

Chương 7

Phụ thuộc hàm và Chuẩn hóa cơ sở dữ liệu

Nội dung trình bày

- Nguyên tắc thiết kế các lược đồ quan hệ.
 - Phụ thuộc hàm.
 - Các dạng chuẩn.
 - Một số thuật toán chuẩn hóa.
-

Nguyên tắc thiết kế

- Nhìn lại vấn đề thiết kế csdl
 - Dựa trên trực quan của người thiết kế.
 - Thiếu một tiêu chuẩn hình thức để đánh giá.
 - Đánh giá chất lượng thiết kế
 - Ngữ nghĩa của các thuộc tính.
 - Giảm các giá trị thừa trong các bộ.
 - Giảm các giá trị null trong các bộ.
 - Không để xuất hiện các bộ không có thực.
-

Ngữ nghĩa của các thuộc tính (1)

NHANVIEN

f.k.

Ten	<u>MaNV</u>	NgSinh	DChi	MaPhong
-----	-------------	--------	------	---------

p.k.

PHONGBAN

f.k.

Ten	<u>MaPB</u>	TrPhong
-----	-------------	---------

p.k.

DUAN

f.k.

Ten	<u>MaDA</u>	Diadiem	PhongQly
-----	-------------	---------	----------

p.k.

TRUSO_PHONG

f.k.

f.k.

<u>MaPB</u>	<u>Truso</u>
-------------	--------------

p.k.

THAMGIA

f.k.

f.k.

<u>MaNV</u>	<u>MaDA</u>	SoGio
-------------	-------------	-------

p.k.

Ngữ nghĩa của các thuộc tính (2)



- Ý nghĩa của các thuộc tính càng dễ hiểu thì lược đồ thiết kế càng tốt.
 - Tránh tổ hợp các thuộc tính của nhiều kiểu thực thể vào cùng một lược đồ.
-

Thông tin thừa trong các bộ (1)

NHANVIEN

Ten	<u>MaNV</u>	NgSinh	DChi	MaPhong
Hung	123456789	09/01/1965	...	5
Nghia	333445555	08/12/1955	...	5
Vuong	999887777	19/01/1968	...	4


PHONGBAN

Ten	<u>MaPB</u>	TrPhong
Nghien cuu	5	333445555

NHANVIEN_PHONGBAN

TenNV	<u>MaNV</u>	NgSinh	DChi	MaPB	TenPB	TrPhong
Hung	123456789	09/10/1965	...	5	Nghien cuu	333445555
Nghia	333445555	08/12/1965	...	5	Nghien cuu	333445555

Dữ liệu bị trùng lặp



Thông tin thừa trong các bộ (2)

- Dị thường khi thêm bộ

NHANVIEN_PHONGBAN

TenNV	<u>MaNV</u>	NgSinh	DChi	MaPB	TenPB	TrPhong
Nghia	333445555	08/12/1965	...	5	Nghien cuu	333445555
Hung	123456789	09/10/1965	...	5	Nghien cuu	999887777
null	null	null	null	4	Hanh chinh	987654321

- Dị thường khi xóa bộ

NHANVIEN_PHONGBAN

TenNV	<u>MaNV</u>	NgSinh	DChi	MaPB	TenPB	TrPhong
Nghia	333445555	08/12/1965	...	5	Nghien cuu	333445555
Hung	123456789	09/10/1965	...	5	Nghien cuu	333445555

Thông tin thừa trong các bộ (3)

- Dị thường khi sửa bộ

NHANVIEN_PHONGBAN

TenNV	<u>MaNV</u>	NgSinh	DChi	MaPB	TenPB	TrPhong
Nghia	333445555	08/12/1965	...	5	Nghien cuu	123456789
Hung	123456789	09/10/1965	...	5	Nghien cuu	123456789

- Tránh xảy ra các dị thường cập nhật dữ liệu.
 - Có thể vi phạm nguyên tắc này để tăng hiệu quả truy vấn dữ liệu. Khi đó các dị thường cần được ghi chú cẩn thận.
-

