



**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

# THIẾT KẾ CSDL QUAN HỆ

GV: Đỗ Bá Lâm

Email: lamdb@soict.hust.edu.vn

Viện Công nghệ thông tin và Truyền thông,  
Trường Đại học Bách Khoa Hà Nội

# Nội dung

Mục đích của chuẩn hoá là gì?



Thế nào là chuẩn hóa?  
Có bao nhiêu dạng chuẩn?

# Ví dụ

- Cho 1 CSDL về các hãng cung ứng.
  - Suppliers(sid, sname, city, Number of Employee, product, quantity)
- Các vấn đề đặt ra
  - Dư thừa dữ liệu
  - Không nhất quán
  - Dị thường khi thêm/sửa bộ

Sid	Sname	City	NOE	Product	quantity
S1	Smith	London	100	Screw	50
S1	Smith	London	100	Nut	100
S2	J&J	Paris	100	Screw	78
S3	Blake	Tokyo	75	Bolt	100

# Mục đích của chuẩn hoá

- Xác định được một tập các lược đồ quan hệ cho phép tìm kiếm thông tin một cách dễ dàng, đồng thời tránh được dư thừa dữ liệu
- Hướng tiếp cận:
  - Tách các lược đồ quan hệ “có vấn đề” thành những lược đồ quan hệ “chuẩn hơn”

# Ví dụ

Suppliers(sid, sname, city, NOE, product, quantity)

Sid	Sname	City	NOE	Product	quantity
S1	Smith	London	100	Screw	50
S1	Smith	London	100	Nut	100
S2	J&J	Paris	100	Screw	78
S3	Blake	Tokyo	75	Bolt	100



Suppliers(sid, sname, city, noe)

Supply(sid, product, quantity)

# Phụ thuộc hàm (*Functional dependencies - FD*)

- Cho

- $R(U)$  là một sơ đồ quan hệ với  $U$  là tập các thuộc tính
- $X, Y \subseteq U$

- **$X$  xác định hàm  $Y$  hay  $Y$  phụ thuộc hàm vào  $X$  nếu**

- $\forall$  quan hệ  $r$  xác định trên  $R(U)$  và với 2 bộ  $t_1$  và  $t_2$  bất kỳ mà  $t_1[X] = t_2[X]$  thì  $t_1[Y] = t_2[Y]$
- Ký hiệu:  $X \rightarrow Y$

# Ví dụ

Suppliers(sid, sname, city, NOE, product, quantity, price, amount)

sid → sname

sid → city

sid → NOE

sid, product → quantity

product → price

quantity, price → amount

# Hệ tiên đề Amstrong

- Cho
  - $R(U)$  là một sơ đồ quan hệ,  $U$  là tập các thuộc tính
  - $X, Y, Z, W \subseteq U$
  - Ký hiệu:  $XY = X \cup Y$
- **Phản xạ** (*reflexivity*)  
Nếu  $Y \subseteq X$  thì  $X \rightarrow Y$
- **Tăng trưởng** (*augmentation*)  
Nếu  $X \rightarrow Y$  thì  $XZ \rightarrow YZ$
- **Bắc cầu** (*transitivity*)  
Nếu  $X \rightarrow Y, Y \rightarrow Z$  thì  $X \rightarrow Z$



# Hệ quả

- Luật hợp (*union*)  
Nếu  $X \rightarrow Y$ ,  $X \rightarrow Z$  thì  $X \rightarrow YZ$
- Luật tựa bắc cầu (*pseudotransitivity*)  
Nếu  $X \rightarrow Y$ ,  $WY \rightarrow Z$  thì  $XW \rightarrow Z$
- Luật tách (*decomposition*)  
Nếu  $X \rightarrow Y$ ,  $Z \subseteq Y$  thì  $X \rightarrow Z$

# Bao đóng của tập phụ thuộc hàm

- Định nghĩa: Bao đóng của tập phụ thuộc hàm  $F$ , ký hiệu là  $F^+$  là **tập hợp tất cả** các phụ thuộc hàm có thể được suy diễn ra từ  $F$  dựa trên các tiên đề Amstrong.
- $F \subseteq F^+$
- $F$  là **họ đầy đủ** (*full family*) nếu  $F = F^+$
- Ví dụ:
  - Cho lược đồ quan hệ  $Q(ABCDEFGH)$  và  $F = \{B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D\}$
  - Khi đó  $F^+ = \{B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D; BC \rightarrow AC; BC \rightarrow D; DA \rightarrow AH; \dots\}$

# Bao đóng của tập thuộc tính

- Định nghĩa: Bao đóng của tập thuộc tính  $X$ , ký hiệu là  $X^+$  là tập hợp tất cả các thuộc tính được xác định hàm bởi  $X$  thông qua tập  $F$ .
- $X^+ = \{A \in U \mid X \rightarrow A \in F^+\}$
- Ví dụ:
  - Cho lược đồ quan hệ  $R = (U, F)$  với  $U = \{A, B, C, D, E, G, H\}$  và  $F = \{AB \rightarrow C, D \rightarrow EG, ACD \rightarrow B, C \rightarrow A, BE \rightarrow C, CE \rightarrow AG, BC \rightarrow D, CG \rightarrow BD, G \rightarrow H\}$
  - $(D)^+ = \{D, E, G, H\}$
  - $(BE)^+ = \{A, B, C, D, E, G, H\}$

# Khoá tối thiểu

- Định nghĩa: Cho lược đồ quan hệ  $R(U)$ , tập các phụ thuộc hàm  $F$ ,  $K \subseteq U$ .  $K$  được gọi là **khóa tối thiểu** của  $R$  nếu như
  - $K \rightarrow U \in F^+$
  - Với  $\forall K' \subset K$  thì  $K' \rightarrow U \notin F^+$
- Nhận xét
  - $K^+ = U$
  - $K$  là tập thuộc tính nhỏ nhất có tính chất như vậy

# Tìm bao đóng của tập thuộc tính

- **Vào:** Tập thuộc tính  $U$ , tập phụ thuộc hàm  $F$  trên  $U$ ,  $X \subseteq U$

- **Ra:**  $X^+$

- **Thuật toán**

$B^0 \quad X^0 = X.$

$B^i$  Tính  $X^i$  từ  $X^{i-1}$

Nếu  $\exists Y \rightarrow Z \in F$ , và  $Y \subseteq X^{i-1}$  và  $A \in Z$  và  $A \notin X^{i-1}$

thì  $X^i = X^{i-1} \cup A$

ngược lại  $X^i = X^{i-1}$ .

