

---

# Chương 4

## 4.2. Thuật toán thiết kế CSDL

---

# Phép tách bảo toàn phụ thuộc

## ■ Một số định nghĩa

- Quan hệ vũ trụ đơn  $R=\{A_1, \dots, A_n\}$  chứa tất cả các thuộc tính của CSDL
- Tập hợp  $F$  các phụ thuộc hàm phải được thỏa mãn trên  $R$
- Thuật toán tách sẽ phân chia  $R$  thành 1 tập các quan hệ  $D= \{R_1, \dots, R_m\}$  gọi là **lược đồ CSDL quan hệ** và  $D$  gọi là một phép tách
- Điều kiện bảo toàn thuộc tính  $\cup R_i=R$
- Tính không đầy đủ của các dạng chuẩn: không đảm bảo thiết kế CSDL tốt

# Phép tách bảo toàn FD... (tt)

- ĐN (đk bảo toàn phụ thuộc) Một phụ thuộc hàm  $X \rightarrow Y$  phải xuất hiện trong  $F$  hoặc suy diễn ra được từ các DF thông qua các phép biến đổi trong  $F$
- ĐN(*phép chiếu của  $F$  trên  $R_i$* - ký hiệu là  $\pi_{R_i}(F)$ ): Cho trước một tập hợp các phụ thuộc  $F$  trên  $R$ , trong đó  $R_i$  là một tập con của  $R$ , là một tập hợp các phụ thuộc hàm  $X \rightarrow Y$  trong  $F^+$  sao cho các thuộc tính trong  $X \cup Y$  đều được chứa trong  $R_i$

# Phép tách bảo toàn FD... (tt)

- Ta nói rằng phép tách  $D = \{R_1, R_2, \dots, R_m\}$  của  $R$  bảo toàn phụ thuộc đối với  $F$  nếu hợp của các phép chiếu của  $F$  trên mỗi  $R_i$  trong  $D$  là tương đương với  $F$ . Điều đó có nghĩa là:

$$(\pi_{R_1}(F) \cup \pi_{R_2}(F) \cup \dots \cup \pi_{R_m}(F))^+ = F^+$$

# Phép tách bảo toàn FD... (tt)

- Để kiểm tra xem một phụ thuộc hàm  $X \rightarrow B$ , trong đó  $X$  là tập thuộc tính thuộc về  $R_i$ ,  $B$  là một thuộc tính thuộc  $R_i$  có thỏa mãn trong  $R_i$  hay không ta làm như sau:
  - Tính  $X^+$ ,
  - Với mỗi thuộc tính  $B$ , kiểm tra
    1.  $B$  là một thuộc tính của  $R_i$
    2.  $B$  là ở trong  $X^+$ :
    3.  $B$  không ở trong  $X$
  - Khi đó phụ thuộc hàm  $X \rightarrow B$  thỏa mãn trong  $R_i$ .

# Phép tách bảo toàn FD... (tt)

- Ví dụ Xét lược đồ quan hệ:

$R = \{ \underline{A}, B, C, D \}$  với các phụ thuộc hàm:

$A \rightarrow BCD; BC \rightarrow DA; D \rightarrow B$

- Lược đồ này có hai khóa dự tuyển là A và BC.
- Lược đồ này vi phạm BCNF. Nó được tách thành:
  - $R1 = \{ \underline{D}, B \}$ , lược đồ này chứa phụ thuộc hàm  $D \rightarrow B$
  - $R2 = \{ \underline{A}, C, D \}$ , lược đồ này chứa phụ thuộc hàm  $A \rightarrow CD$
- Rõ ràng sau khi tách, phụ thuộc hàm  $BC \rightarrow DA$  bị mất.

# Phép tách bảo toàn FD... (tt)

- Thuật toán 5.1: Tạo một phép tách bảo toàn phụ thuộc  $D = \{R_1, R_2, \dots, R_n\}$  của một quan hệ vũ trụ  $R$  dựa trên một tập phụ thuộc hàm  $F$  sao cho mỗi  $R_i$  trong  $D$  là ở 3NF.
- Thuật toán này chỉ đảm bảo tính chất **bảo toàn phụ thuộc**, không đảm bảo tính chất **nối không mất mát**.

