



HỆ CƠ SỞ DỮ LIỆU

GV: ThS.Trịnh Thị Ngọc Linh





CHƯƠNG 6. CHUẨN HOÁ

1

Mục đích của việc chuẩn hoá

2

Dư thừa thông tin và cập nhật dị thường

3

Phụ thuộc hàm

4

Chuẩn hoá lược đồ quan hệ



Mục đích của việc chuẩn hoá

- ❖ Chuẩn hoá là một kỹ thuật để tạo ra một tập hợp các quan hệ thích hợp để hỗ trợ các yêu cầu dữ liệu của một hoạt động
- ❖ Về cơ bản, các quy tắc chuẩn hoá loại bỏ các dư thừa dữ liệu và những quan hệ phụ thuộc mâu thuẫn nhau giữa các bảng



Dư thừa thông tin và cập nhật dữ thường

- ❖ Dư thừa dữ liệu là sự trùng lặp thông tin trong cơ sở dữ liệu
- ❖ Các dị thường cập nhật dữ liệu
 - **Dị thường do dữ liệu lặp:** Một số thông tin có thể được lặp lại một cách vô ích
 - **Dị thường chèn bộ:** Không thể chèn bộ mới vào quan hệ, nếu không có đầy đủ dữ liệu
 - **Dị thường xoá bộ:** Trường hợp này ngược với dị thường chèn bộ. Việc xoá bộ có thể kéo theo mất thông tin
 - **Dị thường sửa bộ:** Việc sửa đổi dữ liệu dư thừa có thể dẫn đến sự không tương thích dữ liệu



Dự thừa thông tin và cập nhật dữ thường

EMP(ENO, ENAME, TITLE, SAL, PNO, RESP, DUR)

PROJ(PNO, PNAME, BUDGET)

- ❖ Xét quan hệ EMP: tên (ENAME), chức vụ (TITLE), và lương (SAL) của nhân viên được lặp lại trong mỗi dự án mà họ tham gia **Dị thường do dữ liệu lặp**
- ❖ Xét quan hệ EMP: một nhân viên mới được nhận vào công ty và chưa được phân công vào dự án nào cả thì không thể nhập tên, chức vụ, lương của nhân viên này **Dị thường chèn bộ**



Dự thừa thông tin và cập nhật dị thường

EMP(ENO, ENAME, TITLE, SAL, PNO, RESP, DUR)

PROJ(PNO, PNAME, BUDGET)

- ❖ Xét quan hệ EMP: một nhân viên làm việc trong một dự án duy nhất. Khi dự án chấm dứt, chúng ta không thể xoá thông tin về dự án đó trong EMP được, vì nếu làm thế ta sẽ mất luôn thông tin về nhân viên đó **Dị thường xoá bộ**
- ❖ Xét quan hệ EMP: Giả sử một nhân viên làm việc trong nhiều dự án. Khi có sự thay đổi về lương, rất nhiều bộ phải cập nhật sự thay đổi này **Dị thường sửa bộ**



Phụ thuộc hàm

- ❖ Cơ sở lý thuyết về chuẩn hoá dữ liệu dựa trên các **khái niệm phụ thuộc hàm và khoá của quan hệ**
- ❖ Phụ thuộc hàm là khái niệm được xây dựng để mô tả các ràng buộc trong cơ sở dữ liệu

- ❖ **Định nghĩa:**

Cho lược đồ quan hệ $R=(A_1, A_2, \dots, A_n)$

và X, Y là các tập con của $\{A_1, A_2, \dots, A_n\}$

Ta nói rằng X xác định hàm Y , hay Y phụ thuộc hàm X , ký hiệu $X \twoheadrightarrow Y$, nếu mọi quan hệ bất kỳ r của lược đồ R thoả mãn:

$$u, v \quad r : u(X) = v(X) \quad u(Y) = v(Y)$$



Phụ thuộc hàm

Ví dụ:

Lược đồ quan hệ DMVT(MaVT, TenVT, DonGia) có phụ thuộc hàm:

MaVT TenVT, DonGia

Ví dụ:

Lược đồ quan hệ CTVT(SoCT, Khach, Hang, SoLuong) có phụ thuộc hàm:

SoCT Khach

SoCT, Khach, Hang SoLuong



Phụ thuộc hàm

Ví dụ: Xét các quan hệ:

EMP(ENO, ENAME, TITLE, SAL, PNO, RESP, DUR)

PROJ(PNO, PNAME, BUDGET)

- Đối với quan hệ PROJ: Ta có thể chấp nhận rằng mỗi dự án có tên và kinh phí xác định

PNO PNAME, BUDGET

- Trong quan hệ EMP ta có

ENO, PNO ENAME, TITLE, SAL, RESP, DUR

ENO ENAME, TITLE, SAL

- Chúng ta có thể cho rằng lương của mỗi chức vụ là cố định, do đó sẽ tồn tại phụ thuộc hàm