Kiểm tra nếu  $X^i \neq X^{i-1}$  thì thực hiện  $B^{i+1}$

ngược lại dừng,  $X^+ = X^i$

- **Ý nghĩa:**

- Kiểm tra lần lượt từng phụ thuộc hàm  $f_i = \alpha \rightarrow \beta$

- nếu  $\alpha \subseteq X^+$  thì kết nạp vế phải (tức  $\beta$ ) vào vào  $X^+$ :  $X^+ = X^+ \cup \beta$

- Lặp lại cho đến khi nào  $X^+ = \text{Const}$

# Tính bao đóng của tập thuộc tính (ví dụ)

- Cho  $R(U)$  ,  $U = \{A, B, C, D, E, F\}$   
 $F = \{AB \rightarrow C, BC \rightarrow AD, D \rightarrow E, CF \rightarrow B\}$   
Tính  $(AB)^+$
- Thực hiện:
  - Bước 0:  $X^0 = AB$
  - Bước 1:  $X^1 = ABC$  ( do  $AB \rightarrow C$  )
  - Bước 2:  $X^2 = ABCD$  (do  $BC \rightarrow AD$ )
  - Bước 3:  $X^3 = ABCDE$  (do  $D \rightarrow E$ )
  - Bước 4:  $X^4 = ABCDE$

# Giải thuật tìm khoá tối thiểu – GT1

- **Vào:** Tập thuộc tính  $U = \{A_1, A_2, \dots, A_n\}$ , tập phụ thuộc hàm  $F$
- **Ra:** Khóa tối thiểu  $K$  xác định được trên  $U$  và  $F$
- **Thuật toán**

**$B^0$**       $K^0 = U, n = |U|$

**$B^i$**      Nếu  $(K^{i-1} \setminus \{A_i\}) \rightarrow U$

thì                 $K^i = K^{i-1} \setminus \{A_i\}$

ngược lại,  $K^i = K^{i-1}$

**$B^n$**       $K = K^i$

# Ví dụ

- Cho  $R(U)$  trong đó  $U = \{A, B, C, D, E, F, G\}$ ,  $F = \{A \rightarrow B, ACD \rightarrow E, EF \rightarrow G\}$
- Tìm một khóa tối thiểu của  $R$
- Thực hiện:

$$K^0 = ABCDEFG$$

$$K^1 = K^0 \text{ do nếu loại } A \text{ thì } (BCDEFG)^+ \neq U$$

$$K^2 = K^1 \setminus \{B\} = ACDEFG \text{ do } (ACDEFG)^+ \equiv U$$

$$K^3 = K^2 \text{ do nếu loại } C \text{ thì } (ADEFG)^+ \neq U$$

$$K^4 = K^3 \text{ do nếu loại } D \text{ thì } (ACEFG)^+ \neq U$$

$$K^5 = K^4 \setminus \{E\} = ACDFG \text{ do } (ACDFG)^+ \equiv U$$

$$K^6 = K^5 \text{ do nếu loại } F \text{ thì } (ACDG)^+ \neq U$$

$$K^7 = K^6 \setminus \{G\} = ACDF \text{ do } (ACDF)^+ \equiv U$$

Vậy khóa tối thiểu cần tìm là ACDF



# Giải thuật tìm khoá tối thiểu – GT2

- **Vào:**  $U = \{A_1, A_2, \dots, A_n\}$ ,  $F$
- **Ra:** Khóa tối thiểu  $K$  xác định được trên  $U$  và  $F$

- **Thuật toán**

**B<sup>0</sup>** VT = tập các thuộc tính nằm ở vế trái các phụ thuộc hàm

VP = tập các thuộc tính nằm ở vế phải của các phụ thuộc hàm

**X = U \ VP:** tập các thuộc tính **chắc chắn nằm trong K**

**Y = VP \ VT:** tập các thuộc tính **chắc chắn không nằm trong K**

**Z = VP ∩ VT:** tập các thuộc tính **có thể nằm trong K**

**B<sup>2</sup>** Nếu  $(X)^+ = U$  thì  $X$  là khóa tối thiểu:  $K = X$ . Kết thúc

**B<sup>3</sup>** Nếu  $(X)^+ \neq U$  thì

$$K^0 = X \cup Z$$

Lặp: lần lượt thử loại bỏ các thuộc tính trong  $Z$ ,  $K = K^i$

# Nhận xét về phụ thuộc hàm

- Từ một tập phụ thuộc hàm có thể suy diễn ra các phụ thuộc hàm khác
  - Trong một tập phụ thuộc hàm cho sẵn có thể có các phụ thuộc hàm bị coi là dư thừa.
- Làm thế nào để có được một tập phụ thuộc hàm tốt?

# Tập phụ thuộc hàm tương đương

- Định nghĩa: Hai tập phụ thuộc hàm  $F$  và  $G$  là tương đương nếu  $F^+ = G^+$ . Kí hiệu  $F \approx G$
- Nếu  $F, G$  tương đương thì ta nói  $F$  phủ  $G$  hay  $G$  phủ  $F$ .
- Kiểm tra tính tương đương của 2 tập phụ thuộc hàm
  - B1. Nếu với  $\forall f \in F, f \in G^+$  thì  $F^+ \subseteq G^+$
  - B2. Tương tự, nếu  $\forall f \in G, f \in F^+$  thì  $G^+ \subseteq F^+$
  - B3. Nếu  $F^+ \subseteq G^+$  và  $G^+ \subseteq F^+$  thì  $F \approx G$

# Tập phụ thuộc hàm không dư thừa

- Định nghĩa: Tập phụ thuộc hàm  $F$  là **không dư thừa** nếu  $\nexists X \rightarrow Y \in F$  sao cho  $F \setminus \{X \rightarrow Y\} \approx F$ .
- **Tìm phủ không dư thừa của tập phụ thuộc hàm**
  - Vào: Tập thuộc tính  $U$ ,  $F = \{L_i \rightarrow R_i : i = 1..n\}$
  - Ra : Phủ không dư thừa  $F'$  của  $F$
  - **$B^0$**   $F^0 = F$ ,  $n = |F|$
  - **$B^i$**  Nếu  $F^{i-1} \setminus \{L_i \rightarrow R_i\} \approx F^{i-1}$   
thì  $F^i = F^{i-1} \setminus \{L_i \rightarrow R_i\}$   
ngược lại,  $F^i = F^{i-1}$
  - **$B^n$**   $F' = F^i$