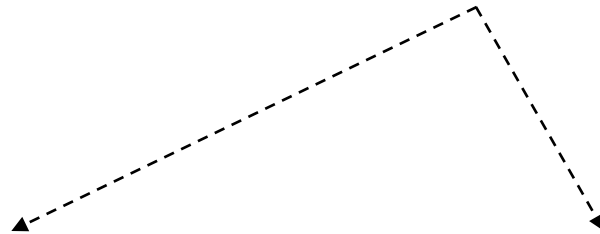
Giá trị null trong các bộ

- Nếu nhiều thuộc tính trong lược đồ nhận giá trị null sẽ
 - Lãng phí không gian lưu trữ.
 - Khó khăn trong thực hiện các phép toán kết.
 - Khó khăn khi sử dụng các hàm tập hợp.
 - Tránh lưu trữ các thuộc tính nhận nhiều giá trị null.
-

Phát sinh các bộ không có thực (1)

NHANVIEN_DUAN

<u>MaNV</u>	<u>MaDA</u>	Gio	TenNV	TenDA	Diadiem
123456789	1	32.5	Hung	San pham X	Tan Binh
123456789	2	7.5	Hung	San pham Y	Thu Duc
333445555	2	10	Nghia	San pham Y	Thu Duc



NHANVIEN_DIADIEM

<u>TenNV</u>	<u>Diadiem</u>
--------------	----------------

p.k.

NHANVIEN_DUAN1

<u>MaNV</u>	<u>MaDA</u>	SoGio	TenDA	Diadiem
-------------	-------------	-------	-------	---------

p.k.

Phát sinh các bộ không có thực (2)

NHANVIEN_DIADIEM

<u>TenNV</u>	<u>Diadiem</u>
Hung	Tan Binh
Hung	Thu Duc
Nghia	Thu Duc

NHANVIEN_DUAN1

<u>MaNV</u>	<u>MaDA</u>	SoGio	TenDA	Diadiem
123456789	1	32.5	San pham X	Tan Binh
123456789	2	7.5	San pham Y	Thu Duc
333445555	2	10	San pham Y	Thu Duc

Kết tự nhiên

MaNV	MaDA	Gio	TenDA	Diadiem	TenNV
123456789	1	32.5	San pham X	Tan Binh	Hung
123456789	2	7.5	San pham Y	Thu Duc	Hung
123456789	2	7.5	San pham Y	Thu Duc	Nghia
333445555	2	10	San pham Y	Thu Duc	Hung
333445555	2	10	San pham Y	Thu Duc	Nghia

Phát sinh các bộ không có thực (3)

- Xây dựng các lược đồ quan hệ sao cho việc thực hiện phép kết bằng giữa chúng chỉ áp dụng trên các thuộc tính khóa chính hoặc khóa ngoại.

Nội dung trình bày

- Nguyên tắc thiết kế các lược đồ quan hệ.
 - Phụ thuộc hàm.
 - Các dạng chuẩn.
 - Một số thuật toán chuẩn hóa.
-

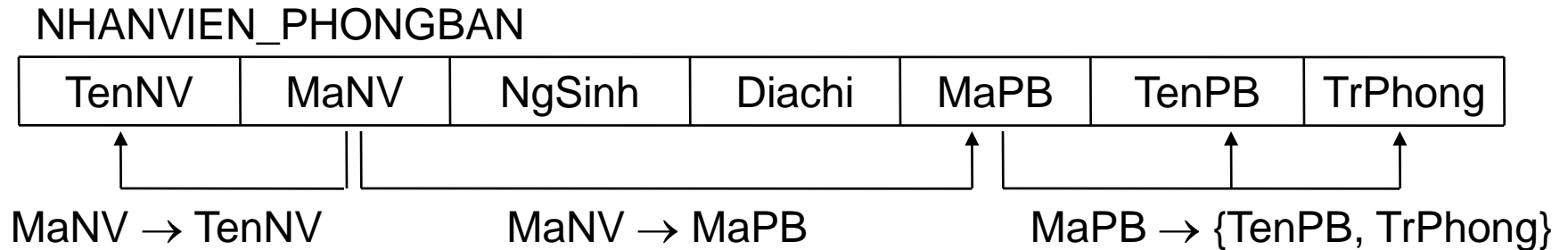
Phụ thuộc hàm (1)

- Xét lược đồ quan hệ gồm n thuộc tính
 - $R(U)$, $U = \{A_1, A_2, \dots, A_n\}$
- PTH giữa hai tập thuộc tính $X, Y \subseteq U$
 - Ký hiệu: $X \rightarrow Y$.
 - $\forall r \in R, \forall t_1, t_2 \in r$ nếu $t_1[X] = t_2[X]$ thì $t_1[Y] = t_2[Y]$.
 - X là vế trái và Y là vế phải của PTH.