TITLE SAL



Các qui tắc phụ thuộc hàm

❖ Hệ tiên đề Armstrong cho các phụ thuộc hàm

Cho $\Omega := \{A_1, A_2, \dots, A_n\}$ là tập thuộc tính khác rỗng

Gọi F là tập các phụ thuộc hàm thỏa trên các quan hệ R trên tập các thuộc tính Ω

Khi đó nếu $\forall A, B, C, D \subseteq \Omega$ thì

- Phản xạ: Nếu với mọi $B \subseteq A \Rightarrow A \rightarrow B$
- Gia tăng: Nếu $A \rightarrow B \Rightarrow AC \rightarrow B, AC \rightarrow BC$
- Bắc cầu: Nếu $A \rightarrow B$ và $B \rightarrow C$ thì suy ra $A \rightarrow C$
- Giả bắc cầu: Nếu $A \rightarrow B$ và $BC \rightarrow Z \Rightarrow AC \rightarrow Z$
- Hợp: Nếu $A \rightarrow B$ và $A \rightarrow C \Rightarrow A \rightarrow BC$
- Tách: Nếu $A \rightarrow BC \Leftrightarrow A \rightarrow B$ và $A \rightarrow C$



Các qui tắc phụ thuộc hàm

❖ Các tính chất của phụ thuộc hàm

- Tính phản xạ: Nếu $B \subseteq A$ khi đó $A \rightarrow B$
- Tính gia tăng: Nếu $A \rightarrow B$ và $C \subseteq \Omega$ khi đó $AC \rightarrow BC$
- Tính bắc cầu: Nếu $A \rightarrow B$ và $B \rightarrow C$ khi đó $A \rightarrow C$
- Quy tắc hợp: Nếu $A \rightarrow B$ và $A \rightarrow C$ khi đó $A \rightarrow BC$
- Quy tắc tách: Nếu $A \rightarrow B$ và $C \subseteq B$ khi đó $A \rightarrow C$



Các qui tắc phụ thuộc hàm

Ví dụ:

Cho lược đồ $R=ABC$ và $F=\{AB \rightarrow C, C \rightarrow A\}$

Hãy chứng minh rằng $BC \rightarrow ABC$

1. $C \rightarrow A$ (theo giả thiết)
2. $BC \rightarrow AB$ (luật 1 thêm B)
3. $AB \rightarrow C$ (giả thiết)
4. $AB \rightarrow ABC$ (luật 3 thêm AB)
5. $BC \rightarrow ABC$ (luật bắc cầu từ 2 đến 4)



Các qui tắc phụ thuộc hàm

Ví dụ:

Cho $\{AB \rightarrow E, AG \rightarrow I, BE \rightarrow I, E \rightarrow G, GI \rightarrow H\}$

Chứng minh $AB \rightarrow GH$

1. $AB \rightarrow E; E \rightarrow G \Rightarrow AB \rightarrow G$

2. $AB \rightarrow G \wedge AB \rightarrow AG \text{ mà } AG \rightarrow I \Rightarrow AB \rightarrow I$

$AB \rightarrow G, AB \rightarrow I \Rightarrow AB \rightarrow GI, \text{ mà } GI \rightarrow H \Rightarrow AB \rightarrow H$

Từ (1) và (2): $AB \rightarrow GH$



Suy diễn lô-gíc

❖ Định nghĩa:

Giả sử F là tập các phụ thuộc hàm trên lược đồ quan hệ R ,
 X và Y là các tập con thuộc tính của R

Ta nói rằng F suy diễn lôgic phụ thuộc hàm $X \twoheadrightarrow Y$ hay phụ
thuộc hàm $X \twoheadrightarrow Y$ được suy diễn lôgic từ F

Ký hiệu $F \models X \twoheadrightarrow Y$

nếu mọi quan hệ r thoả các phụ thuộc hàm trong F cũng
thoả phụ thuộc hàm $X \twoheadrightarrow Y$

Ví dụ: $\{A \twoheadrightarrow B, B \twoheadrightarrow C\} \models A \twoheadrightarrow C$



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 suy diễn logic từ F :

$$F^+ = \{X \rightarrow Y \mid F \models X \rightarrow Y\}$$

- ❖ **Ví dụ:**

Cho $F = \{A \rightarrow B, B \rightarrow C, A \rightarrow D, B \rightarrow D\}$. Tìm F^+ ?

- Từ $A \rightarrow B, B \rightarrow C$, suy ra $A \rightarrow C \in F^+$
- Vì $B \rightarrow C$ và $B \rightarrow D$, suy ra $B \rightarrow DC \in F^+$
- Vì $A \rightarrow B$ và $A \rightarrow C \in F^+$, suy ra $A \rightarrow BC \in F^+$
- Vì $A \rightarrow B$ và $A \rightarrow D$, suy ra $A \rightarrow BD \in F^+$
- Vì $A \rightarrow B$ và $B \rightarrow D$, suy ra $A \rightarrow D \in F^+$



Bao đóng của tập phụ thuộc hàm

❖ Ví dụ:

Cho $F = \{A \rightarrow B, C \rightarrow X, BX \rightarrow Z\}$. Khi đó $AC \rightarrow Z \in F^+ ?$

- Vì $A \rightarrow B \Rightarrow AX \rightarrow BX$

- Từ $AX \rightarrow BX$, kết hợp $BX \rightarrow Z$, suy ra $AX \rightarrow Z$

- Từ $C \rightarrow X \Rightarrow AC \rightarrow AX$

- Áp dụng tính chất bắc cầu, $AC \rightarrow AX$ và $AX \rightarrow Z$

suy ra $AC \rightarrow Z \in F^+$



Bao đóng của tập phụ thuộc hàm

❖ Ví dụ:

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

Chứng tỏ rằng $A \rightarrow D \in F^+$?