# Phủ tối thiểu của tập phụ thuộc hàm

- Định nghĩa:  $F_c$  được gọi là **phủ tối thiểu** của 1 tập phụ thuộc hàm  $F$  nếu thỏa mãn 3 điều kiện sau:

**Điều kiện 1:** Với  $\forall f \in F_c$ ,  $f$  có dạng  $X \rightarrow A$ , trong đó  $A$  là 1 thuộc tính

**Điều kiện 2:** Với  $\forall f = X \rightarrow Y \in F_c$ ,  $\exists A \in X$  ( $A$  là 1 thuộc tính):  
 $(F_c \setminus f) \cup \{(X \setminus A) \rightarrow Y\} \approx F_c$

**Điều kiện 3:**  $\exists X \rightarrow A \in F_c : F_c \setminus \{X \rightarrow A\} \approx F_c$

# Tìm phủ tối thiểu

- **Vào:** Tập thuộc tính  $U$ ,  $F = \{L_i \rightarrow R_i: i = 1..n\}$
- **Ra:** Phủ tối thiểu  $F_c$  của tập phụ thuộc hàm  $F$
- **Thuật toán**

## B1. Tách phụ thuộc hàm.

Biến đổi  $F$  về dạng  $F_1 = \{L_i \rightarrow A_j\}$  với  $A_j$  là một thuộc tính để đáp ứng điều kiện 1

## B2. Loại bỏ thuộc tính thừa trong vế trái của các phụ thuộc hàm

Lần lượt giản ước từng thuộc tính trong vế trái của từng phụ thuộc hàm trong  $F_1$  thu được  $F_1'$ . Nếu  $F_1' \approx F_1$  thì loại bỏ thuộc tính đang xét.

Thu được  $F_2$  thỏa mãn điều kiện 2

## B3. Loại bỏ phụ thuộc hàm dư thừa

Lần lượt kiểm tra từng phụ thuộc hàm  $f$ . Nếu  $F_2 \setminus f \approx F_2$  thì loại bỏ  $f$

Thu được  $F_3$  thỏa mãn điều kiện 3

## B4. $F_c = F_3$

# Tìm phủ tối thiểu (ví dụ)

- $U = \{A, B, C\}$ ,  $F = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$

- Tìm phủ tối thiểu của  $F$ ?

- **B1.**  $F_1 = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, AB \rightarrow C\}$

- **B2.** Loại bỏ thuộc tính dư thừa trong vế trái

Xét các phụ thuộc hàm trong  $F_1$  mà vế trái có nhiều hơn một thuộc tính. Đó là  $AB \rightarrow C$ .  
Giả sử lược  $A$  thì ta có  $B \rightarrow C$  có trong  $F_1 \Rightarrow A$  là thuộc tính thừa.  $F_2 = \{A \rightarrow B, A \rightarrow C, B \rightarrow C\}$

- **B3.** Loại bỏ phụ thuộc hàm dư thừa

- Giả sử loại  $A \rightarrow B$ ,  $F'_2 = \{A \rightarrow C, B \rightarrow C\}$ ,  $A^+_{F'_2} = \{A, C\} \Rightarrow A \rightarrow B \notin F'_2+$ . Phụ thuộc hàm này không dư thừa

- Giả sử loại  $A \rightarrow C$ :  $F'_2 = \{A \rightarrow B, B \rightarrow C\}$ ,  $A^+_{F'_2} = \{A, B, C\}$   $A \rightarrow C \in F'_2+$ . Phụ thuộc hàm này dư thừa, loại  $A \rightarrow C$

- Giả sử loại  $B \rightarrow C$ ,  $F'_2 = \{A \rightarrow B\}$ ,  $B^+_{F'_2} = \{B\} \Rightarrow B \rightarrow C \notin F'_2+$ . Phụ thuộc hàm này không dư thừa.  $F_3 = \{A \rightarrow B, B \rightarrow C\}$

- **B4.**  $F_c = \{A \rightarrow B, B \rightarrow C\}$

# Phép tách các lược đồ quan hệ

- Mục đích

- Thay thế một sơ đồ quan hệ  $R(A_1, A_2, \dots, A_n)$  bằng một tập các sơ đồ con  $\{R_1, R_2, \dots, R_k\}$  sao cho có thể tạo lại  $R$  từ phép toán kết nối tự nhiên các quan hệ đã tách, tức  $R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_k$

