

# CƠ SỞ DỮ LIỆU

( Databases )

## Chương 5: Dạng chuẩn và Chuẩn hóa

# Nội dung

1. Các dạng chuẩn
2. Phân rã lược đồ quan hệ
3. Chuẩn hóa lược đồ CSDL
4. Bài tập

# 5.1. Các dạng chuẩn

## ❑ Chuẩn hóa là gì?

- Chuẩn hóa là kỹ thuật dùng để tạo ra một tập các quan hệ có các đặc điểm mong muốn dựa vào các yêu cầu về dữ liệu của 1 enterprise
- Chuẩn hóa là 1 cách tiếp cận từ dưới lên (bottom-up approach) để thiết kế CSDL, bắt đầu từ các mối liên hệ giữa các thuộc tính

## ❑ Mục đích của chuẩn hóa

- Loại bỏ các bất thường của 1 quan hệ để có được các quan hệ có cấu trúc tốt hơn, nhỏ hơn

## ❑ Quan hệ có cấu trúc tốt (well-structured relation):

- Là quan hệ có sự dư thừa dữ liệu là tối thiểu và cho phép người dùng thêm, sửa, xóa mà không gây ra mâu thuẫn dữ liệu

# 5.1.1. Sự dư thừa dữ liệu

## ❑ Sự phụ thuộc giữa các thuộc tính gây ra sự dư thừa

– Ví dụ:

- Điểm các môn học → Điểm trung bình → xếp loại
- Địa chỉ → zip code

TENPHG	MAPHG	TRPHG	NG_NHANCHUC	MANV	TENNV	HONV	...
Nghien cuu	5	333445555	05/22/1988	333445555	Tung	Nguyen	...
Dieu hanh	4	987987987	01/01/1995	987987987	Hung	Nguyen	...
Quan ly	1	888665555	06/19/1981	888665555	Vinh	Pham	...

# 5.1.1.Sự dư thừa dữ liệu (tt)

❑ Thuộc tính đa trị trong lược đồ ER → nhiều bộ số liệu trong lược đồ quan hệ

❑ Ví dụ:

NHANVIEN(TENNV, HONV, NS,DCHI,GT,LUONG, BANGCAP)

TENNV	HONV	NS	DCHI	GT	LUONG	BANGCAP
Tung	Nguyen	12/08/1955	638 NVC Q5	Nam	40000	Trung học
Nhu	Le	06/20/1951	291 HVH QPN	Nu	43000	Trung học
Nhu	Le	06/20/1951	291 HVH QPN	Nu	43000	Đại học
Hung	Nguyen	09/15/1962	Ba Ria VT	Nam	38000	Thạc sỹ

# 5.1.1.Sự dư thừa dữ liệu (tt)

## □ Sự 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 thêm:

TENPHG	MAPHG	TRPHG	NG_NHANCHUC	MANV	TENNV	HONV	...
Nghien cuu	5	333445555	05/22/1988	333445555	Tung	Nguyen	...
Dieu hanh	4	987987987	01/01/1995	987987987	Hung	Nguyen	...
Quan ly	1	888665555	06/19/1981	888665555	Vinh	Pham	...

# 5.1.1.Sự dư thừa dữ liệu (tt)

## ❑ Các giá trị không xác định

- Đặt thuộc tính Trưởng phòng vào quan hệ NHANVIEN thay vì vào quan hệ PHONGBAN

## ❑ Các bộ giả

- Khi sử dụng các phép nối

# 5.1.1.Sự dư thừa dữ liệu (tt)

## □ Một số quy tắc khi thiết kế CSDL quan hệ

- **NT1**: Rõ ràng về mặt ngữ nghĩa, tránh các sự phụ thuộc giữa các thuộc tính với nhau
- **NT2**: Tránh sự trùng lặp về nội dung → đảm bảo tránh được các dị thường khi thao tác cập nhật dữ liệu
  - Phải có một số thao tác khi thêm mới và cập nhật vào lược đồ quan hệ, cũng như có thể gây sai hỏng trong trường hợp xóa bỏ các bộ
- **NT3**: Tránh sử dụng các thuộc tính có nhiều giá trị Null
  - Khó thực hiện các phép nối và kết hợp
- **NT4**: Thiết kế các lược đồ quan hệ sao cho chúng có thể được nối với điều kiện bằng trên các thuộc tính là khoá chính hoặc khoá ngoài theo cách đảm bảo không sinh ra các bộ “giả”



## 5.1.2. Dạng chuẩn

- ❑ Mỗi một dạng chuẩn là một tập các điều kiện trên lược đồ nhằm đảm bảo các tính chất của nó (liên quan tới dư thừa và bất thường trong cập nhật)
- ❑ Chuẩn hóa dữ liệu: quá trình phân tích lược đồ quan hệ dựa trên các FD và các khóa chính để đạt được
  - Cực tiểu sự dư thừa
  - Cực tiểu các phép cập nhật bất thường

# 5.1.2. Dạng chuẩn (tt)

## □ Các dạng chuẩn

- Dạng chuẩn 1 (1NF – first normal form)
- Dạng chuẩn 2 (2NF – second normal form)
- Dạng chuẩn 3 (3NF – third normal form)
- Dạng chuẩn BCNF (Boyce-Codd normal form)

# Dạng chuẩn 1

## □ Định nghĩa:

Quan hệ R được gọi là ở dạng 1NF nếu miền giá trị của một thuộc tính (bất kỳ) chỉ chứa giá trị nguyên tố đơn (đơn trị, không phân chia được) và giá trị của mỗi thuộc tính cũng là một giá trị đơn lấy từ miền giá trị của nó

## □ Ví dụ: 1 phòng ban có thể có nhiều địa điểm

PHONGBAN( MaPhg, TenPhg, **DDIEM**)

Thuộc  
tính đa trị

# Dạng chuẩn 1 (tt)

❑ Vấn đề còn tồn tại trong 1NF

❑ Xét lược đồ

DDIEM\_PHG(MaPHG, DDIEM)

– Vẫn bị lặp lại

MAPHG	DIADIEM
1	TP HCM
4	HA NOI
5	VUNGTAU
5	NHATRANG
5	TP HCM

– Có thể ẩn chứa các phụ thuộc hàm bộ phận

# Dạng chuẩn 2 (2NF)

□ **Phụ thuộc hàm đầy đủ**: Một phụ thuộc hàm  $X \rightarrow Y$  là một phụ thuộc hàm đầy đủ nếu loại bỏ bất kỳ thuộc tính  $A$  nào ra khỏi  $X$  thì phụ thuộc hàm không còn đúng nữa.

$\forall A, A \in X, (X - \{A\}) \rightarrow Y$  : là sai.

□ **Phụ thuộc hàm bộ phận**: Một phụ thuộc hàm  $X \rightarrow Y$  là phụ thuộc bộ phận nếu có thể bỏ một thuộc tính  $A \in X$ , ra khỏi  $X$  phụ thuộc hàm vẫn đúng, điều đó có nghĩa là với

$\exists A \in X, (X - \{A\}) \rightarrow Y$

# Dạng chuẩn 2 (tt)

## □ Định nghĩa dạng chuẩn 2

Một quan hệ được gọi là ở dạng chuẩn 2 (2NF) nếu:

- Thỏa mãn dạng chuẩn 1NF
- Các thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào khóa chính.

## □ Nhận xét

- Với các quan hệ có khóa là 1 thuộc tính đơn thì đương nhiên các thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào khóa.
- Chỉ cần kiểm tra các lược đồ có chứa phụ thuộc hàm bộ phận

# Dạng chuẩn 2 (tt)

## □ Ví dụ

NV\_DA(MaNV, MaDA, Sogio, TenDA, DDiemDA)

Chỉ phụ thuộc vào MaDA

FD<sub>1</sub>: MaNV, MaDA → Sogio, TenDA, DDiemDA

FD<sub>2</sub>: MaDA → TenDA

FD<sub>3</sub>: MaDA → DDiemDA

**Chưa đạt 2NF**

# Dạng chuẩn 3 (3NF)

❑ Dạng chuẩn 3NF dựa trên khái niệm phụ thuộc bắc cầu.

