ABCD	AB	AB
C	C	ABCD	C	C
AC	AC	ABCD	AC	
BC	BC	ABCD	BC	
ABC	ABC	ABCD	ABC	
D	D	BD		
AD	AD	ABCD	AD	AD
BD	BD	BD		
ABD	ABD	ABCD	ABD	
CD	CD	ABCD	CD	
ACD	ACD	ABCD	ACD	

$K_1 = \{AB\}; K_2 = \{AD\}; K_3 = \{C\}$  là các khóa  $\Rightarrow$  mọi phụ thuộc hàm  $X \rightarrow A \in F$  đều có  $A^6$  là  $^{65}$ thuộc tính khóa. Vậy  $Q$  đạt chuẩn 3

# Dạng chuẩn Boyce-Codd (BCNF)

- Một quan hệ ở dạng BCNF nếu mọi determinant (định thuộc) đều là candidate key
- Cho 1 lược đồ quan hệ  $R(U,F)$  với  $U$  là tập thuộc tính,  $F$  là tập phụ thuộc hàm. Lược đồ ở dạng chuẩn BCNF nếu với mỗi phụ thuộc hàm  $X \rightarrow Y \in F$  nếu 1 trong 2 điều kiện sau là đúng:
  - $Y \subseteq X$  (phụ thuộc hàm tầm thường)
  - $X$  là siêu khóa của  $R$



Quan hệ này đạt chuẩn 3NF  
nhưng không đạt chuẩn BCNF

# Dạng chuẩn Boyce-Codd (BCNF)

- Hệ quả
  - Hệ quả 1: Nếu Q đạt chuẩn BC thì Q đạt chuẩn 3 (hiển nhiên do định nghĩa)
  - Hệ quả 2: Mỗi lược đồ có hai thuộc tính đều đạt chuẩn BC (xét phụ thuộc hàm có thể có của Q )
- Định lý
  - *Q là lược đồ quan hệ*
  - *F là tập các phụ thuộc hàm có vẽ phải một thuộc tính.*
  - *Q đạt chuẩn BC nếu và chỉ nếu mọi phụ thuộc hàm  $X \rightarrow A$  với  $A \notin X$  đều có  $X$  là siêu khóa*

## **Thuật toán kiểm tra dạng chuẩn BC**

- *Vào: lược đồ quan hệ Q, tập phụ thuộc hàm F*
- *Ra: khẳng định Q đạt chuẩn BC hay không đạt chuẩn BC.*
- **Bước 1:** Tìm tất cả khóa của Q
- **Bước 2:** Từ F tạo tập phụ thuộc hàm tương đương  $F_{1\text{tt}}$  có vẽ phải một thuộc tính
- **Bước 3:** Nếu mọi phụ thuộc hàm  $X \rightarrow A \in F_{1\text{tt}}$  với  $A \notin X$  đều có X là siêu khóa thì Q đạt chuẩn BC ngược lại Q không đạt chuẩn BC

# Dạng chuẩn Boyce-Codd (BCNF)

- Ví dụ: Q(A,B,C,D,E,I)  $F=\{ACD \rightarrow EBI; CE \rightarrow AD\}$ . Hỏi Q có đạt chuẩn BC không?
- C

# Dạng chuẩn Boyce-Codd (BCNF)

- Ví dụ 8:  $Q(SV, MH, THAY)F = \{SV, MH \rightarrow THAY; THAY \rightarrow MH\}$   
Quan hệ trên đạt chuẩn 3 nhưng không đạt chuẩn BC..
- Ví dụ 9:  
Chẳng hạn cho  $Q(A, B, C, D)$  và  $F=\{AB \rightarrow C; D \rightarrow B; C \rightarrow ABD\}$   
thì  $Q$  là 3NF nhưng không là BCNF  
Nếu  $F=\{B \rightarrow D, A \rightarrow C, C \rightarrow ABD\}$  là 2 NF nhưng không là 3 NF

# Chuyển đổi thành BCNF

- Một quan hệ ở BCNF thì nó cũng ở dạng 3NF
- Có thể biến đổi trực tiếp bảng từ 1NF thành BCNF, mà không cần phải qua các bước chuẩn hóa 2NF, 3NF
  - Loại bỏ các định thuộc không phải là siêu khoá
  - Tạo các quan hệ mới tương ứng với các định thuộc sao cho định thuộc trở thành siêu khoá của quan hệ mới