- Không phải mọi phép tách đều tốt. Ví dụ

Tách *employee*(*ID*, *name*, *street*, *city*, *salary*) thành

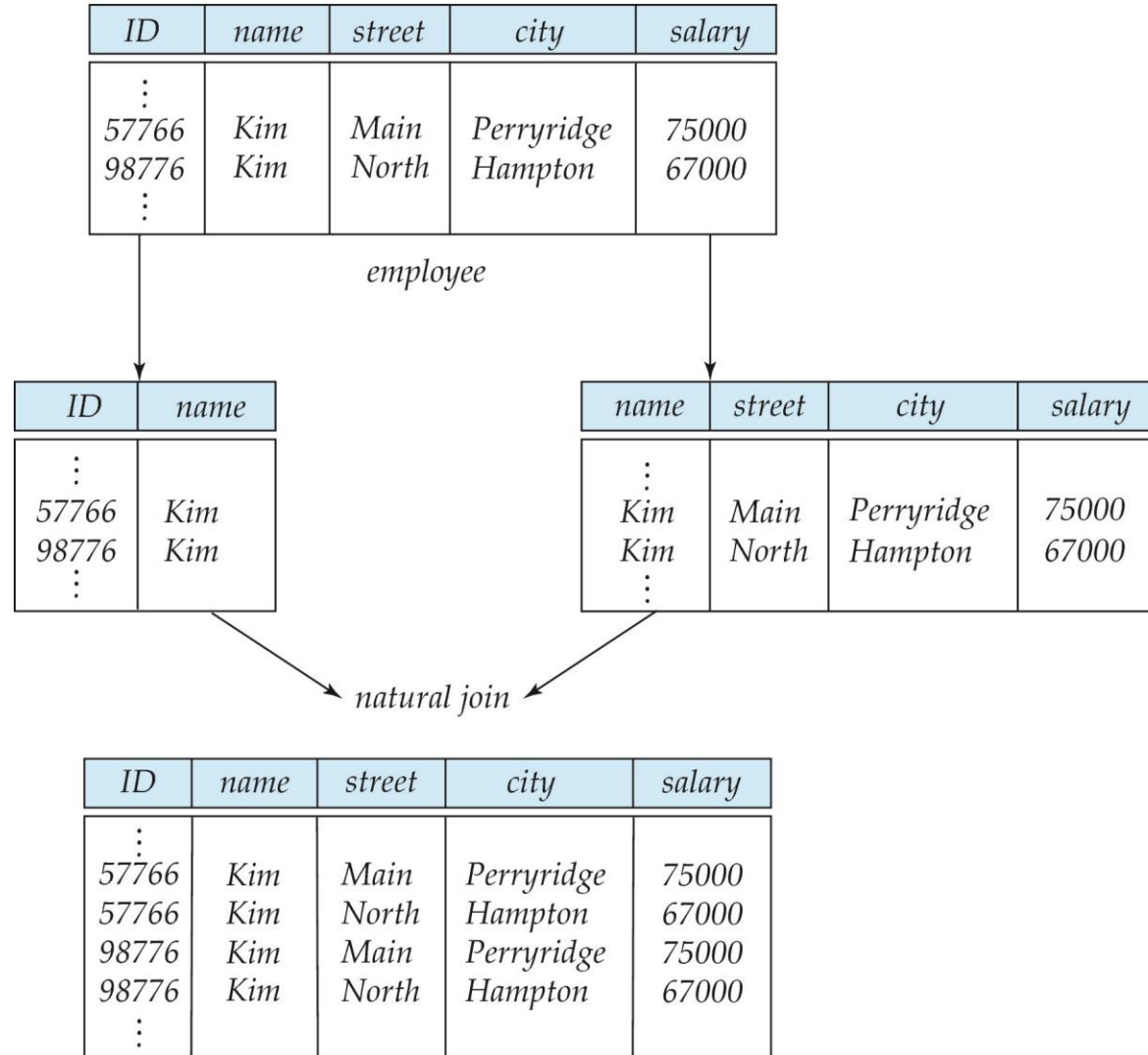
*employee1* (*ID*, *name*)

*employee2* (*name*, *street*, *city*, *salary*)

Chúng ta sẽ mất thông tin vì không thể tạo lại quan hệ gốc *employee*



# Ví dụ phép tách không tốt



# Phép tách không mất mát thông tin (*Lossless join*)

- Định nghĩa: Cho lược đồ quan hệ  $R(U)$  phép tách  $R$  thành các sơ đồ con  $\{R_1, R_2, \dots, R_k\}$  được gọi là **phép tách không mất mát thông tin** đối với một tập phụ thuộc hàm  $F$  nếu với mọi quan hệ  $r$  xác định trên  $R$  thỏa mãn  $F$  thì:

$$r = \Pi_{R_1}(r) \bowtie \Pi_{R_2}(r) \bowtie \dots \bowtie \Pi_{R_k}(r)$$

- Ví dụ:

Suppliers(sid, sname, city, NOE, pname, colour, quantity)

⇒ S1(sid, sname, city, NOE)

SP1(sid, pname, colour, quantity)

# Định lý tách đôi

- Cho lược đồ quan hệ  $R(U)$ , tập phụ thuộc hàm  $F$ . Phép tách  $R$  thành  $R_1(U_1)$ ,  $R_2(U_2)$  là một phép tách **không mất mát thông tin** nếu 1 trong 2 phụ thuộc hàm sau là thỏa mãn trên  $F^+$ :

$$U_1 \cap U_2 \rightarrow U_1 - U_2$$

$$U_1 \cap U_2 \rightarrow U_2 - U_1$$

- **Hệ quả:** Cho lược đồ quan hệ  $R(U)$  và phụ thuộc hàm  $X \rightarrow Y$  thỏa mãn trên  $R(U)$ . Phép tách  $R$  thành 2 lược đồ con  $R_1(U_1)$ ,  $R_2(U_2)$  là một phép tách không mất mát thông tin với:

$$U_1 = XY$$

$$U_2 = XZ \text{ với } Z = U \setminus XY$$

# Kiểm tra tính không mất mát thông tin

- **Vào:**  $R(A_1, A_2, \dots, A_n)$ , tập phụ thuộc hàm  $F$ , phép tách  $\{R_1, R_2, \dots, R_k\}$
- **Ra:** Cho biết phép tách là mất mát thông tin hay không

- **Thuật toán**

**B1.** Thiết lập một bảng  $k$  hàng,  $n$  cột

Xét cột  $j$ , và hàng  $i$ :

Nếu  $A_j$  là thuộc tính của  $R_i$  thì điền  $a_j$  vào ô  $(i,j)$ , ngược lại điền  $b_{ij}$

**Bi.** Xét  $f = X \rightarrow Y \in F$

Nếu  $\exists$  2 hàng  $t1, t2$  thuộc bảng:  $t1[X] = t2[X]$

thì  $t1[Y] \neq t2[Y]$ , ưu tiên đồng nhất về giá trị  $a$

Lặp cho tới khi không thể thay đổi được giá trị nào trong bảng

**Bn.** Nếu bảng có 1 hàng gồm các kí hiệu  $a_1, a_2, \dots, a_n$

thì phép tách là không mất mát thông tin

ngược lại, phép tách không bảo toàn thông tin

# Kiểm tra ...

- $R = ABCD$  được tách thành  $R_1=AB$ ,  $R_2 =BD$ ,  $R_3=ABC$ ,  $R_4=BCD$ .  $F = \{A \rightarrow C, B \rightarrow C, CD \rightarrow B, C \rightarrow D\}$
- B1: Tạo bảng gồm 4 hàng, 4 cột

	A	B	C	D
$R_1$	$a_1$	$a_2$	$b_{13}$	$b_{14}$
$R_2$	$b_{21}$	$a_2$	$b_{23}$	$a_4$
$R_3$	$a_1$	$a_2$	$a_3$	$b_{34}$
$R_4$	$b_{41}$	$a_2$	$a_3$	$a_4$

# Kiểm tra

B1

	A	B	C	D
R <sub>1</sub>	a <sub>1</sub>	a <sub>2</sub>	b <sub>13</sub>	b <sub>14</sub>
R <sub>2</sub>	b <sub>21</sub>	a <sub>2</sub>	b <sub>23</sub>	a <sub>4</sub>
R <sub>3</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	b <sub>34</sub>
R <sub>4</sub>	b <sub>41</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>