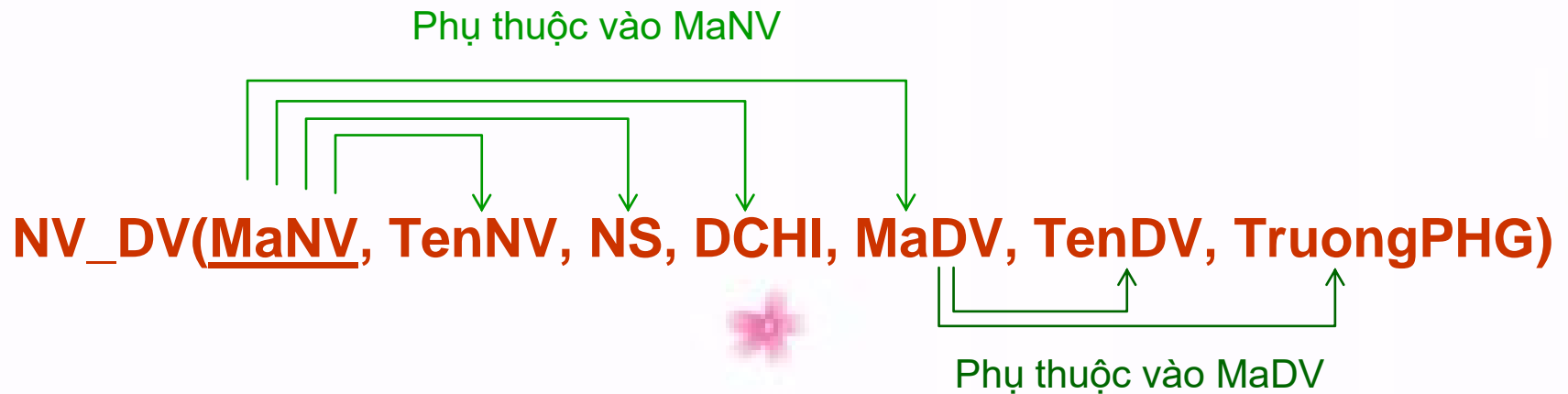
❑ Định nghĩa:

Một lược quan hệ R được coi đạt dạng chuẩn 3 (3NF) nếu nó:

- Thỏa mãn là 2NF
- Không có thuộc tính không khoá nào của R phụ thuộc bắc cầu vào khoá chính.



# Dạng chuẩn 3 (tt)



- Mọi thuộc tính trong quan hệ đều phụ thuộc vào khóa
- Có thuộc tính không khóa phụ thuộc bắc cầu vào khóa khi và chỉ khi tồn tại phụ thuộc hàm giữa các thuộc tính không khóa.

# Tóm tắt 3 dạng chuẩn

	Nhận biết	Cách chuẩn hóa
<b>1NF</b>	Quan hệ ko chứa thuộc tính đa trị hoặc thuộc tính kết hợp	Chuyển tất cả các thuộc tính đa trị hoặc thuộc tính kết hợp thành 1 quan hệ mới
<b>2NF</b>	Có thuộc tính không khóa phụ thuộc 1 phần vào khóa chính	Tách thuộc tính phụ thuộc 1 phần thành lược đồ mới, đảm bảo quan hệ với lược đồ liên quan
<b>3NF</b>	Tồn tại phụ thuộc hàm giữa các thuộc tính ko phải là khóa	Tách các thuộc tính đó thành lược đồ mới

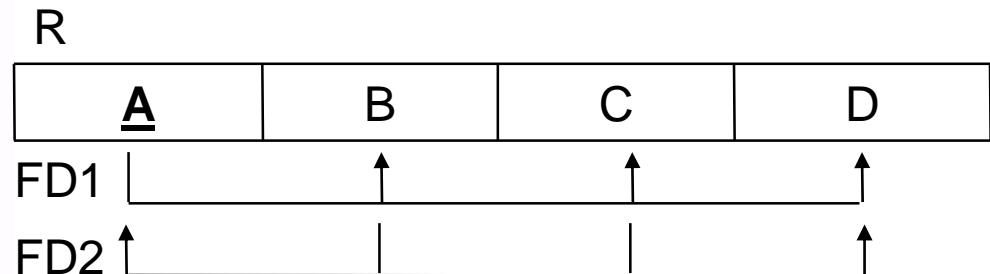
# Dạng chuẩn Boyce-Codd

- ❑ Một lược đồ quan hệ R được gọi là ở dạng chuẩn Boyce-Codd (BCNF) nếu nó
  - Thỏa mãn dạng 3NF
  - $\forall X \rightarrow Y \in F^+$  thì X là siêu khóa (chứa khóa của quan hệ) hoặc  $Y \subset X$ . Nói cách khác, quan hệ đạt R sẽ không đạt BCNF nếu tồn tại phụ thuộc hàm mà vế trái không phải là khóa

## ❑ Ví dụ 1

f1:  $A \rightarrow BCD$

f2:  $BC \rightarrow AD$



# Dạng chuẩn Boyce-Codd(tt)

## □ Ví dụ 2:

Cho quan hệ R (A1,A2,A3,A4,A5)

Với các phụ thuộc hàm:

$$f_1: A1, A2 \rightarrow A3, A4, A5$$

$$f_2: A4 \rightarrow A2$$

- R đạt dạng chuẩn 3NF (Không có thuộc tính không khóa phụ thuộc bắc cầu vào khóa)
- R chưa đạt BCNF vì  $f_2: A4 \rightarrow A2$  có vế trái không phải siêu khóa

# Dạng chuẩn Boyce-Codd(tt)

□ **Ví dụ 3:** Cho quan hệ  
SV\_MH\_GV(MaSV, MONHOC, GIANGVIEN)

Với các phụ thuộc hàm:

$f_1: \text{MaSV} \rightarrow \text{MonHoc}, \text{GiangVien}$

$f_2: \text{MonHoc} \rightarrow \text{Giangvien}$

- R chưa đạt BCNF vì  $f_2$  có vế trái không phải siêu khóa

# Dạng chuẩn Boyce-Codd(tt)

- ❑ Nếu một lược đồ quan hệ không thỏa mãn điều kiện BCNF, thủ tục chuẩn hóa bao gồm:
  - Loại bỏ các thuộc tính khóa phụ thuộc hàm vào thuộc tính không khóa ra khỏi quan hệ
  - tách chúng thành một quan hệ riêng có khoá chính là thuộc tính không khóa gây ra phụ thuộc.