- Vì $C \subset B$, áp dụng tính chất phản xạ, suy ra $B \rightarrow C$
- Từ $A \rightarrow B$ và $B \rightarrow C$ suy ra $A \rightarrow C$
- Từ $A \rightarrow C$ và $C \rightarrow D$ suy ra $A \rightarrow D \in F^+$



Bao đóng của tập thuộc tính

❖ Bao đóng của tập thuộc tính X R (đối với tập phụ thuộc hàm F), ký hiệu là X^+ , là tập hợp tất cả các thuộc tính phụ thuộc hàm vào X : $X^+ = \{A \mid X \twoheadrightarrow A \mid F^+\}$

❖ Ví dụ:

Cho $R=(A,B,C)$

$F = \{A \twoheadrightarrow B, B \twoheadrightarrow C\}$

Khi đó $B^+ = \{B,C\}$



Bao đóng của tập thuộc tính

❖ Ví dụ:

Cho bảng Chứng từ vật tư có các trường như sau
CTVT(A, B, C, D, E, F)

Và các phụ thuộc hàm:

A B, C

C D

A, C, E F

Với tập thuộc tính $X = \{A, C, E\}$ thì:

$X^+ = \{A, B, C, D, E, F\} = \text{CTVT}$



Thuật toán tìm bao đóng

- ❖ Đầu vào: Tập các thuộc tính R , tập các phụ thuộc hàm F trên R và tập $X \subseteq R$
- ❖ Đầu ra: X^+ (Bao đóng X^+ của X đối với F)
- ❖ Phương pháp: Ta tính lần lượt dãy các tập thuộc tính X_0, X_{i-1}, \dots, X_n như sau:
 - Đặt $X_0 = X$
 - Tính X_i như sau: $X_i = X_{i-1} \cup A$ nếu có $X_{i-1} \not\models A$, nếu không $X_i = X_{i-1}$
 - Kiểm tra điều kiện kết thúc: $X_i = R$ hoặc không có phụ thuộc hàm nào thỏa mãn



Thuật toán tìm bao đóng

❖ Ví dụ:

Cho $R=ABCDEF$

$F = \{A \ BC, C \ D, AC \ F\}$, $X = ACE$. Hãy tính X^+

Ta có:

$X_0 = ACE$

$X_1 = ACEB$ vì $A \ BC$

$X_2 = ABCED$ vì $C \ D$

$X_3 = ABCDE$ vì $ACE \ F$

Vậy $X^+ = ABCDEF$



Khóa và siêu khóa

- ❖ Cho lược đồ quan hệ $R=(A_1, \dots, A_n)$ và tập phụ thuộc hàm F trên R .
 - Tập con $X \subseteq \{A_1, \dots, A_n\}$ là khóa của R nếu $X \twoheadrightarrow A_1, \dots, A_n$
 F^+ là phụ thuộc hàm nguyên tố.
 - Tập $S \subseteq \{A_1, \dots, A_n\}$ là siêu khóa của R nếu S chứa khóa.

❖ Ví dụ:

Xét lược đồ quan hệ $R=(A, B, C)$

với tập phụ thuộc hàm $F=\{A \twoheadrightarrow B, B \twoheadrightarrow C\}$

Ta có khóa duy nhất là (A) , vì $A \twoheadrightarrow (A, B, C)$. Mọi tập thuộc tính chứa A là siêu khóa



Khóa và siêu khóa

❖ Ví dụ:

Cho $R=ABCDEG$

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

$K= ABD$ là siêu khoá của R

$(ABD)^+ = ABDGEC$



Phép tách lược đồ quan hệ

❖ **Định nghĩa:** Cho lược đồ quan hệ $R = A_1A_2\dots A_n$.

Tách lược đồ quan hệ R là thay thế R bằng các lược đồ con R_1, R_2, \dots, R_m sao cho $R_1 \quad R_2 \quad \dots \quad R_m = R$

và $R_i \neq R_j$ khi $i \neq j$



Phép tách bảo toàn thông tin

❖ $(R) = (R_1, R_2, \dots, R_m)$ bảo toàn thông tin

$$r(R) = r_{R_1}(r) * r_{R_2}(r) * \dots * r_{R_m}(r)$$

SoCT	Ngày	Khach	Hang	SoLuong
01	20/3	K01	H01	4
02	21/3	K03	H01	5
03	21/3	K03	H04	2

Tách

Nối

$\pi(r)$
Tiêu đề

SoCT	Ngày	Khach
01	20/3	K01
02	21/3	K03

$\pi(r)$
Chi tiết

SoCT	Hang	SoLuong
01	H01	4
02	H01	5
03	H04	2



Thuật toán kiểm tra phép tách bảo toàn thông tin

- ❖ Đầu vào: $R = A_1A_2\dots A_n$ và $(R) = (R_1, R_2, \dots, R_m)$
- ❖ Đầu ra: (R) bảo toàn thông tin hay không?
- ❖ Phương pháp:
 - Bước 1:
 - Lập bảng gồm m dòng và n cột. Dòng thứ i tương ứng lược đồ con R_i , cột thứ j tương ứng thuộc tính A_j
 - Tại vị trí (i,j) ta ký hiệu a_j nếu $A_j \in R_i$, ngược lại ký hiệu $b(i,j)$
 - Bước 2: Dựa vào các phụ thuộc hàm để làm bảng theo nguyên tắc: Xét $X \rightarrow Y$, nếu trên các dòng mà giá trị X bằng nhau ưu tiên cho ký hiệu a_j
 - Lặp lại bước 2 cho đến khi
 - Có một dòng chứa toàn ký hiệu a_j . Khi đó kết luận (R) bảo toàn thông tin
 - Không áp dụng được phụ thuộc hàm nào nữa. Khi đó kết luận (R) mất thông tin



Thuật toán kiểm tra phép tách bảo toàn thông tin

❖ Ví dụ:

Cho $R = ABCDE$ và $F = \{A \rightarrow BC, ACD \rightarrow E\}$

$(R) = (ABC, ADE)$ có bảo toàn thông tin hay không?