B2.  $A \rightarrow C$

	A	B	C	D
R <sub>1</sub>	a <sub>1</sub>	a <sub>2</sub>	<b>a<sub>3</sub></b>	b <sub>14</sub>
R <sub>2</sub>	b <sub>21</sub>	a <sub>2</sub>	b <sub>23</sub>	a <sub>4</sub>
R <sub>3</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	b <sub>34</sub>
R <sub>4</sub>	b <sub>41</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>

B4.  $C \rightarrow D$

	A	B	C	D
R <sub>1</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	<b>a<sub>4</sub></b>
R <sub>2</sub>	b <sub>21</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>
R <sub>3</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	<b>a<sub>4</sub></b>
R <sub>4</sub>	b <sub>41</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>

B3.  $B \rightarrow C$

	A	B	C	D
R <sub>1</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	b <sub>14</sub>
R <sub>2</sub>	b <sub>21</sub>	a <sub>2</sub>	<b>a<sub>3</sub></b>	a <sub>4</sub>
R <sub>3</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	b <sub>34</sub>
R <sub>4</sub>	b <sub>41</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>

Ta có 2 hàng có toàn các giá trị a. Chứng tỏ phép tách đã cho là không mất mát thông tin

# Phép tách bảo toàn tập phụ thuộc hàm

- Hình chiếu của tập phụ thuộc hàm

Cho sơ đồ quan hệ  $R$ , tập phụ thuộc hàm  $F$ , phép tách  $\{R_1, R_2, \dots, R_k\}$  của  $R$  trên  $F$

Hình chiếu  $F_i$  của  $F$  trên  $R_i$  là tập tất cả  $X \rightarrow Y \in F^+ : XY \subseteq R_i$

- Phép tách sơ đồ quan hệ  $R$  thành  $\{R_1, R_2, \dots, R_k\}$  là một phép tách bảo toàn tập phụ thuộc hàm  $F$  nếu

$$(F_1 \cup F_2 \dots \cup F_k)^+ = F^+$$

hay hợp của tất cả các phụ thuộc hàm trong các hình chiếu của  $F$  lên các sơ đồ con sẽ suy diễn ra các phụ thuộc hàm trong  $F$ .

# Bài tập

- **Ví dụ 1:** Cho  $R(U)$ ,  $U = \{A, B, C\}$   $F = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$  được tách thành  $R_1(A, B)$ ,  $R_2(B, C)$ .

Phép tách này có phải là bảo toàn tập phụ thuộc hàm không?

- **Ví dụ 2:** Cho  $R(U)$ ,  $U = \{A, B, C\}$ ,  $F = \{AB \rightarrow C, C \rightarrow B\}$  được tách thành  $R_1(A, B)$ ,  $R_2(B, C)$ . Phép tách này có bảo toàn tập pth không, có mất mát thông tin không?
- **Ví dụ 3:** Cho  $R(U)$ ,  $U = \{A, B, C, D\}$ ,  $F = \{A \rightarrow B, C \rightarrow D\}$  được tách thành  $R_1(A, B)$ ,  $R_2(C, D)$ . Phép tách này có bảo toàn tập pth không, có mất mát thông tin không?



- Một phép tách có bảo toàn tập phụ thuộc hàm thì không đảm bảo là nó sẽ không mất mát thông tin và ngược lại

# Nội dung

Mục đích của chuẩn hoá là gì?



Thế nào là chuẩn hóa?  
Có bao nhiêu dạng chuẩn?

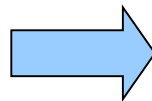
# Các dạng chuẩn

- Mục đích:
  - Mỗi dạng chuẩn đảm bảo ngăn ngừa (giảm thiểu) một số các dạng dư thừa hay dị thường dữ liệu
- Các dạng chuẩn hay sử dụng
  - Dạng chuẩn 1 (1NF)
  - Dạng chuẩn 2 (2NF)
  - Dạng chuẩn 3 (3NF)
  - Dạng chuẩn Boye-Code (BCNF)
  - Dạng chuẩn 4 (4NF)

# Dạng chuẩn 1 (1NF)

- Định nghĩa: Một sơ đồ quan hệ R được gọi là ở dạng chuẩn 1 nếu *tất cả các miền giá trị của các thuộc tính trong R đều chỉ chứa giá trị nguyên tố*.
- Giá trị nguyên tố là giá trị mà không thể chia nhỏ ra được nữa
- Ví dụ: Quan hệ không ở 1NF và quan hệ sau khi chuẩn hóa về 1NF

sname	city	product	
		name	price
Blake	London	Nut	100
Blake	London	Bolt	120
Smith	Paris	Screw	75



sname	city	item	price
Blake	London	Nut	100
Blake	London	Bolt	120
Smith	Paris	Screw	75

# Dạng chuẩn 2 (2NF)

- Định nghĩa: Một sơ đồ quan hệ R được coi là ở dạng chuẩn 2 nếu
  - Sơ đồ quan hệ này ở 1NF
  - Tất cả các thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào khóa chính
- Giải thích:
  - A là một thuộc tính khóa nếu A thuộc một khóa tối thiểu nào đó của R. Ngược lại A là thuộc tính không khóa
  - Phụ thuộc hàm đầy đủ (slide tiếp theo)

# Phụ thuộc hàm đầy đủ

- Định nghĩa: Cho lược đồ quan hệ  $R(U)$ ,  $F$  là tập phụ thuộc hàm trên  $R$ .  $X, Y \subseteq U$ .  $Y$  được gọi là **phụ thuộc đầy đủ** vào  $X$  nếu:
  - $X \rightarrow Y$  thuộc  $F^+$
  - $\nexists X' \subset X : X' \rightarrow Y \in F^+$
- Ý nghĩa
  - $Y$  phụ thuộc hàm vào  $X$
  - $Y$  phụ thuộc hàm vào bất kì một tập con nào của  $X$
- Các phụ thuộc hàm không đầy đủ còn gọi là **phụ thuộc bộ phận**

# Ví dụ

- Cho sơ đồ quan hệ Sales(sid, sname, city, item, price), tập phụ thuộc hàm  $F = \{sid \rightarrow (sname, city), (sid, item) \rightarrow price\}$ , và khóa chính bao gồm (sid, item)
- Nhận xét: sname, city không phụ thuộc hàm đầy đủ vào khóa chính
  - $\Rightarrow$  Sales không thuộc 2NF
  - $\Rightarrow$  Chuẩn hoá
    - S(sid, sname, city)
    - Sales (sid, item, price)

# Dạng chuẩn 3 (3NF)

- Định nghĩa: một sơ đồ quan hệ R được coi là ở dạng chuẩn 3 nếu:
  - Sơ đồ quan hệ này ở 2NF
  - Mọi thuộc tính không khóa đều **không** phụ thuộc bắc cầu vào khóa chính
- Ví dụ về phụ thuộc bắc cầu:
  - Cho:
    - $A \rightarrow B$
    - $B \rightarrow C$
  - Thì C phụ thuộc bắc cầu vào A



# Ví dụ

- Ví dụ 1: S (sid, sname, city)

Sales(sid, item, price)

$F = \{sid \rightarrow sname, city\}$

$\Rightarrow$  S, Sales thuộc dạng chuẩn 3, NF3

- Ví dụ 2: ItemInfo(item, price, discount).