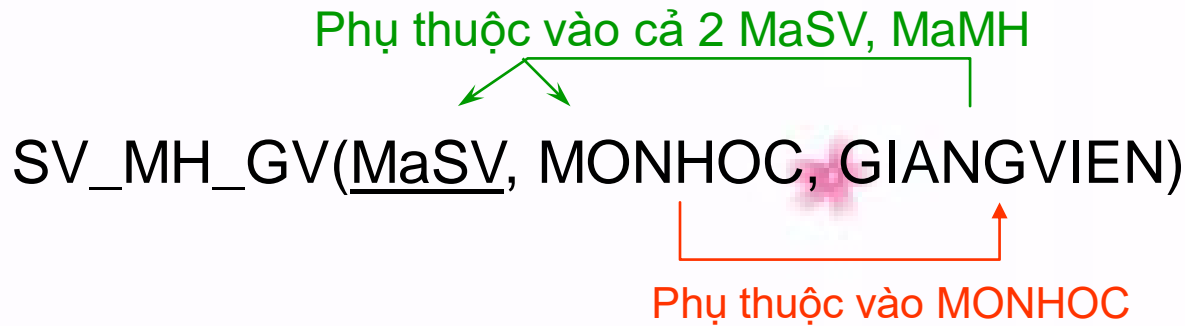
- ❑ Ví dụ trên: R (A1, A2, A3, A4, A5)  
Với các phụ thuộc hàm:

- $A1, A2 \rightarrow A3, A4, A5$
- $A4 \rightarrow A2$

- ❑ lược đồ được tách ra như sau:
  - R1( A4, A2)
  - R2(A1, A4, A3, A5)

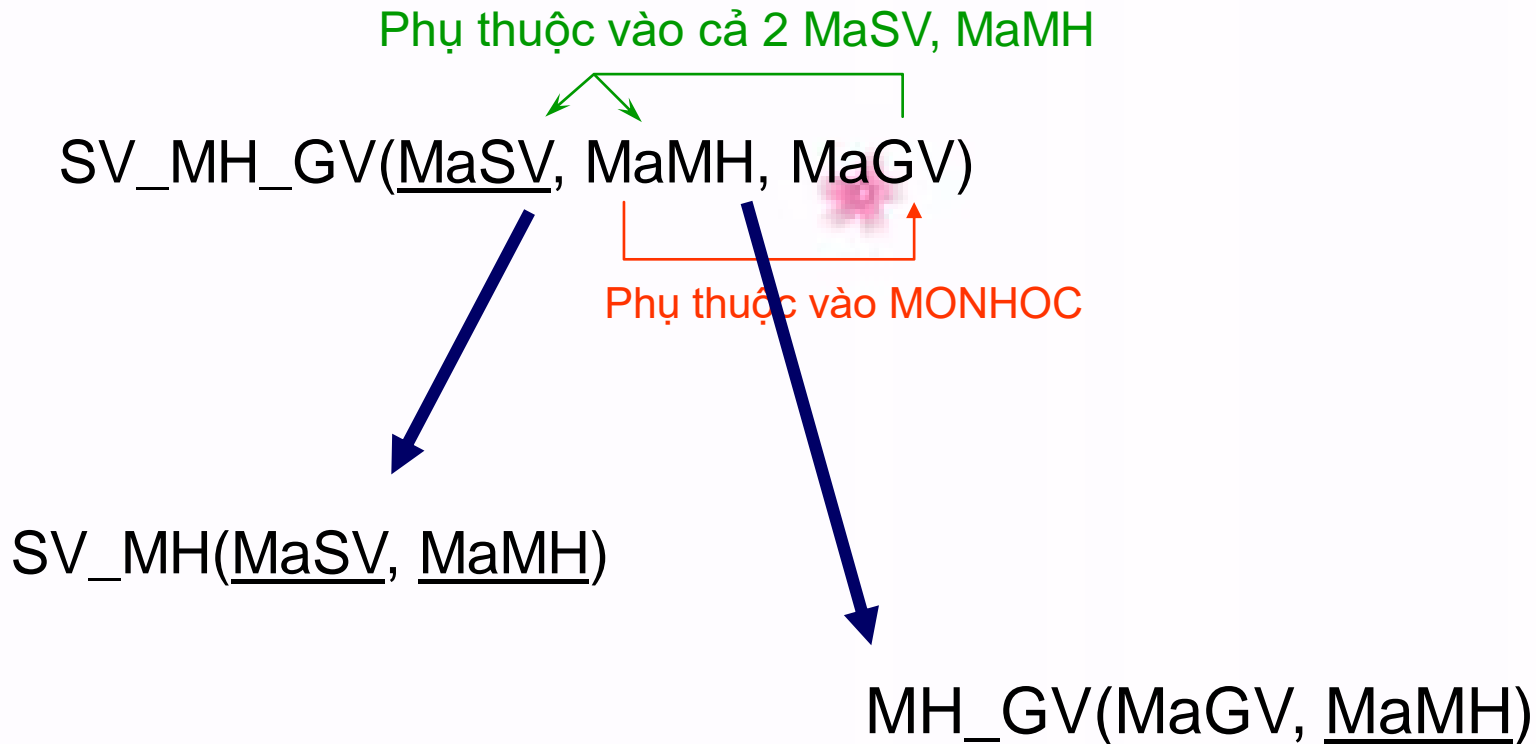
# Dạng chuẩn Boyce-Codd(tt)

## ■ Ví dụ



# Dạng chuẩn Boyce-Codd(tt)

## ■ Ví dụ





## 5.1.3. Phân rã lược đồ quan hệ

### □ Lược đồ quan hệ chung $R(A_1, \dots, A_n)$

- Tập hợp tất cả các thuộc tính của các thực thể.

### □ Xác định tập phụ thuộc hàm $F$ trên $R$ .

### □ Phân rã

- Sử dụng các thuật toán chuẩn hóa để tách  $R$  thành tập các lược đồ  $D = \{R_1, \dots, R_m\}$ .

### □ Yêu cầu

- Bảo toàn thông tin
- Các lược đồ  $R_i$  phải ở dạng chuẩn 3 hoặc BCNF.

## 5.2. Phân rã lược đồ quan hệ

- ❑ Phân rã lược đồ quan hệ là việc tách lược đồ quan hệ kém chất lượng ban đầu (chưa đạt chuẩn) cùng với tập phụ thuộc hàm của nó thành những lược đồ quan hệ chất lượng hơn.
- ❑ Sau phân rã, CSDL không còn lược đồ quan hệ  $R$  mà chỉ lưu lại các lược đồ quan hệ chiếu của nó  $R_1, R_2, \dots, R_n$ .
- ❑ **Hai vấn đề cần quan tâm:**
  - Phân rã bảo toàn thông tin (khôi phục được thông tin ban đầu từ các lược đồ đã tách?)
  - Phân rã bảo toàn Phụ thuộc hàm (Bảm đảm khôi phục được các PTH gốc)

## 5.2. Phân rã lược đồ quan hệ (tt)

### □ Ví dụ: Cho lược đồ quan hệ

- GIANGDAY(Monhoc, Sotiet, Lop, Giangvien, Hocvi, Diachi)
- Tập phụ thuộc hàm: { Monhoc  $\rightarrow$  Sotiet; Monhoc, Lop  $\rightarrow$  GV; GV  $\rightarrow$  Hocvi, Diachi }
- Xét bảng dữ liệu

Monhoc	Sotiet	Lop	Giangvien	Hocvi	Diachi
CSDL	60	CNTT1	Đ.T. Hiền	TS	HN
CSDL	60	CNTT2	Đ.T. Hiền	TS	HN
TRR	45	CNTT1	H.V.Anh	ThS	HCM
TRR	45	CNTT2	H.V.Anh	ThS	HCM

- Giả sử phân rã thành:
  - TKB (Monhoc, Sotiet, Lop)
  - GV(Lop, GV, Hocvi, Diachi)

## 5.2. Phân rã lược đồ quan hệ (tt)

### □ Ví dụ (tt)

Monhoc	Sotiet	Lop
CSDL	60	CNTT1
CSDL	60	CNTT2
TRR	45	CNTT1
TRR	45	CNTT2

Lop	GV	Hovi	ĐC
CNTT1	Đ.T.Hiền	TS	HN
CNTT2	Đ.T.Hiền	TS	HN
CNTT1	H.V.Anh	ThS	HCM
CNTT2	H.V.Anh	ThS	HCM

– Để trả lời câu truy vấn “Cho biết Giảng viên dạy CSDL ở lớp CNTT1”

→ Dùng phép kết tự nhiên theo cột lớp!