	A	B	C	D	E
ABC	a_1	a_2	a_3	b_{14}	b_{15}
ADE	a_1	b_{22}	b_{23}	a_4	a_5

$A \rightarrow BC$
→

	A	B	C	D	E
ABC	a_1	a_2	a_3	b_{14}	b_{15}
ADE	a_1	a_2	a_3	a_4	a_5

Vậy (R) bảo toàn thông tin



Thuật toán kiểm tra phép tách bảo toàn thông tin

❖ Ví dụ:

Cho $R = ABCD$ và $F = \{A \rightarrow B, AC \rightarrow D\}$

$(R) = (AB, ACD)$ có bảo toàn thông tin hay không?

	A	B	C	D
AB	a_1	a_2	b_{13}	b_{14}
ACD	a_1	b_{22}	a_3	a_4

$A \rightarrow B$
 \longrightarrow

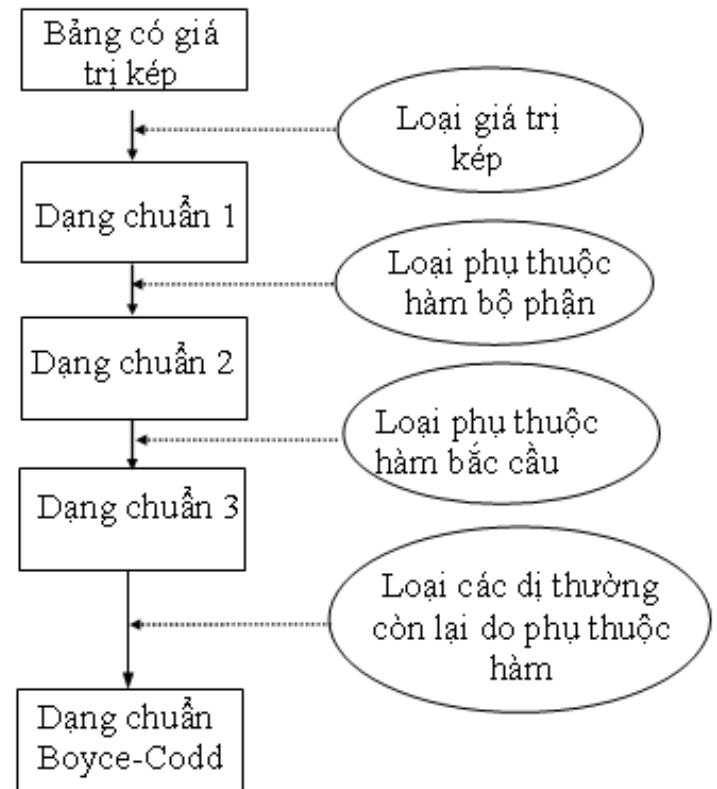
	A	B	C	D
AB	a_1	a_2	b_{13}	b_{14}
ACD	a_1	a_2	a_3	a_4

Vậy (R) bảo toàn thông tin

Qui trình chuẩn hoá

❖ Khi thiết kế và cài đặt các hệ CSDL, chuẩn hoá là quá trình khảo sát danh sách các thuộc tính và áp dụng tập các quy tắc phân tích vào danh sách đó, biến đổi chúng thành nhiều tập nhỏ hơn sao cho:

- Tối thiểu việc lặp lại
- Tránh dị thường thông tin
- Xác định và giải quyết được sự không rõ ràng, nhập nhằng trong suy diễn





Dạng chuẩn một (1NF)

- ❖ **Định nghĩa:** Một lược đồ quan hệ R được gọi là ở dạng chuẩn thứ nhất nếu và chỉ nếu toàn bộ các miền có mặt trong R đều chỉ chứa các giá trị nguyên tố (không phân chia được nữa)

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

Chưa ở dạng chuẩn 1



Dạng chuẩn một (1NF)

❖ Đưa về dạng chuẩn 1:

- Biến cột đa trị thành đơn trị
- Điền đủ dữ liệu vào các cột khác

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



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



Dạng chuẩn thứ 2 (2NF)

❖ Giả sử K là khóa của lược đồ R

Khi đó mọi thuộc tính không khóa A của R đều phụ thuộc hàm vào khóa K : $K \twoheadrightarrow A$

Nếu A không phụ thuộc đầy đủ vào K thì tồn tại tập con thực sự H của K xác định A , tức $H \twoheadrightarrow A$. Khi đó phụ thuộc hàm $H \twoheadrightarrow A$ gọi là phụ thuộc hàm bộ phận

❖ **Định nghĩa:** Một lược đồ quan hệ R là ở dạng chuẩn thứ 2 nếu nó ở dạng chuẩn thứ 1 và không có phụ thuộc hàm bộ phận, tức là mọi thuộc tính không khóa đều phụ thuộc đầy đủ vào các khóa của lược đồ



Dạng chuẩn thứ 2 (2NF)

❖ Chú ý:

- Chỉ kiểm tra các quan hệ có đạt 2NF nếu quan hệ đó có khoá chính gồm 2 thuộc tính trở lên
- Để chuyển quan hệ từ dạng 1NF sang dạng 2NF, chúng ta dùng phép chiếu



Dạng chuẩn thứ 2 (2NF)

❖ **Ví dụ:** Xét các lược đồ quan hệ sau:

EMP(ENO, ENAME, TITLE, SAL, PNO, RESP, DUR)

PROJ(PNO, PNAME, BUDGET)

Lược đồ của EMP có khóa là (ENO, PNO)

- Phụ thuộc hàm ENO ENAME, TITLE là phụ thuộc hàm bộ phận vì vế phải là tập con thực sự của khóa. Vậy EMP không ở dạng chuẩn thứ 2

- Lược đồ của PROJ không có phụ thuộc hàm bộ phận, vậy nó ở dạng chuẩn 2



Dạng chuẩn thứ 2 (2NF)

❖ Ví dụ: Bảng R có các phụ thuộc hàm sau:

MF → Tenfim, NSX, Giathue, HSX, NPP

MaKH → TenKH, Diachi

MF, MaKH → Ngaydat

Khóa chính: MF, MaKH.

Các thuộc tính Tenfim, Giathue, TenKH, Diachi,...là các thuộc tính không khóa, chỉ phụ thuộc vào một bộ phận của khóa

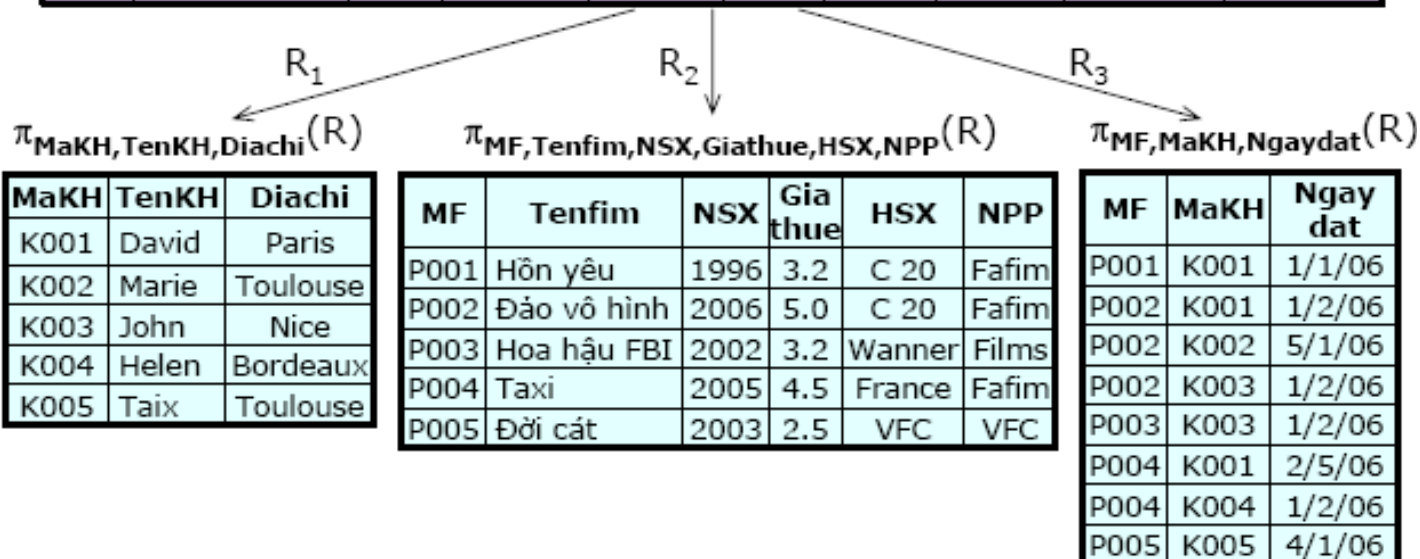
→ R không đạt chuẩn 2



Dạng chuẩn thứ 2 (2NF)

❖ Để chuyển về dạng chuẩn 2, sử dụng phép chiếu:

MF	Tenfim	NSX	Giathue	HSX	NPP	MaKH	TenKH	Điachi	Ngaydat
P001	Hôn yêu	1996	3.2	C 20	Fafim	K001	David	Paris	1/1/06
P002	Đảo vô hình	2006	5.0	C 20	Fafim	K001	David	Paris	1/2/06
P002	Đảo vô hình	2006	5.0	C 20	Fafim	K002	Marie	Toulouse	5/1/06
P002	Đảo vô hình	2006	5.0	C 20	Fafim	K003	John	Nice	1/2/06
P003	Hoa hậu FBI	2002	3.2	Wanner	Films	K003	John	Nice	1/2/06
P004	Taxi	2005	4.5	France	Fafim	K001	David	Paris	2/5/06
P004	Taxi	2005	4.5	France	Fafim	K004	Helen	Bordeaux	1/2/06
P005	Đời cát	2003	2.5	VFC	VFC	K005	Taix	Toulouse	4/1/06





Dạng chuẩn thứ 3 (3NF)