$F = \{item \rightarrow price, price \rightarrow discount\}$

$\Rightarrow$  Thuộc tính không khóa discount phụ thuộc bắc cầu vào khóa chính item.

$\Rightarrow$  ItemInfo không ở 3NF.

➤ Chuẩn hoá:

ItemInfo(item, price)

Discount(price, discount)

# Dạng chuẩn Boye-Codd

- Định nghĩa một sơ đồ quan hệ  $R(U)$  với một tập phụ thuộc hàm  $F$  được gọi là ở dạng chuẩn Boye-Codd (BCNF) nếu với  $\forall X \rightarrow A \in F^+$  thì:
  - $A$  là thuộc tính xuất hiện trong  $X$  hoặc
  - $X$  chứa một khóa của quan hệ  $R$ .
- Ví dụ
  - $R = \{\underline{A}, B, C\}$  ;  $F = \{AB \rightarrow C, C \rightarrow B\}$ .
  - $R$  không phải ở BCNF vì  $\exists C \rightarrow B$ ,  $C$  không phải là khóa
- Chú ý:
  - Một quan hệ thuộc 3NF thì chưa chắc đã thuộc BCNF
  - Nhưng một quan hệ thuộc BCNF thì thuộc 3NF

# Tách bảo toàn tập phụ thuộc hàm về 3NF

- **Vào:**  $R(U)$ ,  $F$  (giả thiết  $F$  là phủ tối thiểu)
- **Ra:** Phép tách bảo toàn tập phụ thuộc hàm về 3NF
- **Thuật toán**
  - B1.** Với các  $A_i \in U$ ,  $A_i \notin F$  thì loại  $A_i$  khỏi  $R$  và lập 1 quan hệ mới cho các  $A_i$
  - B2.** Nếu  $\exists f \in F$ ,  $f$  chứa tất cả các thuộc tính của  $R$  thì kết quả là  $R$
  - B3.** Ngược lại, với mỗi  $X \rightarrow A \in F$ , xác định một quan hệ  $R_i(XA)$ .  
Nếu  $\exists X \rightarrow A_i, X \rightarrow A_j$  thì tạo một quan hệ chung  $R'(XA_iA_j)$

# Ví dụ

Cho  $R = \{A, B, C, D, E, F, G\}$

$F = \{A \rightarrow B, ACD \rightarrow E, EF \rightarrow G\}$

- Xác định phép tách bảo toàn tập phụ thuộc hàm về 3NF

[**B0.** Kiểm tra  $F$  có phải là phủ tối thiểu chưa]

**B1.** Không lập được quan hệ nào mới

**B2.**  $\nexists f \in F: f$  chứa tất cả các thuộc tính của  $R$

**B3.**  $A \rightarrow B \quad \Rightarrow R_1(AB)$

$ACD \rightarrow E \quad \Rightarrow R_2(ACDE)$

$EF \rightarrow G \quad \Rightarrow R_3(EFG)$

# Tách không mất mát thông tin và bảo toàn tập phụ thuộc hàm về 3NF

## •Yêu cầu:

- Bảo toàn tập phụ thuộc hàm (như thuật toán trên)
- Đảm bảo là có một lược đồ con chứa khóa của lược đồ được tách

## •Các bước tiến hành

B1. Tìm một khóa tối thiểu của lược đồ quan hệ R đã cho

B2. Tách lược đồ quan hệ R theo phép tách bảo toàn tập phụ thuộc hàm (cần phải tìm phủ tối thiểu của F)

B3. Nếu một trong các sơ đồ con có chứa khóa tối thiểu thì kết quả của B2 là kết quả cuối cùng.

Ngược lại, thêm vào kết quả đó một sơ đồ quan hệ được tạo bởi khóa tối thiểu tìm được ở 1.

# Ví dụ

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

$$F = \{A \rightarrow B, ACD \rightarrow E, EF \rightarrow G\}$$

**B1.** Khóa tối thiểu cần tìm là ACDF (xem slide 16)

**B2.** Phép tách bảo toàn tập phụ thuộc hàm R cho 3 sơ đồ con  $R_1(AB)$ ,  $R_2(ACDE)$ ,  $R_3(EFG)$  (xem slide 44)

Lưu ý: ở đây F đã là phủ tối thiểu, nếu không cần phải xác định phủ tối thiểu của tập phụ thuộc hàm F

**B3.** Do khóa ACDF không nằm trong bất kỳ một sơ đồ con nào trong 3 sơ đồ con trên, ta lập một sơ đồ con mới  $R_4(ACDF)$

Kết quả cuối cùng ta có phép tách R thành 4 sơ đồ con  $\{R_1, R_2, R_3, R_4\}$  là một phép tách không mất mát thông tin và bảo toàn tập phụ thuộc hàm

# Tách không mất mát thông tin về BCNF

- **Vào:** Sơ đồ quan hệ  $R(U)$ , tập phụ thuộc hàm  $F$
- **Ra:** Phép tách không mất mát thông tin bao gồm một tập các sơ đồ con ở BCNF với các phụ thuộc hàm là hình chiếu của  $F$  lên sơ đồ đó.
- Cách tiến hành
  - B1.**  $KQ = \{R\}$ ,
  - B2.** Với mỗi  $S \in KQ$ ,  $S$  không ở BCNF, xét  $X \rightarrow A \in F_S$ , với điều kiện  $X$  không chứa khóa của  $S$  và  $A \notin X$ . Thay thế  $S$  bởi  $S_1, S_2$  với  $S_1 = A \cup \{X\}$ ,  $S_2 = \{S\} \setminus A$
  - B3.** Lặp (B2) cho đến khi  $\forall S \in KQ$  đều ở BCNF  
KQ gồm các sơ đồ con của phép tách yêu cầu