→ Xuất hiện 2 người dạy CSDL

## 5.2.1. Phân rã bảo toàn thông tin

□ Phân rã lược đồ  $R = (U, F)$  thành 1 tập hợp các lược đồ:  $R_1 = (U_1, F_1)$   $R_2 = (U_2, F_2)$ ....  $R_n = (U_n, F_n)$

□ Phân rã không mất mát thông tin nếu với mỗi thể hiện  $r$  hợp lệ của  $R$  thì:

$$r = r_1 \bowtie r_2 \bowtie \dots \bowtie r_n$$

$$\text{Với } r_1 = \pi_{U_1}(r) \quad r_2 = \pi_{U_2}(r), \dots \quad r_n = \pi_{U_n}(r)$$

## 5.2.1. Phân rã bảo toàn thông tin (tt)

- ❑ Thực tế sẽ nhận được nhiều bộ (tuple) từ phép kết các  $r_1, r_2, \dots, r_n$  hơn là các bộ gốc ban đầu → **Vậy tại sao lại gọi là mất mát (lossy) ??**
- ❑ Tuy nhiều bộ hơn nhưng lại thiếu thông tin và không có cách nào biết được bộ nào là đúng, bộ nào là không đúng với bộ gốc.
- ❑ Nhiều bộ hơn nhưng không đúng  $\equiv$  mất mát thông tin

# 5.2.1. Phân rã bảo toàn thông tin (tt)

## □ Phân rã nhị phân - Định lý 5.1

- Phân rã  $D = \{R_1(U_1), R_2(U_2)\}$  của  $R(U)$  không mất mát thông tin đối với tập phụ thuộc hàm  $F$  nếu và chỉ nếu nó thỏa mãn 1 trong 2 phụ thuộc hàm:
  - $(U_1 \cap U_2) \rightarrow (U_1 \setminus U_2) \in F^+$ , hoặc
  - $(U_1 \cap U_2) \rightarrow (U_2 \setminus U_1) \in F^+$

## □ Ví dụ:

- Cho quan hệ  $R(M, S, L, G, H, D)$
- $F = \{M \rightarrow S; ML \rightarrow G; G \rightarrow HD\}$
- Phân rã thành:
  - $R_1(M, S, L, G)$
  - $R_2(G, H, D)$

## 5.2.1. Phân rã bảo toàn thông tin (tt)

### □ Ví dụ:

- Cho quan hệ  $R(M, S, L, G, H, D)$
- $F = \{M \rightarrow S; ML \rightarrow G; G \rightarrow HD\}$
- Phân rã thành:
  - $R_1(M, S, L, G)$
  - $R_2(G, H, D)$

### □ Rõ ràng:

$$U1 \cap U2 = G$$

$$U1 \setminus U2 = \{M, S, L\}$$

$$U2 \setminus U1 = \{H, D\}$$

$$U1 \cap U2 \rightarrow U2 \setminus U1 \\ \text{thuộc } F^+$$

Phân rã này bảo  
toàn thông tin theo  
định lý 5.1



## 5.2.1. Phân rã bảo toàn thông tin (tt)

### □ Kiểm tra bảo toàn thông tin bằng bảng Tableau

- Cho quan hệ  $R(U)$  với tập phụ thuộc hàm  $F$
- Phân có phân rã  $D = \{R_1, R_2, \dots, R_m\}$
- Bảng Tableau có dạng:
  - Gồm  $n$  cột, mỗi cột ứng với 1 thuộc tính trong  $U$
  - Gồm  $m$  hàng, mỗi hàng ứng với các quan hệ  $R_i$  đã phân rã

	$u_1$	$u_2$	$u_3$	...	$u_n$
$R_1$					
$R_2$					
$R_3$					
....					
$R_M$					

## 5.2.1. Phân rã bảo toàn thông tin (tt)

- Tại ô  $(i, j)$ 
  - Điền giá trị  $a_j$  nếu  $R_i$  có chứa thuộc tính thứ  $j$  của  $R$
  - Điền giá trị  $b_k$  nếu  $R_i$  không chứa thuộc tính thứ  $j$  của  $R$   
(chú ý:  $k$  tăng dần)
- Biến đổi bảng tableau  $T$  ban đầu thành bảng  $T^*$  theo quy tắc  
**While**  $(X \rightarrow Y \in F)$   
{  
    *Chọn dòng  $t_1$  và  $t_2$  sao cho  $t_1.X = t_2.X$*   
    **If**  $(t_1.Y \neq t_2.Y)$   
    {  
        Nếu  $t_1.Y = a_j$  và  $t_2.Y = b_k$  thay  $t_2.Y = a_j$   
        Nếu  $t_1.Y = b_k$  và  $t_2.Y = a_j$  thay  $t_1.Y = a_j$   
    }  
}

# 5.2.1. Phân rã bảo toàn thông tin (tt)

## ❑ Ví dụ:

- Cho quan hệ  $R(\text{Môn}, \text{Giảng\_Viên}, \text{Giờ}, \text{Phòng}, \text{Sinh\_Viên}, \text{Hạng})$  với tập phụ thuộc hàm  $F = \{M \rightarrow GV; G, P \rightarrow M; G, GV \rightarrow P; M, SV \rightarrow H; G, SV \rightarrow P\}$
- Giả sử có phân rã quan hệ  $R$  trên thành:
  - $R1(\underline{G}, \underline{P}, M)$  với tập PTH là  $F1 = \{G, P \rightarrow M\}$
  - $R2(\underline{M}, GV)$  với tập PTH là  $F2 = \{M \rightarrow GV\}$
  - $R3(\underline{M}, \underline{SV}, H)$  với tập PTH là  $F3 = \{M, SV \rightarrow H\}$
- Lập bảng Tableau (T) và biến đổi bảng theo thuật toán.

	M	GV	G	P	SV	H
R1	a1	b1	a3	a4	b2	b3
R2	a1	a2	b4	b5	b6	b7
R3	a1	b8	b9	b10	a5	a6

	M	GV	G	P	SV	H
R1	a1	a2	a3	a4	b2	b3
R2	a1	a2	b4	b5	b6	b7
R3	a1	b8	b9	b10	a5	a6

	M	GV	G	P	SV	H
R1	a1	a2	a3	a4	b2	b3
R2	a1	a2	b4	b5	b6	b7
R3	a1	a2	b9	b10	a5	a6

*Phân rã không  
bảo toàn thông tin*

## 5.2.1. Phân rã bảo toàn thông tin (tt)

### □ Định lý 5.2

- Nếu phân rã  $D = \{R_1, \dots, R_m\}$  của  $R$  không mất mát thông tin đối với  $F$  và
- Phân rã  $D_i = \{Q_1, \dots, Q_k\}$  của  $R_i$  không mất mát thông tin đối với  $\pi_{R_i}(F)$  thì
- Phân rã  $D' = \{R_1, \dots, R_{i-1}, Q_1, \dots, Q_k, R_{i+1}, \dots, R_m\}$  của  $R$  cũng không mất mát thông tin.

## 5.2.2. Phân rã bảo toàn phụ thuộc hàm

### □ Phép chiếu của tập phụ thuộc hàm

Xét lược đồ quan hệ  $R = (U, F)$  và tập  $S \subseteq U$

Phép chiếu của  $F$  lên tập các thuộc tính  $S$  được định nghĩa như sau:

$$\pi_S(F) = \{X \rightarrow Y \mid X \rightarrow Y \in F^+ \text{ và } X \cup Y \subseteq S\}$$

## 5.2.2. Phân rã bảo toàn phụ PTH (tt)

□ Cho lược đồ  $R(U, F)$

$D = \{R_1(U_1, F_1), R_2(U_2, F_2), \dots, R(U_n, F_n)\}$  là phân rã của  $R$ .

