

BÀI GIẢNG CƠ SỞ DỮ LIỆU



Chương 3 PHỤ THUỘC HÀM VÀ CHUẨN HÓA

Giáo viên: Nguyễn Thị Uyên Nhi
Email: nhintu@due.edu.vn

TIN HỌC QUẢN LÝ

Mục tiêu

2

- Làm thế nào để đưa được từ một lược đồ cơ sở dữ liệu chưa tốt thành một lược đồ cơ sở dữ liệu tốt hơn nhằm
 - Tránh được dư thừa dữ liệu
 - Tránh được các bất thường khi thêm, sửa, xóa dữ liệu đồng thời cho phép tìm kiếm thông tin một cách dễ dàng.

Mục tiêu

3

▪ GIỚI HẠN CỦA LƯỢC ĐỒ ER

- ✓ Cung cấp một tập các hướng dẫn → không đưa tới một lược đồ CSDL duy nhất
- ✓ Không đưa ra cách đánh giá giữa các lược đồ khác nhau
- ✓ Lý thuyết về chuẩn hóa CSDL quan hệ cung cấp **kỹ thuật** để phân tích và chuyển hóa từ lược đồ ER sang lược đồ quan hệ

Sự dư thừa

4

■ Dư thừa → Dị thường

- ✓ Thao tác sửa đổi: cập nhật tất cả các giá trị liên quan
- ✓ Thao tác xóa: người cuối cùng của đơn vị → mất thông tin về đơn vị
- ✓ Thao tác chèn

TENPHG	MAPHG	TRPHG	NG_NHAMCHUC	MANV	TENNV	HONV
Nghiên cứu	5	123456789	01/02/2012	123456789	Tùng	Nguyễn
Điều hành	4	333444555	01/01/2010	333444555	Hưng	Nguyễn
Quản lý	1	999888777	01/06/2012	999888777	Vĩnh	Phạm

→ Các giá trị không xác định: Đặt thuộc tính trường phòng vào quan hệ NHANVIEN NHANVIEN thay vì vào quan hệ PHONGBAN

Nội dung chi tiết

5

I. Phụ thuộc hàm

II. Chuẩn hóa lược đồ quan hệ

Một số vấn đề gặp phải khi tổ chức dữ liệu

6

Cho lược đồ quan hệ:

THI (MASV, HOTENSV, MONHOC, DIEMTHI)

MaSV	HoTenSV	MonHoc	DiemThi
00DTH189	Nguyễn Văn Thành	Mạng	7
00DTH189	Nguyễn Văn Thành	Pascal	9
00DTH211	Trần Thu Hà	Pascal	5
00DTH189	Nguyễn Văn Thành	Pascal	8

Một số **vấn đề nảy sinh** khi tổ chức dữ liệu theo lược đồ trên:

- Dư thừa dữ liệu
- Bất thường khi sửa
- Bất thường khi thêm
- Bất thường khi xóa

Một số vấn đề gặp phải khi tổ chức dữ liệu

7

Hướng giải quyết:

Phân rã lược đồ quan hệ **THI** thành 2 lược đồ quan hệ

SINHVIEN (MASV, HOTENSV)

KETQUA (MASV, MONHOC, DIEMTHI)

I. Phụ thuộc hàm (PTH)

8

1. Định nghĩa
2. Biểu diễn PTH bằng đồ thị
3. Suy diễn logic các phụ thuộc hàm
4. Hệ tiên đề Amstrong
5. Bao đóng
6. Bao đóng của tập thuộc tính
7. Khóa - Thuật toán tìm khóa

Phụ thuộc hàm (PTH)

9

Phụ thuộc hàm (Functional Dependency)

- Là sự biểu diễn RBTV dưới hình thức toán học
- Bảo đảm thông tin không bị tổn thất khi phân rã hoặc kết nối giữa các quan hệ

Định nghĩa

10

Định nghĩa:

- ❖ Quan hệ **R** được định nghĩa trên tập thuộc tính $U = \{ A_1, A_2, \dots, A_n \}$
- ❖ $A, B \subset U$ là 2 tập con của tập thuộc tính U
- ❖ Nếu tồn tại một ánh xạ **f**: **A** \rightarrow **B** thì ta nói rằng **A** xác định hàm **B**, hay **B** phụ thuộc hàm vào **A**.
- ❖ **Ký hiệu:** **A** \rightarrow **B**

Định nghĩa

11

Định nghĩa hình thức của PTH:

Quan hệ $Q(A, B, C)$ có phụ thuộc hàm A xác định B (ký hiệu là $A \rightarrow B$) nếu:

$$\forall q, q' \in Q: \quad q.A = q'.A \text{ thì } q.B = q'.B$$

- Nghĩa là: *Ứng với 1 giá trị của A thì có một giá trị duy nhất của B*
- A là vế trái của phụ thuộc hàm, B là vế phải của phụ thuộc hàm.
- PTH $A \rightarrow A$ được gọi là PTH hiển nhiên

Ví dụ

12

Ví dụ: Trong quan hệ:

KETQUA (SINHVIEN, MONHOC, TINCHI, HOCKY, DIEM)

SINHVIEN	MONHOC	TINCHI	HOCKY	DIEM
Hoàng Thị Hà	Toán CC	3	1	8
Hoàng Thị Hà	SQL Server	2	2	3
Hoàng Thị Hà	SQL Server	2	3	6
Nguyễn Văn Hai	Toán CC	3	1	7

Một số điều kiện ràng buộc:

- Nếu biết được môn học thì sẽ biết được tín chỉ của môn đó
Monhoc → Tinchi
- Nếu biết được sinh viên, môn học, học kỳ thì sẽ biết được điểm của môn mà sinh viên học trong học kỳ đó
Sinhvien, Monhoc, Hocky → Diem

Ví dụ

13

Trong quan hệ:

SINHVIEN (MASV, HOTEN, PHAI, NGSINH, QUEQUAN, DIACHI)

Có các phụ thuộc hàm sau:

- **MASV → QUEQUAN, DIACHI**
- **MASV, HOTEN → NGSINH, QUEQUAN**

Không có các phụ thuộc hàm sau:

- **HOTEN → NGSINH, QUEQUAN**

Biểu diễn PTH bằng đồ thị

14

PTH có thể biểu diễn bằng đồ thị có hướng:

❖ **Các nút trong đồ thị chia thành 2 loại:**

- Nút thuộc tính: biểu diễn bằng tên thuộc tính
- Nút phụ thuộc hàm: biểu diễn bằng hình tròn có số thứ tự của PTH.