# So sánh 3NF và BCNF

- BCNF được xem là trường hợp đặc biệt của 3NF
- Với quan hệ có nhiều candidate key phức hợp thì BCNF sẽ tránh được hai bất thường có thể xảy ra ở 3NF
  - 1 phần của khóa xác định 1 phần của khóa khác
  - Cột không khóa xác định 1 phần của khóa

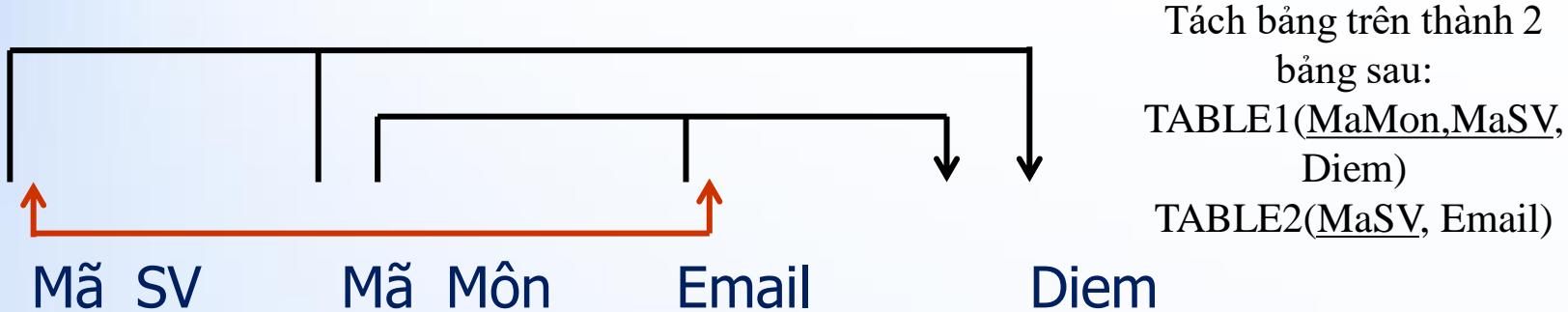
# Candidate key và BCNF

- Một quan niệm sai lầm khi cho rằng một bảng với nhiều candidate key sẽ vi phạm chuẩn BCNF.
- Nhiều candidate key không vi phạm BCNF hay 3NF, không cần phải phân chia bảng chỉ vì nó có nhiều candidate key

# Candidate key và BCNF

## Ví dụ

- Xét lược đồ phụ thuộc sau:



- Hai candidate key: Ma\_SV+Ma\_Mon; Email+Ma\_Mon
- Chỉ có 1 thuộc tính không khóa là Diem
- Bất thường 1:** 1 phần của khóa này xác định 1 phần của khóa khác.
- Bảng thuộc 3NF nhưng không là BCNF
- Làm thế nào để chuẩn hóa thành BCNF???

# Thuật toán kiểm tra dạng chuẩn của một lược đồ quan hệ

- Vào: lược đồ quan hệ Q, tập phụ thuộc hàm F
- Ra: khẳng định Q đạt chuẩn gì?
- **Bước 1:** Tìm tất cả khóa của Q
- **Bước 2:** Kiểm tra chuẩn BC nếu đúng thì Q đạt chuẩn BC, kết thúc thuật toán ngược lại qua bước 3
- **Bước 3:** Kiểm tra chuẩn 3 nếu đúng thì Q đạt chuẩn 3, kết thúc thuật toán ngược lại qua bước 4
- **Bước 4:** Kiểm tra chuẩn 2 nếu đúng thì Q đạt chuẩn 2, kết thúc thuật toán ngược lại Q đạt chuẩn 1

**Định nghĩa: Dạng chuẩn của một lược đồ cơ sở dữ liệu là dạng chuẩn thấp nhất trong các dạng chuẩn của các lược đồ quan hệ con.**

1. Cho biết dạng chuẩn cao nhất của các LDQH sau:

- a) Q(ABCDEG) F = {A ->BC, C->DE, E->G}
- b) Q(ABCDEFGH) F = {C->AB, D->E, B->G}
- c) Q(ABCDEFGH) F = {A->BC, D->E, H->G}
- d) Q(ABCDEG) F = {AB->C, C->B, ABD->E, G->A}
- e) Q(ABCDEFGHI) F = {AC->B, BI->ACD, ABC->D  
H->I, ACE->BCG, CG->A}

2. Cho Q(CDEGHK) và F = {CK->H, C->D, E->C,  
E->G, CK->E}

- a) Chứng minh EK->DH
- b) Tìm tất cả các khóa của Q
- c) Xác định dạng chuẩn cao nhất của Q