□ Phân rã  $D$  được gọi là bảo toàn phụ thuộc hàm nếu và chỉ nếu  $F$  tương đương với  $F' = \cup F_i$

□ Nếu 1 phụ thuộc hàm  $f \in F$  nhưng không thuộc bất kỳ  $F_i$  nào không có nghĩa là phân rã đó không bảo toàn phụ thuộc hàm (vì  $f$  có thể được suy diễn từ  $\cup F_i$ )

– Chỉ khi nào  $f$  không thể suy diễn từ  $\cup F_i$  thì phân rã đó mới không bảo toàn PTH

## 5.2.2. Phân rã bảo toàn phụ PTH (tt)

### □ Ví dụ:

- $R(ABCD)$ ,  $F = \{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow A\}$
- Phân rã thành:
  - $R1(AB)$
  - $R2(BC)$
  - $R3(CD)$
- Hỏi phân rã trên có mất mát thông tin không?
- Phân rã trên có bảo toàn phụ thuộc hàm không?

## 5.2.2. Phân rã bảo toàn phụ PTH (tt)

### □ Ví dụ (tiếp)

Ta có:

- $F1 = \pi_{AB}(F) = \{A \rightarrow B, B \rightarrow A\}$
- $F2 = \pi_{BC}(F) = \{B \rightarrow C, C \rightarrow B\}$
- $F3 = \pi_{CD}(F) = \{C \rightarrow D, D \rightarrow C\}$

Vậy:

$$F' = F1 \cup F2 \cup F3 = \{A \rightarrow B, B \rightarrow A, B \rightarrow C, C \rightarrow B, C \rightarrow D, D \rightarrow C\}$$

Kiểm tra ta thấy  $F$  phủ  $F'$  và  $F'$  phủ  $F$

→  $F \equiv F'$  → Phân rã đã cho bảo toàn phụ thuộc hàm



## 5.2.2. Phân rã bảo toàn phụ PTH (tt)

### □ Ví dụ 2:

- Cho quan hệ  $R(U, F)$  với  $U = \{ABCDEFGH\}$ ,  
 $F = \{ABH \rightarrow C, A \rightarrow DE, BGH \rightarrow F, F \rightarrow ADH, BH \rightarrow GE\}$
- Phân rã  $R$  thành
  - $R_1(ADE, \{A \rightarrow DE\})$
  - $R_2(ABCFGH, \{ABH \rightarrow C, BGH \rightarrow F, F \rightarrow AH, BH \rightarrow G\})$

Phân rã trên có bảo toàn phụ thuộc hàm không?

## 5.2.2. Phân rã bảo toàn phụ PTH (tt)

Ví dụ 3:

❑ Xét lược đồ quan hệ:

HAS\_ACCOUNT(ClientID, OfficeID, AccountNumber)

❑ Với các Phụ thuộc hàm:

ClientID, OfficeID → AccountNumber  
AccountNumber → OfficeID

❑ Nếu phân rã lược đồ trên thành 2 lược đồ sau:

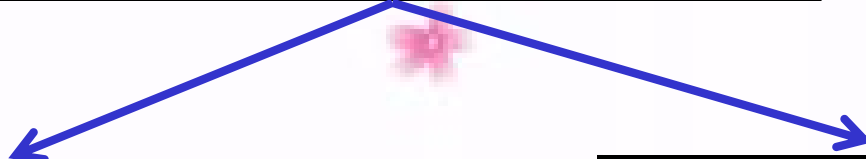
ACCT\_OFFICE (AccountNumber, OfficeID)

ACCT\_CLIENT (AccountNumber, ClientID)

❑ Phân rã trên có bảo toàn PTH?

## 5.2.2. Phân rã bảo toàn phụ PTH (tt)

AccountNumber	ClientID	OfficeID
B123	111111	SB01
A908	123456	MN08



AccountNumber	OfficeID
B123	SB01
A908	MN08

Account Number	ClientID
B123	111111
A908	123456

## 5.2.2. Phân rã bảo toàn phụ PTH (tt)

- ❑ Chèn thêm 1 hàng vào các phân rã của lược đồ HAS\_ACCOUNT

AccountNumber	ClientID	OfficeID
B123	111111	SB01
B567	111111	SB01
A908	123456	MN08

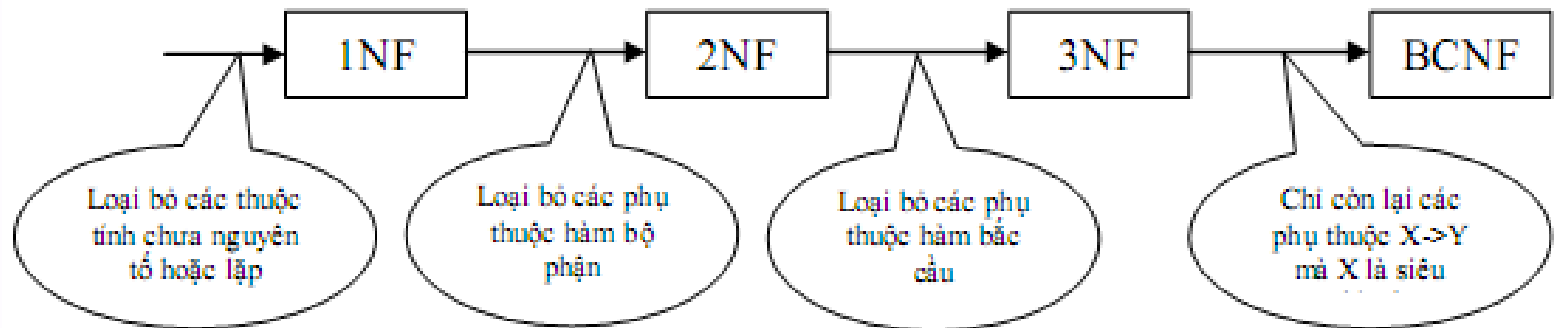
AccountNumber	OfficeID
B123	SB01
B567	SB01
A908	MN08

Account Number	ClientID
B123	111111
B567	111111
A908	123456

Sau khi join 2 lược đồ phân rã lại, phụ thuộc hàm  
ClientID, OfficeID → AccountNumber  
bị vi phạm

## 5.3. Chuẩn hóa CSDL

- ❑ Quá trình chuẩn hóa được thực hiện qua nhiều bước. Mỗi bước tương ứng một dạng chuẩn
- ❑ Bước chuẩn hóa
  - Bước 1: Đưa về dạng 1NF, loại bỏ các thuộc tính đa trị
  - Bước 2: Đưa về dạng 2NF, loại bỏ phụ thuộc hàm bộ phận vào khóa
  - Bước 3: Đưa về dạng 3NF, loại bỏ phụ thuộc bắc cầu vào khóa
  - Bước 4: Đưa về dạng BCNF: Mọi phụ thuộc hàm phải có vế trái là siêu khóa.



# Giải thuật phân rã thành 3NF

## Cho lược đồ $R(U,F)$

- **Bước 1:** Tìm phủ tối thiểu  $G$  của  $F$
- **Bước 2:** Phân hoạch  $G$  thành các tập phụ thuộc hàm  $G_1, \dots, G_n$  sao cho mỗi  $G_i$  chứa các PTH có cùng vế trái
- **Bước 3:** với mỗi  $G_i$ , tạo 1 lược đồ  $(R_i, G_i)$  với  $R_i$  chứa tất cả thuộc tính trong  $G_i$
- **Bước 4:** Nếu một trong các  $R_i$  thỏa  $(R_i)^+_{F_i} = K$  là khóa tối thiểu của  $R$  thì kết thúc, ngược lại đặt  $R_0 = (K, \{\})$  là 1 lược đồ mới. Khi đó  $R_0, R_1, \dots, R_n$  là kết quả phân rã.