❖ **Các cung trong đồ thị cũng có 2 loại:**

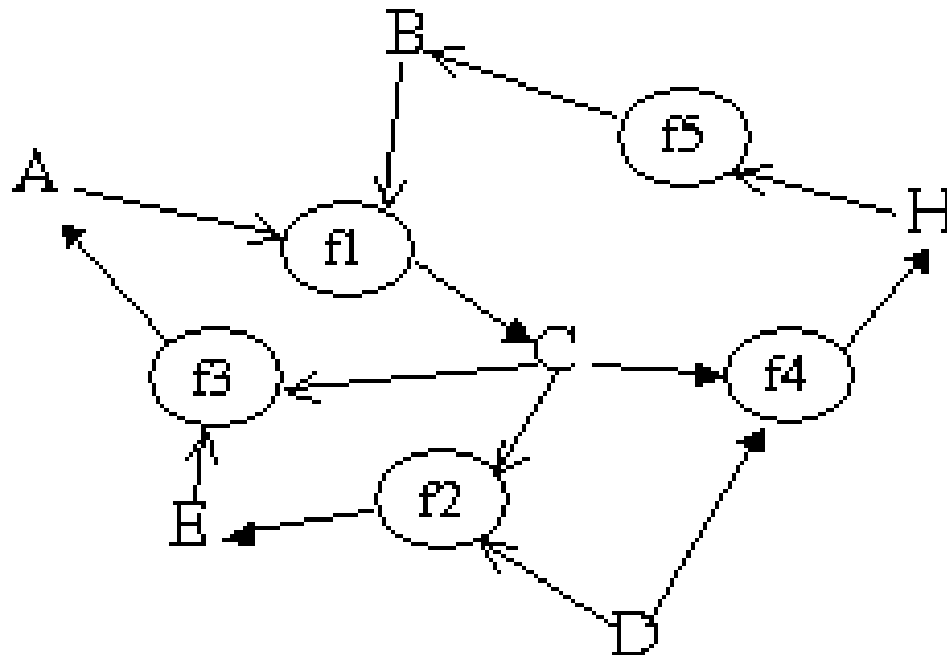
- Cung đến PTH: xuất phát từ các thuộc tính ở vế trái của các PTH
- Cung rời PTH: hướng đến các thuộc tính ở vế phải của các PTH

Biểu diễn PTH bằng đồ thị

15

Ví dụ: Cho $R(A, B, C, D, E, H)$

Với $F = \{AB \rightarrow C, CD \rightarrow E, EC \rightarrow A, CD \rightarrow H, H \rightarrow B\}$

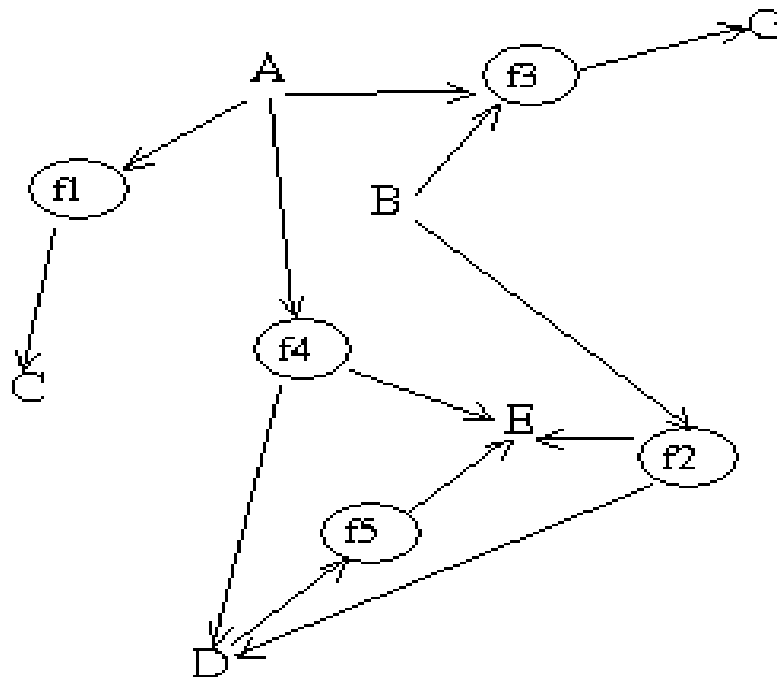


Biểu diễn PTH bằng đồ thị

16

Ví dụ: Cho R (A, B, C, D, E, G)

Với $F = \{A \rightarrow C; B \rightarrow DE; AB \rightarrow G; A \rightarrow ED; D \rightarrow E\}$



Biểu diễn PTH bằng đồ thị

17

BÀI TẬP 1:

Cho $R(U, F)$ với tập thuộc tính $U = \{A, B, C, D, E, G, H, I, J, K\}$ và tập PTH $F = \{A \rightarrow B; A \rightarrow D; B \rightarrow EG; C \rightarrow HI; C \rightarrow J, AC \rightarrow K\}$

Vẽ cây phụ thuộc hàm

Biểu diễn PTH bằng đồ thị

18

BÀI TẬP 2:

Cho lược đồ quan hệ:

$Q(A,B,C,D,E,G,H,K,L,M,N,O,P,U,V,S);$

trong đó mỗi thuộc tính là một ký tự và tập các PTH:

$F=\{D \rightarrow B,C; D,S \rightarrow A; L \rightarrow G,H; G \rightarrow E; N \rightarrow K,L,M; S \rightarrow N,O,P,U,V\}$

Hãy vẽ cây PTH.