- ❖ Phụ thuộc hàm $X \twoheadrightarrow A$ gọi là phụ thuộc hàm bắc cầu, nếu nó là phụ thuộc hàm nguyên tố, A là thuộc tính không khóa, $A \not\rightarrow X$, và X chứa thuộc tính không khóa. Khi đó với mọi khóa K ta có các phụ thuộc hàm không tầm thường $K \rightarrow X$ & $X \rightarrow A$. Mặt khác không thể có $X \rightarrow K$ vì X chứa các thuộc tính không khóa và không chứa khóa (vì $X \twoheadrightarrow A$ là nguyên tố)
- ❖ **Định nghĩa:** Một lược đồ quan hệ gọi là ở dạng chuẩn thứ 3 nếu nó ở dạng chuẩn thứ 2 và không có phụ thuộc hàm bắc cầu



Dạng chuẩn thứ 3 (3NF)

❖ Ví dụ:

Xét lược đồ quan hệ EMP(ENO, ENAME, TITLE, SAL, PNO, RESP, DUR)

Lược đồ của quan hệ có TITLE SAL là phụ thuộc hàm bậc cầu. Vậy EMP không ở dạng chuẩn thứ 3

Lược đồ của quan hệ PROJ(PNO, PNAME, BUDGET) không có phụ thuộc hàm bậc cầu, vậy nó ở dạng chuẩn 3



Thuật toán đưa về dạng chuẩn 3 bảo toàn thông tin

❖ Thuật toán 1:

- Đầu vào: $\langle R, F \rangle$
- Đầu ra: (R) thỏa 3NF bảo toàn thông tin
- Phương pháp:
 - Bước 1: Loại bỏ trong R những thuộc tính không thuộc về phụ thuộc hàm nào
 - Bước 2: Thu gọn các phụ thuộc hàm
Nếu $X \rightarrow A_1, X \rightarrow A_2, \dots, X \rightarrow A_n$ Thì $X \rightarrow A_1A_2\dots A_n$
 - Bước 3: Mọi phụ thuộc hàm chuyển thành một lược đồ con

❖ Ví dụ:

$R=ABCDEFGHIJ$ và $F=\{A \rightarrow BC, D \rightarrow AF, DG \rightarrow H, G \rightarrow IJ\}$

$A \rightarrow BC \quad R_1=ABC$

$D \rightarrow AF \quad R_2=DAF$

$DG \rightarrow H \quad R_3=DGH$

$G \rightarrow IJ \quad R_4=GIJ$

$(R) = (ABC, DAF, DGH, GIJ)$



Thuật toán đưa về dạng chuẩn 3 bảo toàn thông tin

❖ Thuật toán 2:

- Đầu vào: $\langle R, F \rangle$
- Đầu ra: (R) thoả 3NF bảo toàn thông tin
- Phương pháp:
 - Bước 1: Tìm khoá của R và giả sử F là đầy đủ và không dư thừa
 - Bước 2: Nếu $X \twoheadrightarrow A$ và X không chứa khoá của R : $R = (XA, R \setminus A)$
Lặp lại bước 2 với $R \setminus A$ cho đến khi không tách được

❖ Ví dụ:

$R = MTGPSL, F = \{M \twoheadrightarrow T, GP \twoheadrightarrow M, GT \twoheadrightarrow P, MS \twoheadrightarrow L, GS \twoheadrightarrow P\}$

Khoá của R là GS

Ta có:

$M \twoheadrightarrow T \quad R_1 = MT, R = MGPSL$

$MS \twoheadrightarrow L \quad R_2 = MSL, R = MGPS$

$GP \twoheadrightarrow M \quad R_{23} = GPM, R = GPS$

$(R) = (MT, MSL, GPM, GPS)$



Dạng chuẩn Boyce-Codd (BCNF)

- ❖ Định nghĩa: Lược đồ quan hệ $s = \langle R, F \rangle$ được gọi là lược đồ dạng chuẩn Boyce - Codd (BCNF), nếu với mọi phụ thuộc $X \rightarrow Y \in F^+$, thì khi đó hoặc $Y \subseteq X$ (phụ thuộc tầm thường), hoặc X là một khoá của lược đồ quan hệ. Tức là nếu $X \rightarrow Y \in F^+$, $Y \not\subseteq X$ thì $X^+ = R$
- ❖ Từ định nghĩa trên có thể suy ra rằng:
 - Các thuộc tính không khoá phụ thuộc hoàn toàn vào khoá
 - Các thuộc tính khoá phụ thuộc hoàn toàn vào tất cả khoá khác



Dạng chuẩn Boyce-Codd (BCNF)

❖ Ví dụ:

Lược đồ của quan hệ PROJ(PNO, PNAME, BUDGET) chỉ có phụ thuộc hàm duy nhất PNO → (PNAME, BUDGET), vậy nó ở dạng chuẩn Boyce-Codd

❖ Chú ý:

- Một quan hệ ở BCNF thì cũng đạt 3NF
- Trong thực hành các quan hệ đạt chuẩn 3NF là đủ. Tuy nhiên một quan hệ ở 3NF không đảm bảo đã loại bỏ được tất cả các lỗi khi thao tác dữ liệu