# Phép tách bảo toàn FD... (tt)

- Input: Một quan hệ vũ trụ  $R$  và một tập phụ thuộc hàm  $F$  trên các thuộc tính của  $R$ .
- 1. Tìm phủ tối thiểu  $G$  của  $F$ .
- 2. Với mỗi vế trái  $X$  của một phụ thuộc hàm xuất hiện trong  $G$ , hãy tạo một lược đồ trong  $D$  với các thuộc tính  $\{X \cup \{A_1\} \cup \{A_2\} \cup \dots \cup \{A_k\}\}$  trong đó  $X \rightarrow A_1, X \rightarrow A_2, \dots, X \rightarrow A_k$  chỉ là các phụ thuộc hàm trong  $G$  với  $X$  là vế trái ( $X$  là khóa của quan hệ này).
- 3. Đặt các thuộc tính còn lại (những thuộc tính chưa được đặt vào quan hệ nào) vào một quan hệ đơn để đảm bảo tính chất bảo toàn thuộc tính.



# Phép tách bảo toàn FD... (tt)

- Ví dụ: Xét lược đồ quan hệ:

$R = \{ \underline{A}, B, C, D \}$  với các phụ thuộc hàm:

$A \rightarrow BCD; BC \rightarrow DA; D \rightarrow B$

Yêu cầu tách lược đồ  $R$  thành tập các lược đồ sao cho bảo toàn các phụ thuộc hàm trong  $R$

# Phép tách bảo toàn FD... (tt)

- Ta thực hiện thuật toán như sau:
  - B1: Tìm G là phủ tối thiểu của F.
    - Theo thuật toán tìm phủ tối thiểu, đầu tiên ta làm cho các vế phải trong G chỉ chứa một thuộc tính, ta có:
      - \*  $G = \{A \rightarrow B; A \rightarrow C; A \rightarrow D; BC \rightarrow D; BC \rightarrow A; D \rightarrow B\}$
    - Sau đó ta bỏ đi các phụ thuộc hàm thừa ( $A \rightarrow B$ )
      - \*  $G = \{A \rightarrow C; A \rightarrow D; BC \rightarrow D; BC \rightarrow A; D \rightarrow B\}.$
  - B2: Lược đồ R sẽ được tách thành:
    - $R1(\underline{A}, C, D)$
    - $R2(\underline{B}, \underline{C}, D, A)$
    - $R3(\underline{D}, B)$

# Phép tách không mất mát

- Phép tách D phải có một tính chất nữa là nối không mất mát (hoặc tính chất nối không phụ thêm), nó đảm bảo rằng không có các bộ giả được tạo ra khi áp dụng một phép nối tự nhiên vào các quan hệ trong phép tách

# Phép tách không mất mát (tt)

## ■ Thuật toán 5.2: kiểm tra nổi không mất mát

Input: Một quan hệ vũ trụ  $R(A_1, A_2, \dots, A_n)$ , một phép tách  $D = \{R_1, R_2, \dots, R_m\}$  của  $R$  và một tập  $F$  các phụ thuộc hàm.

1. Tạo một ma trận  $S$  có  $m$  hàng,  $n$  cột. Mỗi cột của ma trận ứng với một thuộc tính, mỗi hàng ứng với mỗi quan hệ  $R_i$
2. Đặt  $S(i, j) = 1$  nếu thuộc tính  $A_j$  thuộc về quan hệ  $R_i$  và bằng 0 trong trường hợp ngược lại.
3. Lặp lại vòng lặp sau đây cho đến khi nào việc thực hiện vòng lặp không làm thay đổi  $S$ : Với mỗi phụ thuộc hàm  $X \rightarrow Y$  trong  $F$ , xác định các hàng trong  $S$  có các ký hiệu 1 như nhau trong các cột ứng với các thuộc tính trong  $X$ . Nếu có một hàng trong số đó chứa 1 trong các cột ứng với thuộc tính  $Y$  thì hãy làm cho các hàng cho các cột tương ứng của các hàng khác cũng chứa 1.
4. Nếu có một hàng chứa toàn ký hiệu “1” thì phép tách có tính chất nổi không mất mát, ngược lại, phép tách không có tính chất đó.

# Ví dụ

- $R = (\text{MaNV}, \text{TenNV}, \text{MaDA}, \text{TenDA}, \text{DDiem}, \text{Sốgiờ})$
- $F = \{\text{MaNV} \rightarrow \text{TenNV},$   
 $\text{MaDA} \rightarrow \{\text{TenDA}, \text{DDiem}\},$   
 $\{\text{MaNV}, \text{MaDA}\} \rightarrow \text{Sốgiờ}\}$
- $R1 = (\text{MaNV}, \text{TenNV})$
- $R2 = (\text{MaDA}, \text{TenDA}, \text{DDiem})$
- $R3 = (\text{MaNV}, \text{MaDA}, \text{Sốgiờ})$

Kiểm tra xem phép tách có mất mát không?