Suy diễn logic các PTH

19

- ❖ Cho lược đồ quan hệ **R** với tập thuộc tính **U** và tập các PTH **F**
- ❖ **$X \rightarrow Y$** là 1 PTH; $X, Y \subseteq U$
- ❖ Ta nói rằng $X \rightarrow Y$ được *suy diễn logic* từ **F** nếu:
 $\forall r \in R$, nếu r thỏa tất cả các PTH trong **F thì r cũng thỏa $X \rightarrow Y$**
- ❖ Ký hiệu: $F \models X \rightarrow Y$

Ví dụ:

Với $F = \{X \rightarrow Y, X \rightarrow Z, Y \rightarrow T\}$

Thì ta có các phụ thuộc hàm:

$X \rightarrow YZ$ và **$X \rightarrow T$**

Hệ tiên đề Amstrong

20

Năm 1974, **Amstrong** đã đưa ra hệ tiên đề (gọi là hệ luật dẫn Amstrong). Cho lược đồ quan hệ Q với tập thuộc tính U .

$X, Y, Z, W \subseteq U$. PTH có các quy tắc **cơ bản** sau:

- Luật **phản xạ**:

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

- Luật **tăng trưởng** (mở rộng hai vế):

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

- Luật **bắc cầu**:

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

Hệ tiên đề Amstrong

21

Ví dụ:

Cho $F = \{AB \rightarrow C, C \rightarrow A\}$. **CMR:** $BC \rightarrow ABC$

- Ta có:
- | | |
|--------------------------|-----------------|
| (1) $C \rightarrow A$ | (giả thiết) |
| (2) $BC \rightarrow AB$ | (tăng trưởng 1) |
| (3) $AB \rightarrow C$ | (giả thiết) |
| (4) $AB \rightarrow ABC$ | (tăng trưởng 3) |
| (5) $BC \rightarrow ABC$ | (bắc cầu 2 & 4) |

Hệ tiên đề Amstrong

22

BÀI TẬP:

Cho $F = \{A \rightarrow B, BC \rightarrow D\}$

CMR: $AC \rightarrow BCD$

Hệ tiên đề Amstrong

23

Các luật bổ sung:

- Cho lược đồ quan hệ Q với tập thuộc tính U .
 $X, Y, Z, W \subseteq U$. PTH có các tính chất sau:
- Luật phân rã:
Nếu $X \rightarrow YZ$ thì $X \rightarrow Y$ và $X \rightarrow Z$
- Luật hợp:
Nếu $X \rightarrow Y$ và $X \rightarrow Z$ thì $X \rightarrow YZ$
- Luật tựa bắc cầu:
Nếu $X \rightarrow Y$ và $YZ \rightarrow W$ thì $XZ \rightarrow W$

Ví dụ:

Cho $F = \{A \rightarrow BC, AC \rightarrow D\}$

CMR: $AC \rightarrow BCD$

Hệ tiên đề Amstrong

24

Ví dụ: Cho $R(A, B, C, D, E, G, H)$. **CMR: $AB \rightarrow E$**

Với $F = \{AB \rightarrow C, B \rightarrow D, CD \rightarrow E, CE \rightarrow GH, G \rightarrow A\}$

Ta có:	(1) $AB \rightarrow C$	(giả thiết)
	(2) $AB \rightarrow AB$	(phản xạ)
	(3) $AB \rightarrow B$	(luật tách)
	(4) $B \rightarrow D$	(giả thiết)
	(5) $AB \rightarrow D$	(bắc cầu 3 & 4)
	(6) $AB \rightarrow CD$	(hợp 1 & 5)
	(7) $CD \rightarrow E$	(giả thiết)
	(8) $AB \rightarrow E$	(bắc cầu 6 & 7)

Hệ tiên đề Amstrong

25

Ví dụ: Cho $R(A, B, C, D, E, G, H, I, J)$

$F = \{AB \rightarrow E, AG \rightarrow J, BE \rightarrow I, E \rightarrow G, GI \rightarrow H\}$

CMR: $AB \rightarrow GH$

- | | | |
|-----|---------------------|------------------|
| 1) | $AB \rightarrow E$ | (giả thiết– f1) |
| 2) | $AB \rightarrow AB$ | (Phản xạ) |
| 3) | $AB \rightarrow B$ | (luật tách) |
| 4) | $AB \rightarrow BE$ | (hợp 1 & 3) |
| 5) | $BE \rightarrow I$ | (giả thiết - f3) |
| 6) | $AB \rightarrow I$ | (bắc cầu 4 & 5) |
| 7) | $E \rightarrow G$ | (giả thiết - f4) |
| 8) | $AB \rightarrow G$ | (bắc cầu 1 & 7) |
| 9) | $AB \rightarrow GI$ | (hợp 6 & 8) |
| 10) | $GI \rightarrow H$ | (cho trước - f5) |
| 11) | $AB \rightarrow H$ | (bắc cầu 9 & 10) |
| 12) | $AB \rightarrow GH$ | (hợp 8 & 11) |

Bao đóng (Closure)

26

- ❖ Gọi **F^+** là **bao đóng** (Closure) của F , tức là tập các phụ thuộc hàm được suy diễn logic từ F .
- ❖ Nếu $F = F^+$ thì ta nói F là họ đầy đủ (full family) của các phụ thuộc hàm.

Bao đóng (Closure)

27

Bài toán thành viên (MemberShip):

Kiểm tra PTH $X \rightarrow Y$ có được suy diễn logic từ F không? (tức là $X \rightarrow Y \in F^+$?)