## □ Tính chất của giải thuật phân rã thành 3NF

- Bảo toàn thông tin
- Bảo toàn phụ thuộc hàm

# Ví dụ – phân rã thành 3NF

Cho quan hệ R(ABCDEFGH) với tập PTH gồm:

$$F = \{ABH \rightarrow C, A \rightarrow DE, BGH \rightarrow F, F \rightarrow ADH, BH \rightarrow GE\}$$

Hãy phân rã lược đồ trên thành các lược đồ đạt 3NF

Giải:

Bước 1: Tìm phủ tối thiểu của F là  $G = \{BH \rightarrow C, A \rightarrow D, C \rightarrow E, F \rightarrow A, E \rightarrow F\}$

Bước 2: phân hoạch G thành 5 nhóm PTH cùng về trái

Bước 3: Tạo lược đồ 5 được đồ

- R1 (BHC;  $\{BH \rightarrow C\}$ )
- R2 (AD;  $\{A \rightarrow D\}$ )
- R3 (CE;  $\{C \rightarrow E\}$ )
- R4 (FA;  $\{F \rightarrow A\}$ )
- R5 (EF;  $\{E \rightarrow F\}$ )

Bước 4: Không có lược đồ phân rã nào thỏa  $(R_i)^+_F = BGH$  (khóa tối thiểu của R)  $\rightarrow$  lập quan hệ mới là R6 (BGH,  $\{ \}$ )

Vậy: Kết quả phân rã đạt 3NF là: R1(BHC), R2(AD), R3(CE), R4(FA), R5(EF), R6(BGH)

# Giải thuật phân rã thành BCNF

□  $R=(U,F)$  là 1 lược đồ quan hệ không ở chuẩn BCNF.

□ Giải thuật:

Thực hiện lặp lại việc phân chia  $R$  thành những lược đồ nhỏ hơn sao cho các lược đồ mới có ít PTH vi phạm BCNF hơn. Giải thuật kết thúc khi tất cả lược đồ kết quả đều ở dạng BCNF



# Giải thuật phân rã thành BCNF

Input  $R = (U, F)$

Decom =  $R$

While có lược đồ  $S = (V, F')$  trong Decom không phải BCNF

*/\*Nếu có  $X \rightarrow Y \in F$  sao cho  $X \cup Y \subseteq S$  và vi phạm BCNF,  
dùng FD này để phân rã\*/*

- Thay  $S$  trong Decom với  $S_1 = (XY, F_1)$
- $S_2 = ((S - Y) \cup X, F_2)$  với  $F_1, F_2$  là tất cả các FD của  $F'$

End

Return Decom

# Ví dụ

□ Cho  $R = (U, F)$

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

$F = \{ABH \rightarrow C, A \rightarrow DE, BGH \rightarrow F, F \rightarrow ADH, BH \rightarrow GE\}$

□ Tìm FD vi phạm BCNF

- $(ABH)^+ = U$ ,  $ABH$  là siêu khóa,  $ABH \rightarrow C$  không vi phạm BCNF
- $A^+ \neq U$ ,  $A \rightarrow DE$  vi phạm BCNF

□ Chia  $R$  thành

- $R_1 = (ADE, \{A \rightarrow DE\})$
- $R_2 = (ABCFGH, \{ABH \rightarrow C, BGH \rightarrow F, F \rightarrow AH, BH \rightarrow G\})$

# Ví dụ (tt)

❑ R1 là BCNF

❑ Với R2 (ABCFGH, { $ABH \rightarrow C$ ,  $BGH \rightarrow F$ ,  $F \rightarrow AH$ ,  $BH \rightarrow G$ })

- $ABH \rightarrow C$ ,  $BGH \rightarrow F$  không vi phạm BCNF ( $ABH$ ,  $BGH$  đều là siêu khóa)
- $F \rightarrow AH$  vi phạm BCNF

Vậy Phân rã R2 thành

- $R_{21} = (FAH, \{F \rightarrow AH\})$
- $R_{22} = (FBCG, \{ \})$

$R_{21}$ ,  $R_{22}$  đều là BCNF nhưng khi đó các Phụ thuộc hàm  $ABH \rightarrow C$ ,  $BGH \rightarrow F$  và  $BH \rightarrow G$  không có mặt nữa và cũng không thể suy dẫn được từ các PTH của  $R_{21}$ ,  $R_{22}$  và  $R_1$

**Phân rã R2 không bảo toàn phụ thuộc hàm**

# Phân rã BCNF thông qua phân rã 3NF

- ❑ Do giải thuật phân rã BCNF có thể không bảo toàn phụ thuộc hàm → Nên áp dụng phân rã BCNF thông qua phân rã 3NF.
- ❑ Nếu lược đồ sau phân rã là BCNF thì dừng, nếu không thì dùng lúc đó mới dùng giải thuật BCNF để phân rã tiếp

# Ví dụ

❑ Xét tập thuộc tính sau: St (Student), C (course), Sem (semester), P (Professor), T (time) và R(room) và tập PTH như sau:

St C Sem  $\rightarrow$  P

P Sem  $\rightarrow$  C

C Sem T  $\rightarrow$  P

P Sem T  $\rightarrow$  C R

P Sem C T  $\rightarrow$  R

P Sem T  $\rightarrow$  C

# Phân rã 3NF bảo toàn FD

□ Phân rã thành 4 lược đồ như sau:

R1 (St C Sem P; {St C Sem  $\rightarrow$  P})

R2 (P Sem C; {P Sem  $\rightarrow$  C})

R3 (C Sem T P; {C Sem T  $\rightarrow$  P})

R4 (P Sem T R; {P Sem T  $\rightarrow$  R})

□ Vì không có phân rã nào hình thành siêu khóa cho lược đồ gốc, nên bổ sung thêm lược đồ mới (bước 4)

R5 ( St T Sem P; {})

# Phân rã thành BCNF

❑ Các phân rã 1 và 3 không phải là BCNF vì  $P \text{ Sem} \rightarrow C$  nằm trong phân rã 2

❑ Phân rã 1 được tách thành 2 lược đồ mới

- $(P \text{ Sem } C; \{P \text{ Sem} \rightarrow C\})$
- $(\text{St Sem } P; \{\})$

➔ Phân rã tuy không mất mát thông tin nhưng không bảo toàn PTH  $\text{St } C \text{ Sem} \rightarrow P$

# Phân rã thành BCNF

## □ Phân rã lược đồ 3 thành

- $(P \text{ Sem } C; \{P \text{ Sem } \rightarrow C\})$
- $(P \text{ Sem } T; \{\})$

➔ Không mất mát thông tin nhưng cũng không bảo toàn PTH  $C \text{ Sem } T \rightarrow P$



# Phân rã thành BCNF

□ Kết quả cuối cùng:

$(P \text{ Sem } C; \{P \text{ Sem} \rightarrow C\})$

$(P \text{ Sem } St)$

$(P \text{ Sem } T)$

$(P \text{ Sem } T \text{ R}; \{P \text{ Sem } T \rightarrow R\})$

$(St \text{ T Sem } P)$

# Bài tập



# Nhận xét

❑ Việc phân rã R thành R1, R21, R22 không phải là duy nhất.