# Tách không mất mát thông tin về BCNF

$$R = \{\underline{A}, B, C\} ; F = \{AB \rightarrow C, C \rightarrow B\}$$

- $KQ = \{R\}$
- R không ở BCNF
  - Xét  $C \rightarrow B \in F$ : C không chứa khóa của R và  $B \notin C$   
→ Thay R bởi  $R_1 = \{C, B\}$  và  $R_2 = \{A, C\}$
  - $KQ = \{R_1, R_2\}$
- $R_1 = \{\underline{C}, B\}, F_{R_1} = \{C \rightarrow B\}$ : BCNF
- $R_2 = \{\underline{A}, C\}, F_{R_2} = \{\}$ : BCNF

**Kết quả cuối cùng:** R được tách thành  $R_1 = \{C, B\}$ ,  $R_2 = \{A, C\}$



# Phụ thuộc đa trị (multivalued dependency)

- Cho  $R(U)$ , một phụ thuộc đa trị  $X \twoheadrightarrow Y$  xác định trong  $R$  nếu như có hai bộ giá trị trong  $R$  có cùng giá trị với  $X$ , thì khi đổi giá trị thành phần  $Y$  của chúng, kết quả là hai bộ cũng thuộc quan hệ
- Ý nghĩa:
  - $X \twoheadrightarrow Y$  trong  $R(X,Y,Z)$  nếu  $(x,y,z)$  và  $(x,y',z')$  là hai bộ trong  $R$  thì  $(x,y',z)$  và  $(x,y,z')$  cũng là hai bộ thuộc  $R$
  - $Y$  và  $Z$  (với  $Z = R - X - Y$ ) là độc lập lẫn nhau

# Phụ thuộc đa trị (multivalued dependency)

- Phụ thuộc đa trị tồn tại khi có hai thuộc tính/tập thuộc tính độc lập lẫn nhau, nhưng cùng phụ thuộc vào một thuộc tính/tập thuộc tính khác
- Ví dụ 1: **Color** và **Manuf\_Year** là độc lập lẫn nhau, nhưng cùng phụ thuộc vào **Bike\_model**.
  - $\text{Bike\_model} \twoheadrightarrow \text{Color}$
  - $\text{Bike\_model} \twoheadrightarrow \text{Manuf\_Year}$
- Ví dụ 2: **Child** và **Salary** là độc lập lẫn nhau, nhưng cùng phụ thuộc vào **Employee**
  - $\text{Employee} \twoheadrightarrow \text{Child}$
  - $\text{Employee} \twoheadrightarrow \text{Salary}$

BIKE_MODEL	MANUF_YEAR	COLOR
M2011	2008	White
M2001	2008	Black
M3001	2013	White
M3001	2013	Black

Employee	Child	Salary	Year
Nam	Hoa	20	2010
Nam	Thao	20	2010
Nam	Hoa	25	2015
Nam	Thao	25	2015

# Hệ tiên đề đối với các phụ thuộc hàm và phụ thuộc đa trị

Cho  $R(U)$ ,  $X, Y, Z, W \subseteq U$  ( $XY = X \cup Y$ )

- A1: Phản xạ đối với FD (*reflexivity*):

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

- A2: Tăng trưởng đối với FD (*augmentation*):

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

- A3: Bắt cầu đối với FD (*transitivity*):

Nếu  $X \rightarrow Y$ ,  $Y \rightarrow Z$  thì  $X \rightarrow Z$ .

- A4: Luật bù đối với MVD (*complementation*):

Nếu  $X \twoheadrightarrow Y$  thì  $X \twoheadrightarrow U \setminus XY$ .

## Hệ tiên đề đối với các phụ thuộc hàm và phụ thuộc đa trị (2)

Cho  $R(U)$ ,  $X, Y, Z, W \subseteq U$  ( $XY = X \cup Y$ )

- A5: Tăng trưởng đối với MVD (*augmentation*):

Nếu  $X \twoheadrightarrow Y$  và  $V \subseteq W$  thì  $WX \twoheadrightarrow VY$ .

- A6: Bắc cầu đối với MVD (*transitivity*):

Nếu  $X \twoheadrightarrow Y$ ,  $Y \twoheadrightarrow Z$  thì  $X \twoheadrightarrow Z \setminus Y$ .

- A7:

Nếu  $X \rightarrow Y$  thì  $X \twoheadrightarrow Y$ .

- A8:

Nếu  $X \twoheadrightarrow Y$ ,  $W \rightarrow Z$  với  $Z \subseteq Y$  và  $W \cap Y = \emptyset$  thì  $X \rightarrow Z$ .

# Các luật suy diễn bổ sung đối với các phụ thuộc đa trị

- Luật hợp (*union*):  
Nếu  $X \twoheadrightarrow Y$ ,  $X \twoheadrightarrow Z$  thì  $X \twoheadrightarrow YZ$ .
- Luật tựa bắc cầu (*pseudotransitivity*):  
Nếu  $X \twoheadrightarrow Y$ ,  $WY \twoheadrightarrow Z$  thì  $WX \twoheadrightarrow Z \setminus WY$ .
- Luật tựa bắc cầu hỗn hợp (*mixed pseudotransitivity*):  
Nếu  $X \twoheadrightarrow Y$ ,  $XY \twoheadrightarrow Z$  thì  $X \twoheadrightarrow Z \setminus Y$ .
- Luật tách (*decomposition*):  
Nếu  $X \twoheadrightarrow Y$ ,  $X \twoheadrightarrow Z$  thì  
 $X \twoheadrightarrow Y \cap Z$ ,  $X \twoheadrightarrow Y \setminus Z$ ,  $X \twoheadrightarrow Z \setminus Y$ .