- Đây là một bài toán khó giải.
- Đòi hỏi phải có một hệ luật dẫn để suy diễn logic các phụ thuộc hàm

Bao đóng (Closure)



Ví dụ:

Cho lược đồ $Q(A, B, C, D, E, G)$

$F = \{AE \rightarrow C, CG \rightarrow A, BD \rightarrow G, GA \rightarrow E\}$

CMR: $BDC \rightarrow Q^+ \in F^+$

$(Q^+ = ABCDEG)$

Ta có:

(1) $BDC \rightarrow BDC$

(phản xạ)

(2) $BD \rightarrow G$

(giả thiết f3)

(3) $CG \rightarrow A$

(giả thiết f2)

(4) $BDC \rightarrow A$

(tựa bắc cầu 2,3)

(5) $BDC \rightarrow GA$

(hợp 2 & 4)

(6) $BDC \rightarrow E$

(bắc cầu 5 & f4)

(7) $BDC \rightarrow Q^+$

(hợp 1,4,5,6)

Bao đóng của tập thuộc tính

29

❖ Định nghĩa:

- Bao đóng của tập thuộc tính X đối với tập các PTH F (ký hiệu: X_F^+) là tập tất cả các thuộc tính A có thể suy dẫn từ X nhờ tập bao đóng của F (F^+)

- $$X_F^+ = \{ A \mid X \rightarrow A \in F^+ \}$$

❖ Nhận xét:

- $X \subseteq X_F^+$
- $X \rightarrow A \in F^+ \Leftrightarrow A \subseteq X_F^+$

Bao đóng của tập thuộc tính

30

Thuật toán Tìm bao đóng của tập thuộc tính

Input: Tập U hữu hạn các thuộc tính & tập các PTH F trên U & $X \subseteq U$.

Output: X_F^+

Phương pháp: Tính liên tiếp X_0, X_1, X_2, \dots theo quy tắc như sau:

Bước 1. $X_0 = X$

Bước 2. Lần lượt xét các pth của F , nếu $Y \rightarrow Z$ có

$Y \subseteq X_i$, thì $\mathbf{X_{i+1} = X_i \cup Z}$, và loại $Y \rightarrow Z$ ra khỏi F

Bước 3. Tiếp tục cho đến khi $X_{i+1} = X_i$

(Vì $X = X_0 \subseteq X_1 \subseteq X_2 \subseteq \dots \subseteq U$, mà U hữu hạn cho nên sẽ tồn tại 1 chỉ số i nào đó mà $X_{i+1} = X_i$)

→ Khi đó $X_F^+ = X_i$

Bao đóng của tập thuộc tính

31

Ví dụ 1: Cho $R(U)$ với $U=ABCDEFG$

$F = \{AB \rightarrow C, C \rightarrow A, BC \rightarrow D, ACD \rightarrow B, D \rightarrow EG, BE \rightarrow C, CG \rightarrow BD, CE \rightarrow AG\}$

Tính X_F^+ , với:

$X = D$

$X = BD$

Bao đóng của tập thuộc tính

32

Ví dụ 1:

$F = \{AB \rightarrow C, C \rightarrow A, BC \rightarrow D, ACD \rightarrow B, D \rightarrow EG,$
 $BE \rightarrow C, CG \rightarrow BD, CE \rightarrow AG\}$

X_i	Tập PTH	X_{i+1}
$X_0 = D$	$D \rightarrow EG$	DEG
$X_1 = DEG$		

$\rightarrow \{D\}^+ = DEG$

$$F = \{AB \rightarrow C, C \rightarrow A, BC \rightarrow D, ACD \rightarrow B, D \rightarrow EG, BE \rightarrow C, CG \rightarrow BD, CE \rightarrow AG\}$$

X_i	Tập PTH	X_{i+1}
$X_0 = BD$	$D \rightarrow EG$	BDEG
$X_1 = BDEG$	$D \rightarrow EG, BE \rightarrow C$	BCDEG
$X_2 = BCDEG$	$C \rightarrow A, BC \rightarrow D,$ $D \rightarrow EG, BE \rightarrow C,$ $CG \rightarrow BD, CE \rightarrow AG$	ABCDEFG
$X_3 = ABCDEG$	Tất cả	ABCDEFG
$X_4 = ABCDEG$		

$$\rightarrow \{BD\}_F^+ = ABCDEG$$

Bao đóng của tập thuộc tính

34

BÀI TẬP: Cho $R(U)$ với $U=ABCDEFGH$

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

Tính X_F^+ , với:

$X = BD; X_F^+ = ABCDEH$

$X = AC; X_F^+ = ACDEH$

Khóa – Thuật toán tìm khóa

35

Định nghĩa:

- R là lược đồ quan hệ định nghĩa trên tập các thuộc tính $U = \{A_1, A_2, \dots, A_n\}$
- Tập các PTH $F = \{f_1, f_2, \dots, f_m\}$ xác định trên R .
- $K \subseteq U$ là khóa của R nếu thỏa mãn hai điều kiện sau đây:
 1. $K \rightarrow U$ (K là siêu khóa)
 2. $\nexists K' \subset K$ mà $K' \rightarrow U$

Khóa – Thuật toán tìm khóa

36

Bài Toán Tìm khóa:

- Xác định tất cả các khóa của 1 lược đồ quan hệ.
- Bài toán này được giải quyết qua 2 giai đoạn:
 1. **Giai đoạn 1:** Xây dựng tập S chứa tất cả các siêu khóa của R
 2. **Giai đoạn 2:** Xây dựng tập K chứa tất cả các khóa của R từ tập S bằng cách loại bỏ khỏi S những siêu khóa không tối thiểu

Bài Toán Tìm khóa:

- Để xác định tất cả các siêu khóa của 1 lược đồ quan hệ R , ta lần lượt xét $(2^n - 1)$ tập hợp con của R^+ : X_1, X_2, \dots
- Nếu 1 tập con X_i của R^+ có bao đóng bằng đúng R^+ thì tập con X_i chính là 1 siêu khóa.
- Nếu R chỉ có 1 siêu khóa S thì siêu khóa đó cũng là khóa của lược đồ quan hệ R
- Trong trường hợp R có nhiều hơn 1 siêu khóa (hữu hạn), để xác định tất cả các khóa chỉ định, ta so sánh 1 cặp siêu khóa S_i và S_j . Nếu $S_i \subset S_j$, ta loại S_j và giữ lại S_i .
- Lần lượt so sánh từng cặp siêu khóa để loại bỏ tập lớn, cuối cùng thu được tập các khóa chỉ định của R