❑ Nếu bắt đầu từ FD  $F \rightarrow ADH$  thì sẽ có

$$R1 = (FADH; \{F \rightarrow ADH\})$$

$$R2 = (FBCEG, \{\})$$

R1, R2 cũng ở chuẩn BCNF và 1 số FD gốc cũng bị mất, không thể suy diễn được

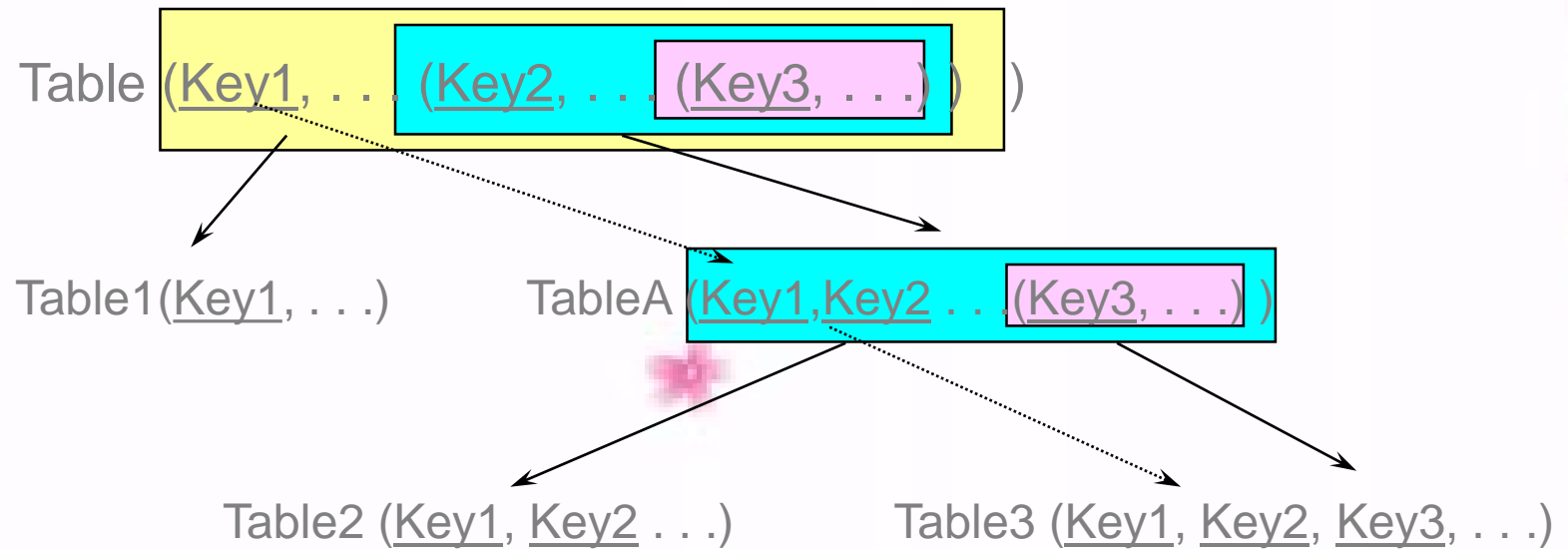
# Tính chất của giải thuật phân rã BCNF

- ❑ Không mất mát thông tin
- ❑ Nhưng có thể không bảo toàn phụ thuộc hàm
- ❑ Là giải thuật không xác định (**nondeterministic**), phụ thuộc vào thứ tự các PTH được chọn để xét phân rã

# Ví dụ

- ❑ Cho  $R = (U, F)$  với  $U = \{ABCDEFGH\}$ ,  $F = \{ABH \rightarrow C, A \rightarrow DE, BGH \rightarrow F, F \rightarrow ADH, BH \rightarrow GE\}$
- ❑ Phủ tối thiểu của  $F$  là:  $G = \{BH \rightarrow C, A \rightarrow D, C \rightarrow E, F \rightarrow A, E \rightarrow F\}$
- ❑ Phân rã thành 5 lược đồ:
  - $R_1 (BHC; \{BH \rightarrow C\})$
  - $R_2 (AD; \{A \rightarrow D\})$
  - $R_3 (CE; \{C \rightarrow E\})$
  - $R_4 (FA; \{F \rightarrow A\})$
  - $R_5 (EF; \{E \rightarrow F\})$
- ❑ Không có lược đồ phân rã nào có  $(R_i)^+_F =$  siêu khóa BCGH của  $R$ , nên bổ sung thêm lược đồ thứ 6
  - $R_6 (BCGH; \{\})$

# Dạng chuẩn 1 (tt)



## ❑ Lược đồ gốc:

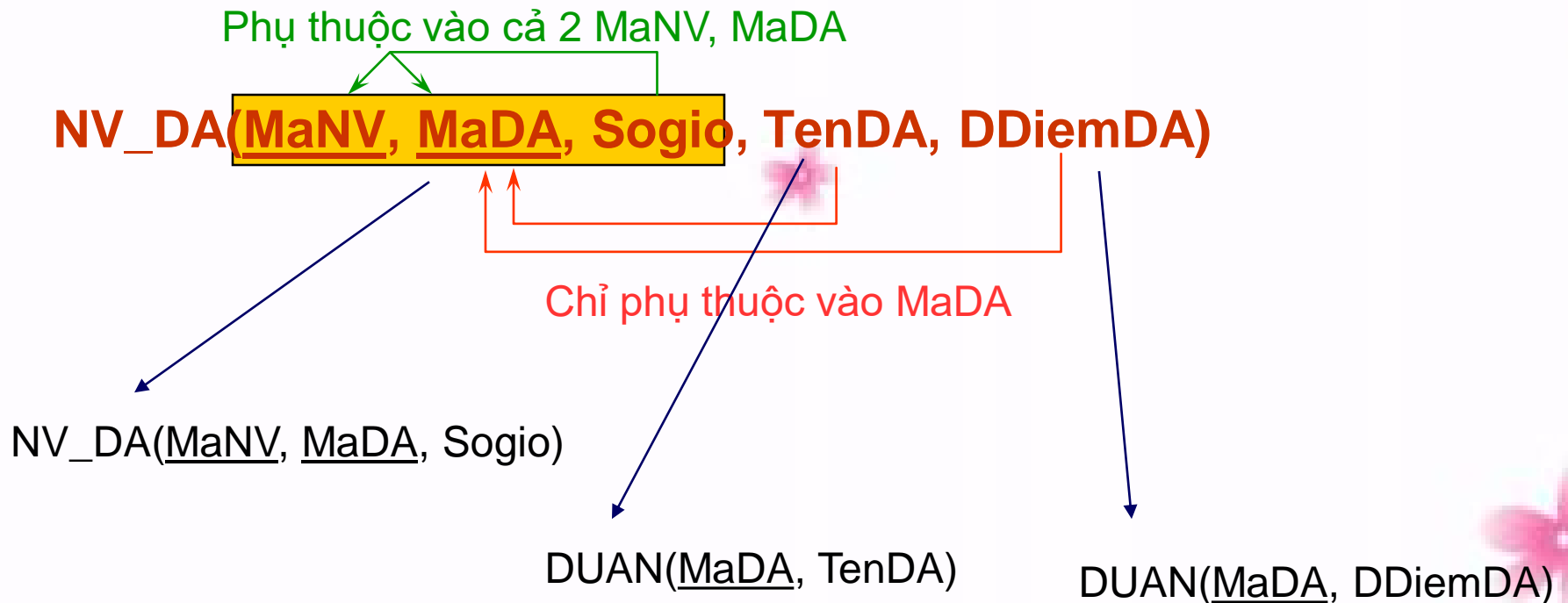
Table (Key1, aaa. . . (Key2, bbb. . . (Key3, ccc. . . ) ) )

## ❑ Để thỏa mãn 1NF chúng ta thực hiện

- Table1(Key1, aaa . . .)
- Table2(Key1, Key2, bbb . . .)
- Table3(Key1, Key2, Key3, ccc. . .)