# Bao đóng của tập phụ thuộc hàm và phụ thuộc đa trị

- Định nghĩa: bao đóng của tập các phụ thuộc hàm và phụ thuộc đa trị  $D$ , ký hiệu  $D^+$  là tập tất cả các phụ thuộc hàm và các phụ thuộc đa trị được suy diễn logic từ  $D$
- Kiểm tra:  $X \twoheadrightarrow Y \in D^+$ ?
  - Tính cơ sở phụ thuộc của  $X$  đối với  $D$
  - Kiểm tra  $Y \setminus X$  có phải là hợp của vài tập hợp trong cơ sở phụ thuộc của  $X$  hay không

# Tính cơ sở phụ thuộc

- **Vào:** Tập các phụ thuộc đa trị  $M$  trên tập thuộc tính  $U$  và tập thuộc tính  $X \subseteq U$
- **Ra:** Cơ sở phụ thuộc của  $X$  đối với  $M$
- Cách tiến hành:
  - B1.** Đặt  $T$  là tập các tập con  $Z$  của  $U$ : với  $W \twoheadrightarrow Y \in M$  mà  $W \subseteq X$  thì  $Z = Y \setminus X$  hoặc  $Z = U \setminus XY$
  - B2.**  $T$  được thiết lập cho tới khi là một tập các tập rời nhau. Nếu có một cặp  $Z_1, Z_2$  không tách rời nhau thì thay chúng bởi  $Z_1 \setminus Z_2, Z_2 \setminus Z_1, Z_1 \cap Z_2$  với điều kiện không ghi nhận tập rỗng. Gọi **S** là tập thu được sau bước này.
  - B3.** Tìm các phụ thuộc có dạng  $V \twoheadrightarrow W \in M$  và một tập  $Y \in S : Y \cap W \neq \emptyset, Y \cap V = \emptyset$   
Thay  $Y$  bằng  $Y \cap W$  và  $Y \setminus W$  cho đến khi không thay đổi  $S$  được nữa
  - B4.** Tập  $S$  thu được sau bước này là cơ sở phụ thuộc của  $X$

# Phép tách không mất thông tin?

- **Vào:**  $R(A_1, A_2, \dots, A_n)$ ,  $F$ ,  $M$ , phép tách  $\{R_1, R_2, \dots, R_k\}$
- **Ra:** phép tách là mất mát thông tin hay không
- **Thuật toán** (*tổng quát hoá thuật toán đã trình bày*)

**B1.** Thiết lập một bảng  $k$  hàng,  $n$  cột

**Bi.** Xét  $f = X \rightarrow Y \in F$ : thực hiện đồng nhất bảng

*Xét  $X \twoheadrightarrow Y$ : nếu  $\exists 2$  hàng  $t1, t2$  thuộc bảng :  $t1[X] = t2[X]$  thì thêm vào bảng đó một hàng mới  $u$ :  $u[X] = t1[X], u[Y] = t1[Y], u[R \setminus XY] = t2[R \setminus XY]$*

Lặp cho tới khi không thể thay đổi được giá trị nào trong bảng

**Bn.** Nếu bảng có 1 hàng gồm các kí hiệu  $a_1, a_2, \dots, a_n$  thì phép tách là không mất mát thông tin. Ngược lại, phép tách không bảo toàn thông tin



## Dạng chuẩn 4 (4NF)

- Một quan hệ R ở dạng chuẩn bốn (4NF) nếu với mỗi phụ thuộc đa trị  $X \twoheadrightarrow Y$  với  $Y \neq \emptyset$ ,  $Y \not\subseteq X$  và XY không chứa tất cả các thuộc tính của R thì X chứa một khóa của R
- Chú ý: nếu R chỉ có các phụ thuộc hàm thì dạng chuẩn bốn chính là dạng chuẩn Boye-Codd và  $X \twoheadrightarrow Y$  phải có nghĩa là  $X \rightarrow Y$

# Tách không làm mất mát thông tin về 4NF

- Tương tự như thuật toán tách về chuẩn Boye-Codd nhưng áp dụng với phụ thuộc đa trị
- Ví dụ:
  - $R(\underline{\text{Employee}}, \underline{\text{Child}}, \underline{\text{Salary}}, \text{Year})$
  - Khóa: Employee, Child, Salary
  - Tách thành
    - $R1(\underline{\text{Employee}}, \underline{\text{Child}})$
    - $R2(\underline{\text{Employee}}, \underline{\text{Salary}}, \underline{\text{Year}})$

Employee	Child	Salary	Year
Nam	Hoa	20	2010
Nam	Thao	20	2010
Nam	Hoa	25	2015
Nam	Thao	25	2015

## Lưu ý:

- Rất hiếm các trường hợp mà 1 lược đồ ở dạng chuẩn 3 mà lại không ở dạng chuẩn 4
- Hầu hết các lược đồ ở dạng chuẩn 3 thì cũng ở dạng chuẩn Boye-Codd

# Kết luận

- Tầm quan trọng của thiết kế CSDL
  - Ảnh hưởng đến chất lượng dữ liệu lưu trữ
  - Hiệu quả của việc khai thác dữ liệu
- Mục đích của thiết kế CSDL: tránh
  - Dư thừa dữ liệu
  - Dị thường dữ liệu khi thêm/sửa đổi
  - Hiệu quả trong tìm kiếm
- Đưa về các dạng chuẩn
  - 2NF: giảm bớt sự dư thừa để tránh các dị thường khi cập nhật
  - 3NF: tránh các dị thường khi thêm/xoá
  - BCNF: tránh các dị thường khi cập nhật
  - 4NF: giảm dư thừa dữ liệu do thuộc tính đa trị



# Bài tập

□ Cho lược đồ quan hệ  $R(U,F)$  với

$$U = \{A, B, C, D, E, G, H, I\}$$

$$F = \{AB \rightarrow DE, BC \rightarrow H, G \rightarrow AI, D \rightarrow E\}$$

1. Hãy xác định một khóa tối thiểu của lược đồ quan hệ trên.
2. Hãy xác định phủ tối thiểu của tập phụ thuộc hàm  $F$  cho ở trên

□ Cho lược đồ quan hệ  $R(U,F)$  với

$$U = \{A, B, C, D, E, G, H, I\}$$

$$F = \{BG \rightarrow E, H \rightarrow I, BC \rightarrow IH, A \rightarrow CD\}$$

Hãy xác định một khóa tối thiểu của lược đồ quan hệ  $R$ .

Hãy xác định phủ tối thiểu của tập phụ thuộc hàm  $F$  cho ở trên

# Bài tập

□ Cho lược đồ quan hệ  $R(U, F)$  với  
 $U = \{A, B, C, D, E, G, H\}$ , tập các phụ thuộc hàm  
 $F = \{A \rightarrow B, E \rightarrow HB, CEB \rightarrow G, AD \rightarrow E\}$

1. Tìm một khóa tối thiểu của  $R(U, F)$
2. Hãy xác định phủ tối thiểu của tập phụ thuộc hàm  $F$  cho ở trên
3. Tách bảo toàn tập phụ thuộc hàm lược đồ trên về dạng chuẩn 3
4. Tách bảo toàn tập phụ thuộc hàm và bảo toàn thông tin lược đồ trên về dạng chuẩn 3