➔ **Thuật toán không khả thi khi n lớn**

Khóa – Thuật toán tìm khóa

38

Bài Toán Tìm khóa – Thuật toán cải tiến:

Chúng ta sẽ cải tiến thuật toán dựa trên việc phân loại tập thuộc tính R^+

- ↪ A gọi là **thuộc tính nguồn** nếu A không xuất hiện ở vế phải của bất kỳ PTH không hiển nhiên nào của F. **Ký hiệu: TN**
- ↪ A gọi là **thuộc tính đích** nếu A không phải thuộc tính nguồn và A không xuất hiện ở vế trái của bất kỳ PTH không hiển nhiên nào của F. **Ký hiệu: TD**

Khóa – Thuật toán tìm khóa

39

Bài Toán Tìm khóa – Thuật toán cải tiến:

- ↪ Tập hợp các thuộc tính không phải nguồn và không phải đích gọi là tập trung gian. **Ký hiệu: TG**
- ↪ Các tập hợp TN, TD, TG rời nhau từng đôi một và $TN \cup TD \cup TG = R^+$

Nhận xét:

- Nếu K là khóa của R thì K chứa tất cả các thuộc tính nguồn và không chứa bất kỳ thuộc tính đích nào.

Khóa – Thuật toán tìm khóa

40

Thuật toán cải tiến:

B1: Xây dựng 2^v tập con của TG: TG_1, TG_2, \dots bằng phương pháp đường chạy nhị phân.

B2: Xây dựng tập K chứa các siêu khóa

$$\Rightarrow K = \emptyset$$

$$\Rightarrow \forall TG_i, X_i = TN \cup TG_i$$

$$\Rightarrow \text{Tính } X_i^+_F. \text{ Nếu } X_i^+_F = R^+ \text{ thì } K = K \cup X_i$$

B3: Loại bỏ dần các siêu khóa lớn

Khóa – Thuật toán tìm khóa

41

Ví dụ 1:

Cho $R(A, B, C, D, E, G)$ với tập PTH

$$F = \{AE \rightarrow C, CG \rightarrow A, BD \rightarrow G, GA \rightarrow E\}$$

Xác định tất cả các khóa của R .

Ta có:

$$TN = \{B, D\}, TD = \emptyset$$

$$TG = \{A, C, E, G\}$$

→ Xây dựng tập thuộc tính TG_i bằng PP đường chạy nhị phân.

ACEG	X_i	$TN \cup X_i$	$(TN \cup X_i)^+$	SIÊU KHÓA	KHÓA
0000		BD	BDG		
0001	G	BDG	BDG		
0010	E	BDE	BDEG		
0011	EG	BDEG	BDEG		
0100	C	BDC	ABDCGE= R^+	BCD	BCD
0101	CG	BDCG			
0110	CE	BDCE			
0111	CEG	BDCEG			
1000	A	BDA	ABCDGE= R^+	BCA	BCA
1001	AG	BDAG		→ Có 2	
1010	AE	BDAE		khóa: BDC,	
1011	AEG	BDAEG		BDA	
1100	AC	BDAC			
1101	ACG	BDACG		→ 4 thuộc	
1110	ACE	BDACE		tính khóa:	
1111	ACEG	BDACEG		A, B, C, D	

Khóa – Thuật toán tìm khóa

43

Ví dụ 2:

Cho $R(A, B, C, D, E, G)$ với tập PTH

$$F = \{EC \rightarrow B, AB \rightarrow C, EB \rightarrow D, BG \rightarrow A, AE \rightarrow G\}$$

Xác định tất cả các khóa của R .

Ta có:

$$TN = \{E\}, TD = \{D\}$$

$$TG = \{A, B, C, G\}$$

→ Xây dựng tập thuộc tính TG_i bằng PP đường chạy nhị phân.

ABCG	L_i	$X_i = N \cup L_i$	$X_i^+_F$
0000		E	E
0001	G	EG	EG
0010	C	EC	ECBD
0011	CG	ECG	ECGBDA= R^+
0100	B	EB	EBD
0101	BG	EBG	EBGDAC= R^+
0110	BC	EBC	EBCD
0111	BCG	EBCG	
1000	A	EA	AEG
1001	AG	EAG	EAG
1010	AC	EAC	EACBDG= R^+
1011	ACG	EACG	
1100	AB	EAB	EABDCG= R^+
1101	ABG	EABG	
1110	ABC	EABC	
1111	ABCG	EABCG	

→ Có 4 siêu khóa: **ECG, EBG, EAC, EAB**

→ Có 4 khóa: **ECG, EBG, EAC, EAB**

II. Chuẩn hóa lược đồ quan hệ

45

1. Đặt vấn đề
2. Dạng chuẩn 1
3. Dạng chuẩn 2
4. Dạng chuẩn 3
5. Dạng chuẩn Boyce-Codd
6. Chuẩn hóa lược đồ CSDL bằng phương pháp phân rã

Đặt vấn đề

46

Xét quan hệ:

ĐẶT_HÀNG (SODH, NGÀYDH, MAKH, MAHH, SOLUONG)

SODH	NGÀYDH	MAKH	MAHH	SOLUONG
DH01	5/1/2019	KH01	H01	50
DH02	13/2/2019	KH05	H02	30
DH02	13/2/2019	KH05	H03	40

Với tập PTH $F = \{ \text{SODH} \rightarrow \text{NGÀYDH, MAKH};$
 $\text{SODH, MAHH} \rightarrow \text{SOLUONG} \}$

\Rightarrow Có Trùng lặp thông tin

Đặt vấn đề

47

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

Đặt vấn đề

48

Tổ chức lại thành 2 quan hệ như sau:

DAT_HANG (SODH, NGAYDH, MAKH)