# Dạng chuẩn 2 (tt)

## □ Ví dụ



# Dạng chuẩn 3 (tt)

Phụ thuộc vào MaNV

NV\_DV(MaNV, TenNV, NS, DCHI, MaDV, TenDV, TruongPHG)

Phụ thuộc vào MaDV

NHANVIEN(MaNV, TenNV, NS, DCHI, MaDV)

DONVI(MaDV, TenDV, TruongPHG)



# Ví dụ phân rã bảo toàn PTH (1)

## □ Cho

❖  $R(ABCDEFGG)$

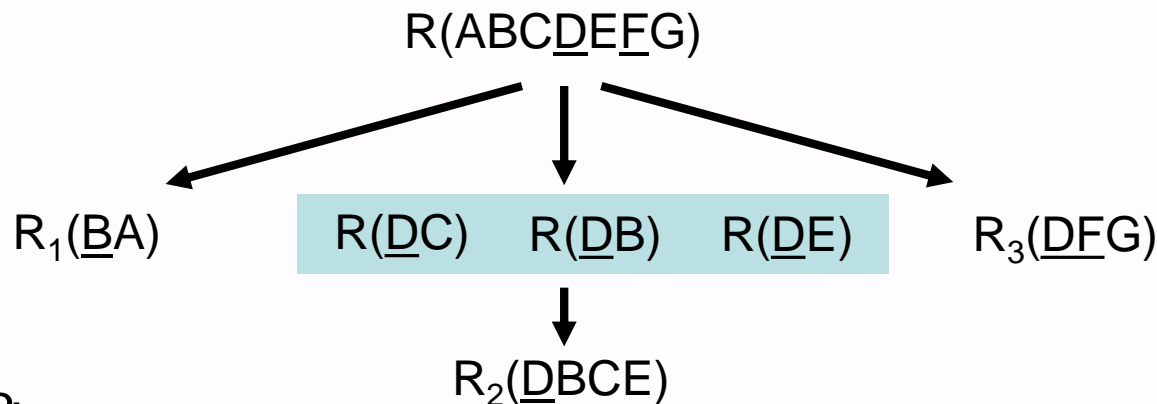
❖  $F = \{B \rightarrow A, D \rightarrow C, D \rightarrow EB, DF \rightarrow G\}$

## □ Tách về dạng chuẩn 3, bảo toàn PTH

– B1:

- Phủ tối thiểu  $G = \{B \rightarrow A, D \rightarrow C, D \rightarrow B, D \rightarrow E, DF \rightarrow G\}$ .

– B2:



– B3:

- Xuất  $D = \{R_1, R_2, R_3\}$ .

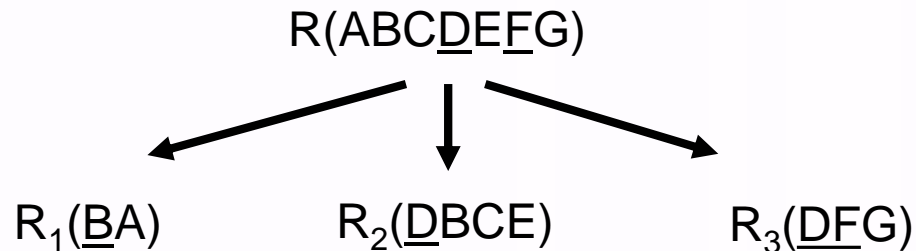
# Ví dụ phân rã bảo toàn PTH (2)

## □ Cho

- $R(ABCDEFGHI)$
- $F = \{B \rightarrow A, D \rightarrow C, D \rightarrow EB, DF \rightarrow G\}$

## □ Tách về dạng chuẩn 3, bảo toàn PTH

- $B1$ :
  - Phủ tối thiểu  $G = \{B \rightarrow A, D \rightarrow C, D \rightarrow B, D \rightarrow E, DF \rightarrow G\}$ .
- $B2$ :



- $B3$ :
  - Vì  $U_1 \cup U_2 \cup U_3 = \{ABCDEF\}$  nên đặt  $R_4(HI)$ .
- $B4$ :
  - $D = \{R_1, R_2, R_3, R_4\}$ .

## 5.1.3.2. Phân rã bảo toàn thông tin (tt)

### □ Thuật toán

❖ Nhập:  $R(U)$ ,  $U = \{A_1, \dots, A_n\}$  và tập PTH  $F$ .

❖ Xuất:  $D = \{R_1, \dots, R_m\}$ ,  $R_i$  ở dạng chuẩn Boyce-Codd.

–  $B1$ :

- $D = \{R\}$ ;

–  $B2$ :

- Nếu có lược đồ  $Q(U_Q) \in D$  không ở dạng chuẩn BC thì

- Tìm  $X \rightarrow Y \in \pi_Q(F)$  làm  $Q$  vi phạm điều kiện BC.

- $D = (D - \{Q\}) \cup Q_1(U_{Q1}) \cup Q_2(U_{Q2})$   
với  $U_{Q1} = U_Q - Y$  và  $U_{Q2} = X \cup Y$ .

- Quay lại B2.

- Ngược lại, chuyển sang B3.

–  $B3$ :

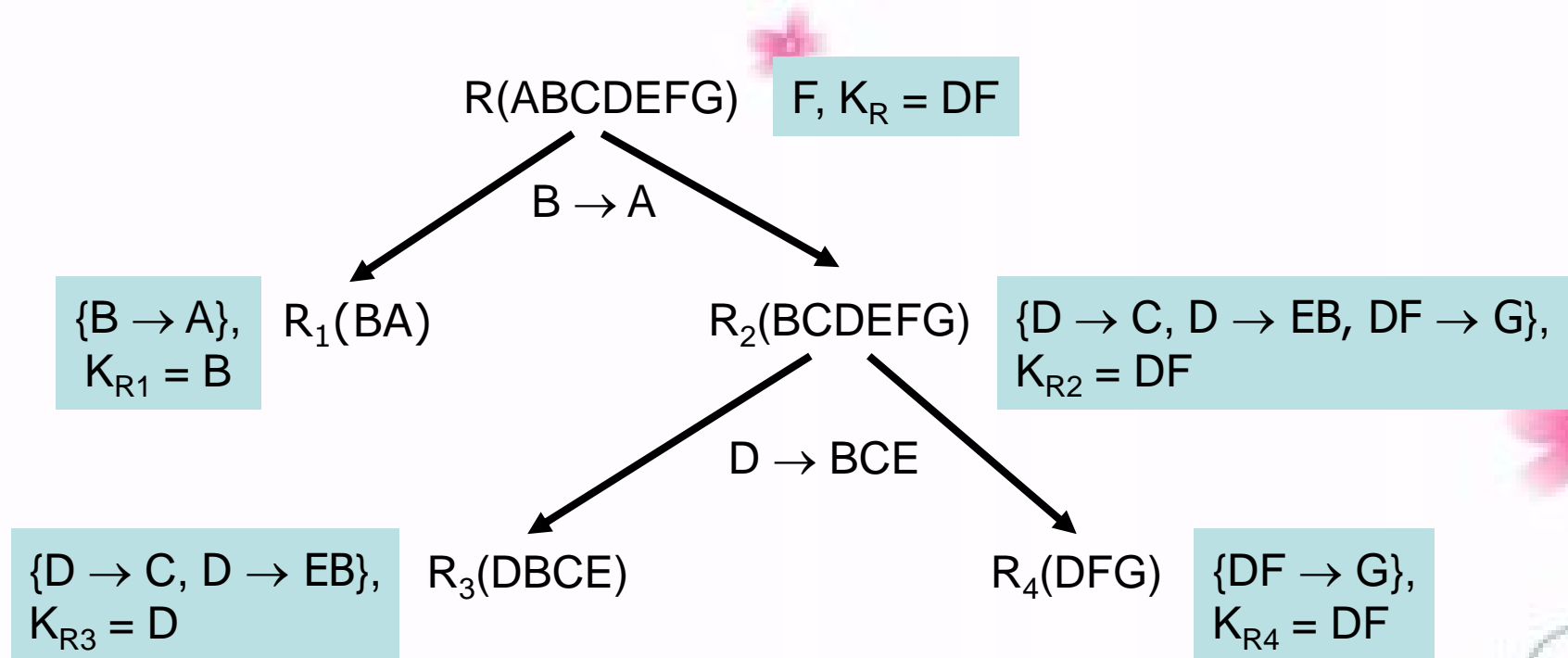
- Xuất  $D$ .

# Ví dụ phân rã bảo toàn thông tin (1)

## □ Cho:

- $R(ABCDEFGG)$
- $F = \{B \rightarrow A, D \rightarrow C, D \rightarrow EB, DF \rightarrow G\}$

## □ Tách về dạng chuẩn BC, không mất thông tin.



# Ví dụ phân rã bảo toàn thông tin (2)