Thuật toán đưa về dạng chuẩn Boyce-Codd bảo toàn thông tin

- ❖ Đầu vào: $\langle R, F \rangle$
- ❖ Đầu ra: (R) thoả BCNF bảo toàn thông tin
- ❖ Phương pháp:
 - Phương pháp chủ yếu của thuật toán là tách lược đồ $s = \langle R, F \rangle$ thành 2 lược đồ
 - Chọn bất kỳ $X \rightarrow A \in F^+$ sao cho X không là khoá và $A \not\subseteq X$. Khi đó lược đồ có tập các thuộc tính XA sẽ có dạng chuẩn BCNF và phụ thuộc hàm $X \rightarrow A$ sẽ thoả trên nó. Lược đồ thứ 2 có tập các thuộc tính $R \setminus A$. Hiển nhiên, khi kết nối lược đồ có tập thuộc tính $R \setminus A$ với lược đồ có tập thuộc tính XA không tổn thất thông tin. Tiếp tục tách $R \setminus A$ cho đến trở thành lược đồ có dạng chuẩn BCNF



Ví dụ:

Cho $\Omega = \text{CTHRSG}$, trong đó:

C : Khoá học, T: Thầy giáo, H: Giờ học R: Phòng học, S : Sinh viên G:
Lớp

Biết rằng:

- Mỗi khoá học chỉ có một thầy dạy
- Một phòng học tại giờ xác định chỉ có một khoá học
- Thầy dạy tại giờ học cụ thể xác định phòng học cụ thể
- Khoá học với một sinh viên cụ thể xác định lớp học cụ thể
- Mỗi một sinh viên học trong một giờ xác định tại phòng học cụ thể

Khi đó $F = \{C \rightarrow T, HR \rightarrow C, HT \rightarrow R, CS \rightarrow G, HS \rightarrow R\}$

Hiển nhiên, $s = \langle \Omega, F \rangle$ không là Boyce Codd, khoá của nó là thuộc tính HS

- ❖ Bước 1: Xét CS \rightarrow G: CS không phải là khóa, có thể tách $s = \langle \Omega, F \rangle$ thành 2 lược đồ quan hệ có dạng như sau:

$s_1 = \langle \Omega_1, F_1 \rangle$	$s_2 = \langle \Omega_2, F_2 \rangle$
$\Omega_1 = \{ \underline{C}, \underline{S}, G \}$	$\Omega_2 = \{ C, T, \underline{H}, \underline{S}, R \}$
$F_1 = \{ CS \rightarrow G \}$	$F_2 = \{ C \rightarrow T, HR \rightarrow C, HT \rightarrow R, HS \rightarrow R \}$.

$s_1 = \langle \Omega_1, F_1 \rangle$ ở dạng Boyce Codd, $s_2 = \langle \Omega_2, F_2 \rangle$ ở dạng 3NF nhưng vẫn chưa ở dạng Boyce Codd



- ❖ Bước 2: Xét $C \rightarrow T$: T không phải là thuộc tính khóa, tách $S_2 = \langle \Omega_2, F_2 \rangle$ thành 2 lược đồ quan hệ sau:

$S_{21} = \langle \Omega_{21}, F_{21} \rangle$	$S_{22} = \langle \Omega_{22}, F_{22} \rangle$
$\Omega_{21} = \{ \underline{C}, T \}$	$\Omega_{22} = \{ C, \underline{H}, \underline{S}, R \}$
$F_{21} = \{ C \rightarrow T \}$	$F_{22} = \{ HR \rightarrow C, HS \rightarrow R \}$



- ❖ Bước 3: Xét $HR \rightarrow C$: HR không phải là thuộc tính khóa, tách $s_{22} = \langle \Omega_{22}, F_{22} \rangle$ thành 2 lược đồ quan hệ sau:

$S_{221} = \langle \Omega_{221}, F_{221} \rangle$	$s_{222} = \langle \Omega_{222}, F_{222} \rangle$
$\Omega_{221} = \{ C, \underline{H}, R \}$	$\Omega_{222} = \{ \underline{H}, S, R \}$
$F_{221} = \{ HR \rightarrow C \}$	$F_{222} = \{ HS \rightarrow R \},$



❖ Kết quả:

$$\Omega_1 = \{ \underline{C}, \underline{S}, \underline{G} \}, F1 = \{ CS \rightarrow G \}$$