Với $F1 = \{SODH \rightarrow NGAYDH, MAKH\}$

SODH	NGAYDH	MAKH
DH01	5/1/99	KH01
DH02	13/2/99	KH05

CHITIET_DH (SODH, MAHH, SOLUONG)

Với $F2 = \{SODH, MAHH \rightarrow SOLUONG\}$

SODH	MAHH	SOLUONG
DH01	H01	50
DH02	H02	30
DH02	H03	40

⇒ *Không còn xảy ra tình trạng trùng lặp thông tin*

Đặt vấn đề

49

- ❖ **Để đánh giá chất lượng thiết kế của lược đồ CSDL**
 - E.F.Codd (tác giả của mô hình dữ liệu quan hệ) đưa ra **3 dạng chuẩn** (Normal Form) – 1NF, 2NF, 3NF
 - 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 phụ thuộc hàm của các thuộc tính**
- ❖ **Ngoài ra còn có các dạng chuẩn:**
 - ❖ Dạng chuẩn 4: dựa trên phụ thuộc đa trị
 - ❖ Dạng chuẩn 5: dựa trên chuỗi kết hàm

Đặt vấn đề

50

- ❖ **Chuẩn hóa dữ liệu:** là quá trình phân tích, tách các lượng đồ quan hệ để đạt mục tiêu xây dựng một LĐQH tốt.
- ❖ **Mục đích của quá trình chuẩn hóa**
 - Để biểu diễn được mọi quan hệ trong CSDL
 - Cực tiểu hóa các sự dư thừa, các cập nhật bất thường
 - 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
- ❖ **Phép tách của lược đồ quan hệ có các tính chất sau:**
 - Tính chất nối không mất mát
 - Tính chất bảo toàn phụ thuộc

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

51

Định nghĩa:

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

Ví dụ: Quan hệ CHUYỀN_MÔN không đạt dạng chuẩn 1

Khắc phục:

CHUYEN_MON (MAGV, MON)

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

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

52

❖ 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: CHUYEN_MON (MAGV, MON)

MAGV	MÔN
GV01	PASCAL, CTDL
GV02	CSDL, PTTKHT

⇒ Môn không là thuộc tính đơn

Dạng chuẩn 2 (2NF)

53

Phụ thuộc đầy đủ:

- Giả sử có 1 lược đồ quan hệ Q và tập phụ thuộc hàm 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 $X' \subset X$, mà $X' \rightarrow A$)

Dạng chuẩn 2 (2NF)

54

Định nghĩa:

- ❖ Một lược đồ quan hệ Q được gọi là đạt chuẩn 2 nếu:
 - Q ở 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 ở dạng chuẩn 2 nếu mọi lược đồ quan hệ con Q_i của nó đều ở dạng chuẩn 2

Ví dụ: Quan hệ

ĐẠT_HANG (SODH, MAHH, NGAYDH, MAKH, SOLUONG)

Với tập PTH

$$F = \{ \text{SODH} \rightarrow \text{NGAYDH, MAKH}; \\ \text{SODH, MAHH} \rightarrow \text{SOLUONG} \}$$

⇒ Không đạt dạng chuẩn 2

Khắc phục: Tách thành 2 quan hệ:

ĐẠT_HANG (SODH, NGAYDH, MAKH)

$$\text{Với } F1 = \{ \text{SODH} \rightarrow \text{NGAYDH, MAKH} \}$$

CHITIET_ĐH (SODH, MAHH, SOLUONG)

$$\text{Với } F2 = \{ \text{SODH, MAHH} \rightarrow \text{SOLUONG} \}$$

Dạng chuẩn 2 (2NF)

56

Nhận xét

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

Dạng chuẩn 3 (3NF)

57

Phụ thuộc bắc cầu:

- Thuộc tính $A \in Q^+$ được gọi là phụ thuộc bắc cầu vào tập thuộc tính X nếu $\exists Y \in Q^+$:
 - $X \rightarrow Y \in F^+$ và $Y \rightarrow A \in F^+$
 - $Y \rightarrow X \notin F^+$
 - $A \notin (X \cup Y)$
- Khi đó $X \rightarrow A$ được gọi là phụ thuộc hàm bắc cầu

Dạng chuẩn 3 (3NF)

58

Định nghĩa:

- ❖ Một lược đồ quan hệ Q được gọi là ở 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 ở dạng chuẩn 3 nếu mọi lược đồ quan hệ con Q_i của nó đều ở dạng chuẩn 3

Ví dụ: Quan hệ

GIANG_DAY (MALOP, MAGV, TENGV, DIACHI)

Với tập PTH:

$F = \{ \text{MALOP} \rightarrow \text{MAGV};$
 $\text{MAGV} \rightarrow \text{TENG V, DIACHI} \}$

\Rightarrow Không đạt dạng chuẩn 3

Khắc phục: Tách thành 2 quan hệ:

GIANG_DAY (MALOP, MAGV)

Với tập PTH $F1 = \{ \text{MALOP} \rightarrow \text{MAGV} \}$

GIAO_VIEN (MAGV, TENGV, DIACHI)

Với tập PTH $F2 = \{ \text{MAGV} \rightarrow \text{TENG V, DIACHI} \}$

Dạng chuẩn 3 (3NF)

60

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

Dạng chuẩn BOYCE-CODD

61

Định nghĩa:

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

Nhận xét:

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

Dạng chuẩn BOYCE-CODD

62

Ví dụ:

DAT_HANG (SODH, NGAYDH, MAKH)

Với $F1 = \{SODH \rightarrow NGAYDH, MAKH\}$

CHITIẾT_ĐH (SODH, MAHH, SOLUONG)

Với $F2 = \{SODH, MAHH \rightarrow SOLUONG\}$

\Rightarrow 2 quan hệ đều đạt dạng chuẩn Boyce-Codd

CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ

63

❖ 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

CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ

64

Sự bảo toàn thông tin

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

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

CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ

65

Định lý Delobel

- Cho lược đồ quan hệ $Q(XYZ)$ và tập phụ thuộc hàm 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

CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ

66

Phương pháp phân rã

Begin

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

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

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;

CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ

67

Ví dụ:

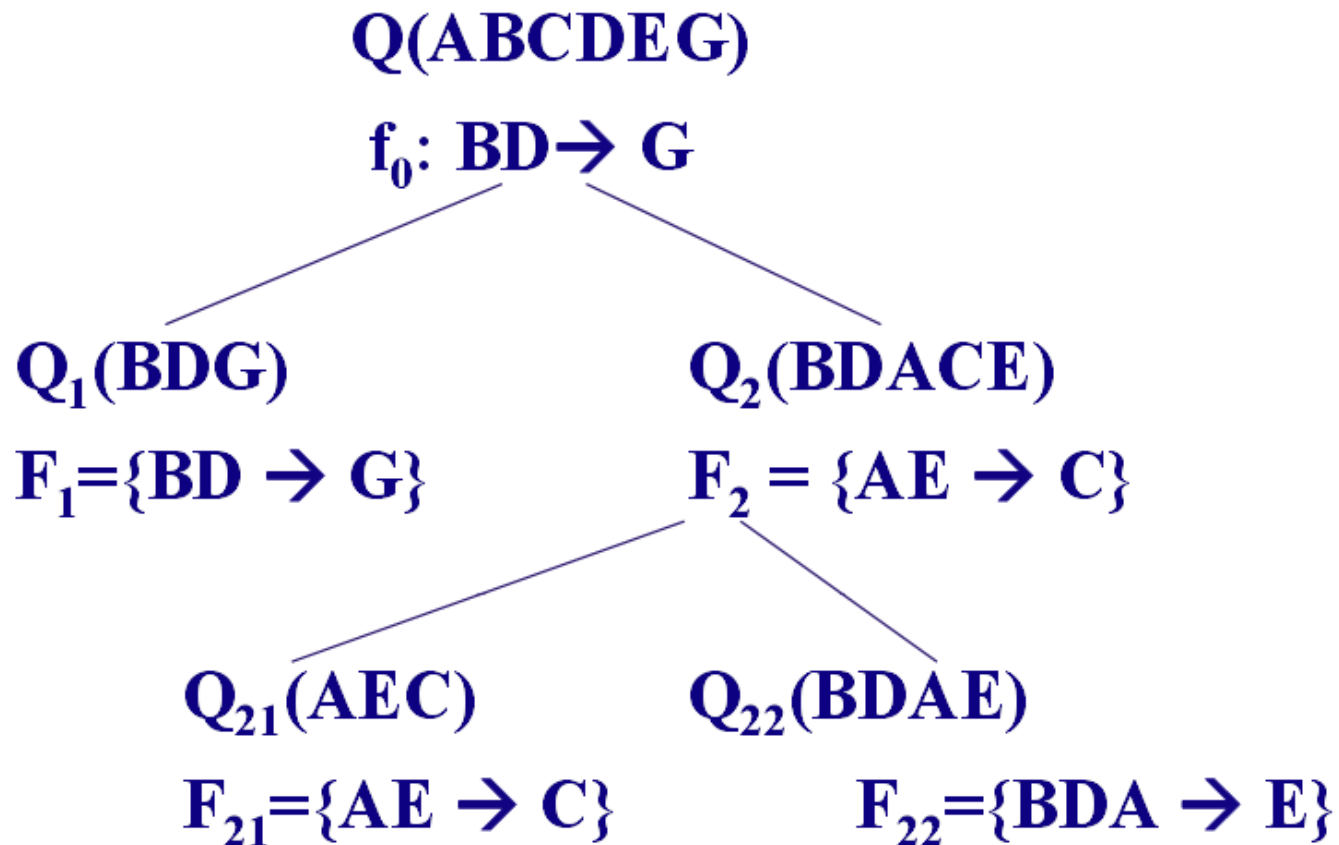
Cho lược đồ quan hệ $Q(ABCDEFG)$ và tập PTH

$$F = \{AE \rightarrow C, CG \rightarrow A, BD \rightarrow G, GA \rightarrow E\}$$

Nhận xét:

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

\Rightarrow Sử dụng phương pháp phân rã



CHUẨN HÓA LƯỢC ĐỒ CSDL BẰNG PHƯƠNG PHÁP PHÂN RÃ

69

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 ít nhất là *dạng chuẩn 3*
- Thuật toán phân rã có thể tạo ra những lược đồ quan hệ con không có nhiều ngữ nghĩa trong thực tế
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)

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 chuẩn 1

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

⇒ Đạt dạng chuẩn 1

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

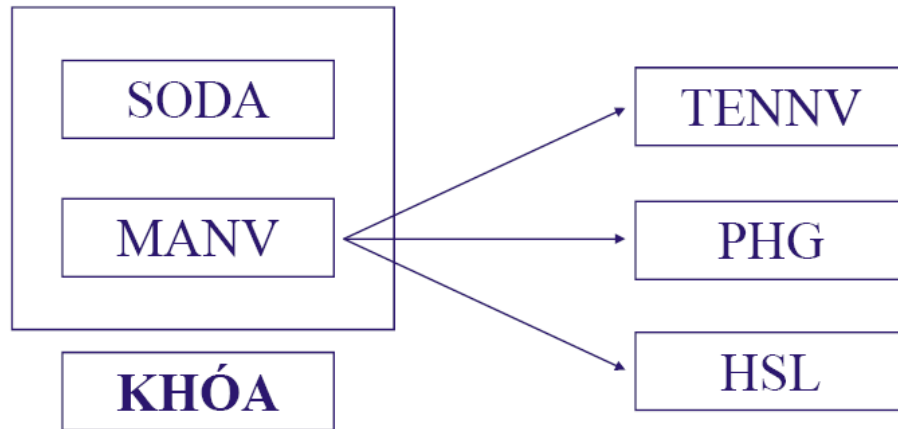
<u>SODA</u>	TENDA	TRGDA
100	TM Dien Tu	789
110	DTao Tu Xa	820
120	Cap Quang	980
130	Trac Nghiem	550

DEAN_NV (SODA, MSNV, TENNV, PHG, HSL, TGIAN)