B1, 2

	MaNV	TenNV	MaDA	TenDA	Ddiem	Số giờ
R1	1	1	0	0	0	0
R2	0	0	1	1	1	0
R3	1	0	1	0	0	1

B3

	MaNV	TenNV	MaDA	TenDA	Ddiem	Số giờ
R1	1	1	0	0	0	0
R2	0	0	1	1	1	0
R3	1	<del>0</del> 1	1	0	0	1

B1, 2

	MaNV	TenNV	MaDA	TenDA	Ddiem	Số giờ
R1	1	1	0	0	0	0
R2	0	0	1	1	1	0
R3	1	0	1	0	0	1

B3

	MaNV	TenNV	MaDA	TenDA	Ddiem	Số giờ
R1	1	1	0	0	0	0
R2	0	0	1	1	1	0
R3	1	1	1	<del>0</del> 1	<del>0</del> 1	1

# Phép tách không mất mát (tt)

- **Thuật toán 5.3:** Tách quan hệ thành các quan hệ BCNF với tính chất nổi không mất mát.
- Đặt  $D^{\text{old}} := \{R\}$  ;
- FOR (lược đồ Q trong D)
  - FOR EACH (Q  $\models X \rightarrow Y$  vi phạm BCNF )
    - $D^{\text{old}} := D^{\text{new}}$
    - $D^{\text{new}} := \{D^{\text{old}}, (Q-Y), (X \cup Y)\}$



# Phép tách không mất mát (tt)

- Kiểm tra lược đồ Q trong D vi phạm BCNF
  - Cách 1: Với mỗi phụ thuộc hàm  $X \rightarrow Y$  trong Q,
    - Tính  $X^+$ .
    - Nếu  $X^+$  không chứa tất cả các thuộc tính trong Q thì  $X \rightarrow Y$  vi phạm BCNF bởi vì X không phải là một siêu khóa.
  - Cách 2: dựa trên quan sát rằng khi một lược đồ quan hệ Q vi phạm BCNF thì có tồn tại một cặp thuộc tính A,B trong Q sao cho  $\{Q - \{A,B\}\} \rightarrow A$ .
    - Tính bao đóng  $\{Q - \{A,B\}\}^+$  cho mỗi cặp thuộc tính  $\{A,B\}$  của Q
    - kiểm tra xem bao đóng có chứa A (hoặc B) hay không, chúng ta có thể xác định được Q có ở BCNF hay không.

# Phép tách không mất mát (tt)

- Ví dụ: Xét lược đồ quan hệ
  - $R = \{ A, B, C, D, E, F \}$
  - $A \rightarrow BCDEF, BC \rightarrow ADEF, B \rightarrow F, D \rightarrow E, D \rightarrow B$
- Có 2 khóa dự tuyển là A và BC
- Ta có  $B \rightarrow F$  vi phạm BCNF vì B không phải là siêu khóa, R được tách thành
  - $R_1(B, F)$  với phụ thuộc hàm  $B \rightarrow F$
- $R_2(A, B, C, D, E)$  với các phụ thuộc hàm  $A \rightarrow CDE, BC \rightarrow ADE, D \rightarrow E, D \rightarrow B$
- Do  $D \rightarrow E$  vi phạm BCNF (D là một thuộc tính không khóa),  $R_2$  được tách thành:
  - $R_{21}(D, E)$  với phụ thuộc hàm  $D \rightarrow E$
  - $R_{22}(ABCD)$  với các phụ thuộc hàm  $A \rightarrow BCD, BC \rightarrow AD, D \rightarrow B$
- Do  $D \rightarrow B$  vi phạm BCNF (D không phải là thuộc tính khóa),  $R_{22}$  được tách thành:
  - $R_{221}(\underline{D}, B)$
  - $R_{222}(\underline{A}, C, D)$  với phụ thuộc hàm  $A \rightarrow CD$  (phụ thuộc hàm  $BC \rightarrow AD$  bị mất)
- Tóm lại, ta có phép tách  $D = \{R_1, R_{21}, R_{221}, R_{222}\}$ . Phép tách này có tính chất nổi không mất thông tin nhưng không bảo toàn phụ thuộc.

# Phép tách không mất mát (tt)

- **Thuật toán 5.4**: tổng hợp quan hệ với tính chất bảo toàn phụ thuộc và nối không mất mát.
  - Tìm phủ tối thiểu  $G$  cho  $F$ .
  - Với mỗi vế trái  $X$  của một phụ thuộc hàm xuất hiện trong  $G$  hãy tạo ra một lược đồ quan hệ trong  $D$  với các thuộc tính  $\{X \cup \{A_1\} \cup \{A_2\} \cup \dots \cup \{A_k\}\}$ , trong đó  $X \rightarrow A_1, X \rightarrow A_2, \dots, X \rightarrow A_k$  chỉ là các phụ thuộc hàm ở trong  $G$  với  $X$  là vế trái ( $X$  là khóa của quan hệ này).
  - Nếu không có lược đồ quan hệ nào trong  $D$  chứa một khóa của  $R$  thì hãy tạo ra thêm một lược đồ quan hệ trong  $D$  chứa các thuộc tính tạo nên một khóa của  $R$ .