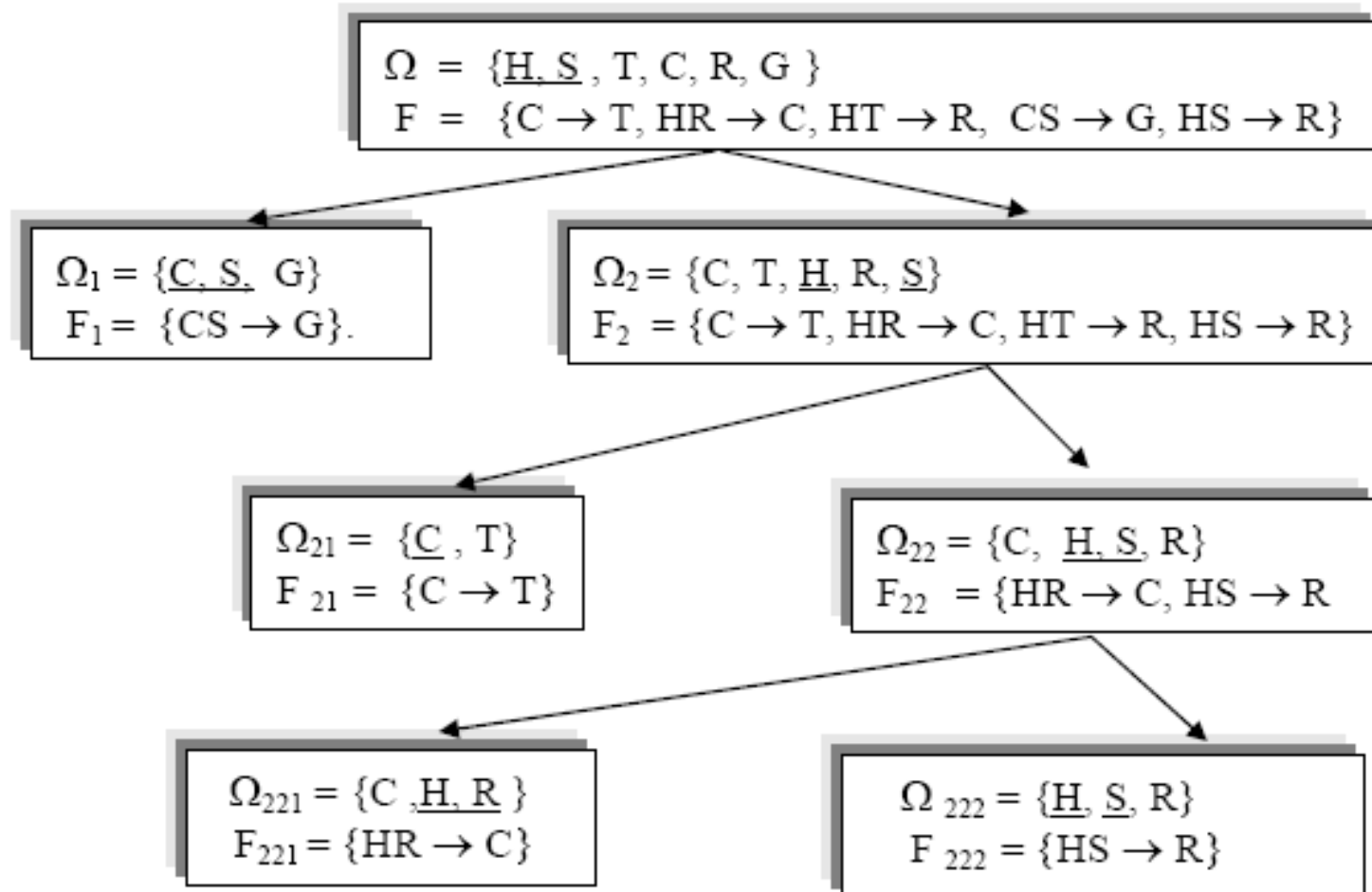
$$\Omega_{21} = \{ \underline{C}, T \}, F21 = \{ C \rightarrow T \}$$

$$\Omega_{221} = \{ C, \underline{H}, \underline{R} \}, F221 = \{ HR \rightarrow C \}$$

$$\Omega_{222} = \{ \underline{H}, \underline{S}, R \}, F222 = \{ HS \rightarrow R \}$$



Thuật toán đưa về dạng chuẩn Boyce-Codd bảo toàn thông tin





Sơ đồ chuẩn hoá

Lược đồ quan hệ



Tìm tập các phụ thuộc hàm

(Dựa vào các thông tin có được và các quy tắc suy diễn)



Tìm khoá

(Dựa vào bao đóng của tập thuộc tính)



Đưa về dạng chuẩn 2

(Loại các phụ thuộc hàm bộ phận, kiểm tra tách có bảo toàn thông tin hay không)



Đưa về dạng chuẩn 3 bảo toàn thông tin

(Loại các phụ thuộc hàm bắc cầu)

Q (ABCDEG)



$F = \{A \rightarrow BC, C \rightarrow DE, E \rightarrow G\}$



Khoá là AC

vì $(AC)^+ = ABCDEG$



Q1(ABC)

Q2(CDEG)

$F = \{A \rightarrow BC\}$

$F2 = \{C \rightarrow DE, E \rightarrow G\}$



R2(CDE)

R3(EG)

$F = \{C \rightarrow DE\}$

$F = \{E \rightarrow G\}$

(Thỏa Boyce Codd)



Tìm khoá

$X_0 = AC$

Vì $A \rightarrow BC$ nên $X_1 = ABC$

Vì $C \rightarrow DE$ nên $X_2 = ABCE$

Vì $E \rightarrow G$ nên $X_3 = ABCEG$

Vì $(AC)^+ = ABCEG$ nên AC là khoá của lược đồ



Kiểm tra tách bảo toàn thông tin

	A	B	C	D	E	G
ABC	a1	a2	a3	b14	b15	b16
CDEG	b21	b22	a3	a4	a5	a6

↓ C → DE

	A	B	C	D	E	G
ABC	a1	a2	a3	a4	a5	b16
CDEG	b21	b22	a3	a4	a5	a6

↓ E → G

	A	B	C	D	E	G
ABC	a1	a2	a3	a4	a5	a6
CDEG	b21	b22	a3	a4	a5	a6



Đưa về dạng chuẩn 3

- ❖ Từ $Q(ABCDEG)$ tách thành $R1(\underline{A}BC)$ và $R2(\underline{C}DEG)$
- ❖ $R2(CDEG)$ chưa đạt chuẩn 3 do có phụ thuộc hàm bậc cầu $C \rightarrow DE, E \rightarrow G$
- ❖ Đưa $R2(CDEG)$ về chuẩn 3:

$R2(CDEG)$ có $F = \{C \rightarrow DE, E \rightarrow G\}$

$C \rightarrow DE$ $Q1(\underline{C}DE)$

$E \rightarrow G$ $Q2(\underline{E}G)$



Kết quả