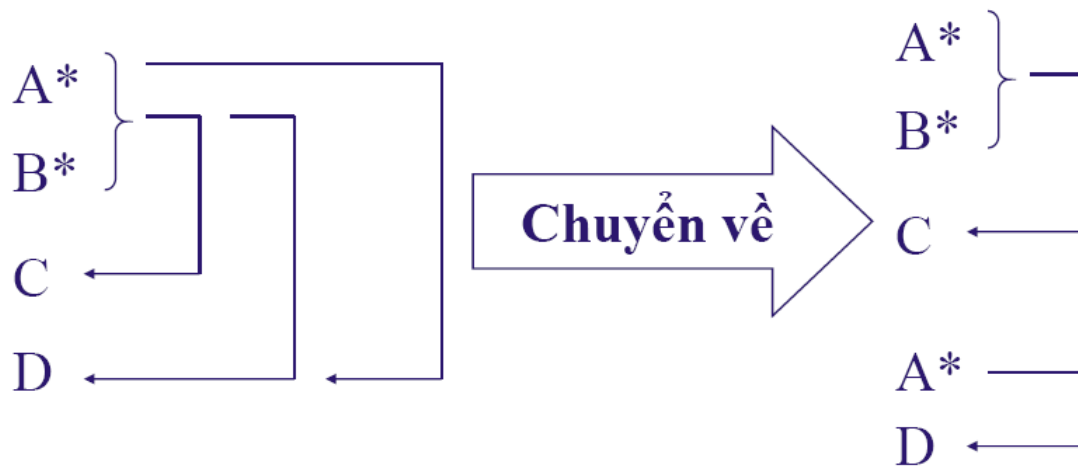
⇒ Đạt dạng chuẩn 1

<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

Quan hệ **DEAN_NV** không đạt chuẩn 2 vì phụ thuộc không đầy đủ



Chuyển về dạng chuẩn 2



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

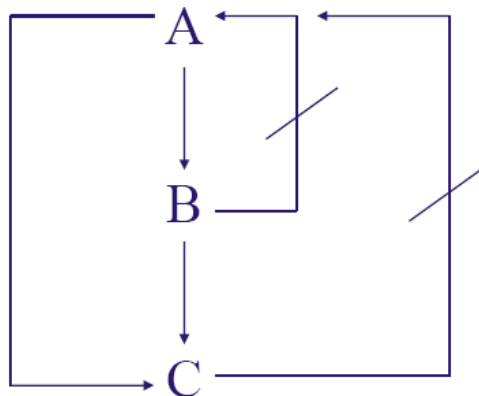
NHANVIEN

<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

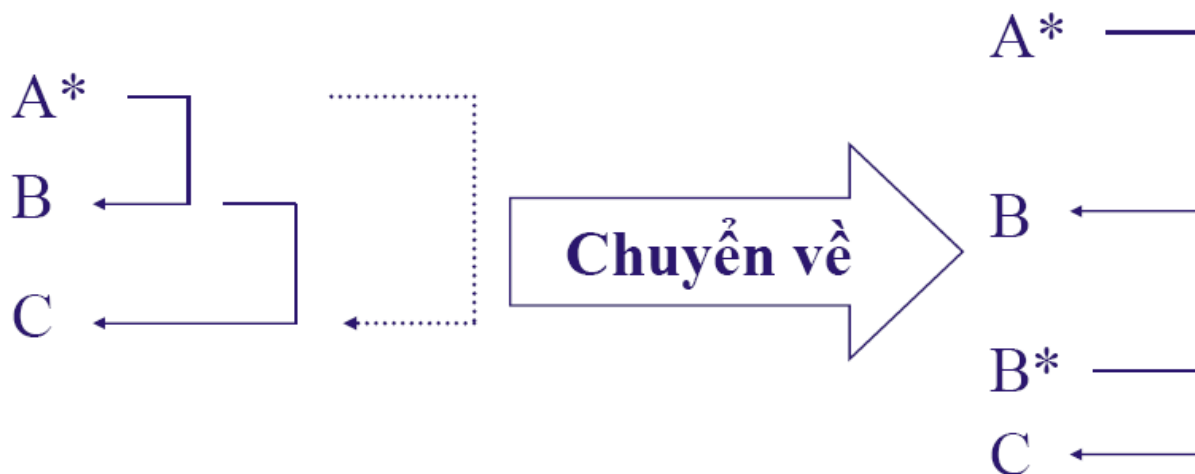
Giả sử có PTH PHG \rightarrow HSL

Quan hệ **NHANVIEN** **không đạt chuẩn 3** vì tồn tại phụ thuộc hàm bậc cao



Điều kiện để xác định phụ thuộc hàm bậc cao

Chuyển về dạng chuẩn 3



NHANVIEN

<u>MSNV</u>	TENNV	PHG
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>	HSL
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

Ví dụ 2: 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

❖ 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 \Rightarrow **Có 2 khóa**

- (MS_NHAMAY, MS_SANPHAM) \Rightarrow **Đạt dạng chuẩn 3**
- TEN_NHAMAY, MS_SANPHAM)

❖ Quan hệ SANXUAT

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

Tách thành 2 quan hệ

NHAMAY

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

NHAMAY_SANPHAM

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

⇒ Đạt dạng chuẩn BC

Các cấp chuẩn hóa quan hệ

79

Dạng chưa được chuẩn hóa



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 – 2NF



Loại bỏ các phụ thuộc bắc cầu

Dạng chuẩn thứ ba – 3NF

BÀI TẬP

Cho $R(A, B, C, D, E, G, H)$. Tập phụ thuộc hàm

$$F = \{A \rightarrow E, B \rightarrow CD, C \rightarrow BG, AE \rightarrow H\}$$

a/ Dùng các luật của hệ tiên đề ArmStrong chứng minh: $AB \rightarrow EG$

b/ Tìm bao đóng của tập thuộc tính AC, BE

c/ Tìm tất cả các khóa của lược đồ quan hệ trên.