# Dạng chuẩn 4- phụ thuộc đa trị

- Ví dụ

NHANVIEN(TENNV, TENDA, TENTHANNHAN)

NHANVIÊN	TênNV	TênDA	TênconNV
	Nam	DA01	Lan
	Nam	DA02	Hoa
	Nam	DA01	Hoa
	Nam	DA02	Lan

# Dạng chuẩn 4-pt đa trị (tt)

- Một phụ thuộc đa trị (MVD), ký hiệu là  $X \twoheadrightarrow Y$ , chỉ ra ràng buộc sau đây trên một trạng thái quan hệ bất kỳ của R: Nếu hai bộ  $t_1$  và  $t_2$  tồn tại trong R sao cho  $t_1[X] = t_2[X]$  thì hai bộ  $t_3$  và  $t_4$  cũng tồn tại trong R với các tính chất sau:
  - $t_3[X] = t_4[X] = t_1[X] = t_2[X]$
  - $t_3[Y] = t_1[Y]$  và  $t_4[Y] = t_2[Y]$
  - $t_3[Z] = t_2[Z]$  và  $t_4[Z] = t_1[Z]$  với  $Z = (R - (X \cup Y))$

# Dạng chuẩn 4-pt đa trị (tt)

- Định nghĩa: Một lược đồ quan hệ R là ở dạng chuẩn 4 (4NF) đối với một tập hợp các phụ thuộc F (gồm các phụ thuộc hàm và phụ thuộc đa trị) nếu với mỗi phụ thuộc đa trị không tầm thường  $X \twoheadrightarrow Y$  trong  $F^+$ , X là một siêu khóa của R.

# Dạng chuẩn 4-pt đa trị (tt)

- **Thuật toán** Tách quan hệ thành các quan hệ 4NF với tính chất nổi không mất mát.
  - Input: Một quan hệ vũ trụ  $R$  và một tập phụ thuộc hàm và phụ thuộc đa trị  $F$ .
- 1. Đặt  $D := \{R\}$ ;
- 2. Khi có một lược đồ quan hệ  $Q$  trong  $D$  không ở 4NF, thực hiện:
  - Chọn một lược đồ quan hệ  $Q$  trong  $D$  không ở 4NF;
  - Tìm một phụ thuộc đa trị không tầm thường  $X \twoheadrightarrow Y$  trong  $Q$  vi phạm 4NF;
  - Thay thế  $Q$  trong  $D$  bằng hai lược đồ  $(Q - Y)$  và  $(X \cup Y)$

# Dạng chuẩn 5- pt nối

- Một *phụ thuộc nối* (JD), ký hiệu là  $JD(R_1, R_2, \dots, R_n)$  trên lược đồ quan hệ  $R$  chỉ ra một ràng buộc trên các trạng thái  $r$  của  $R$ . Ràng buộc đó tuyên bố rằng mỗi trạng thái hợp pháp  $r$  của  $R$  phải có phép tách có tính chất nối không mất mát thành  $R_1, R_2, \dots, R_n$ . Điều đó nghĩa là:

$$* (\pi_{R_1}(r), \pi_{R_2}(r), \dots, \pi_{R_n}(r)) = r$$

- Một phụ thuộc nối  $JD(R_1, R_2, \dots, R_n)$  là một phụ thuộc nối tầm thường nếu một trong các lược đồ quan hệ  $R_i$  ở trong  $JD(R_1, R_2, \dots, R_n)$  là bằng  $R$ .



# Dạng chuẩn 5- pt nối (tt)

- ĐN: Một lược đồ quan hệ  $R$  là ở dạng chuẩn 5 (5NF) (hoặc dạng chuẩn nối chiếu PJNF – Project-Join normal form) đối với một tập  $F$  các phụ thuộc hàm, phụ thuộc đa trị và phụ thuộc nối nếu với mỗi phụ thuộc nối không tầm thường  $JD(R_1, R_2, \dots, R_n)$  trong  $F^+$ , mỗi  $R_i$  là một siêu khóa của  $R$ .

---

# Tài liệu tham khảo

- ***Giáo trình CSDL***
  - Chương 4
- ***Database management system***
  - Chapter 15
- **Fundamentals of Database Systems**
  - Chapter 15