$r(R)$	A	B
	1	4
	1	5
	3	7

r không thỏa $A \rightarrow B$, nhưng thỏa $B \rightarrow A$

Phụ thuộc hàm (2)



- $r \in R$ thỏa các ràng buộc PTH được gọi là trạng thái hợp lệ của R .
 - Nhận xét
 - Các PTH xuất phát từ các ràng buộc trong thế giới thực.
 - $\forall \forall r \in R, \forall t \in r, t[X]$ là duy nhất thì X là một khóa của R .
 - Nếu K là một khóa của R thì K xác định hàm tất cả các tập thuộc tính của R .
 - PTH dùng để đánh giá một thiết kế CSDL.
-

Bao đóng của tập PTH

- F là tập PTH trên R
 - $F = \{ \text{MaNV} \rightarrow \text{TenNV}, \text{MaPB} \rightarrow \{ \text{TenPB}, \text{TrPhong} \}, \text{MaNV} \rightarrow \text{MaPB} \}$.
 - $\forall \forall r \in R$ thỏa F và $\text{MaNV} \rightarrow \{ \text{TenPB}, \text{TrPhong} \}$ cũng đúng với r thì $\text{MaNV} \rightarrow \{ \text{TenPB}, \text{TrPhong} \}$ gọi là được suy diễn từ F .
 - Bao đóng của F , ký hiệu F^+ , gồm
 - F và
 - Tất cả các PTH được suy diễn từ F .
 - F gọi là đầy đủ nếu $F = F^+$.
-

Luật suy diễn

- Luật suy diễn dùng để suy diễn một PTH mới từ một tập PTH cho trước.
 - Hệ luật suy diễn Armstrong
 - Phản xạ: $Y \subseteq X \Rightarrow X \rightarrow Y$.
 - Tăng trưởng: $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$, với $XZ = X \cup Z$.
 - bắc cầu: $X \rightarrow Y, Y \rightarrow Z \Rightarrow X \rightarrow Z$.
 - Các luật khác:
 - Phân rã: $X \rightarrow YZ \Rightarrow X \rightarrow Y, X \rightarrow Z$.
 - Hợp: $X \rightarrow Y, X \rightarrow Z \Rightarrow X \rightarrow YZ$.
 - bắc cầu giả: $X \rightarrow Y, WY \rightarrow Z \Rightarrow WX \rightarrow Z$.
 - Nhận xét
 - Hệ luật Armstrong là đầy đủ.
-

Bao đóng của tập thuộc tính

- Làm thế nào để biết một PTH $X \rightarrow Y$ được suy diễn từ tập PTH F cho trước?
 - Bao đóng của tập thuộc tính X đối với F , ký hiệu X^+ , là
 - Tập các thuộc tính PTH vào X .
 - $X^+ = \{A \in U \mid X \rightarrow A \in F^+\}$
 - Nhận xét
 - $X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X^+$.
 - Nếu K là khóa của R thì $K^+ = U$.
-

Thuật toán tìm X^+

- Nhập: U , F và $X \subseteq U$
 - Xuất: X^+
 - Thuật toán 7.1
 - $B1$: $X^+ = X$;
 - $B2$: Nếu tồn tại $Y \rightarrow Z \in F$ và $Y \subseteq X^+$ thì
$$X^+ := X^+ \cup Z;$$
và tiếp tục $B2$. Ngược lại qua $B3$.
 - $B3$: xuất X^+ .
-

Ví dụ tìm X^+

- Cho:

- $F = \{AB \rightarrow C, BC \rightarrow D, D \rightarrow EG\}$.
- $X = BD$.

- Tính X^+ :

- $X^+ = BD$.
 - Lặp 1:
 - Tìm các PTH có vế trái là tập con của $X^+ = BD$
 - + $D \rightarrow EG$, thêm EG vào X^+ ta được $X^+ = BDEG$.
 - Lặp 2:
 - Tìm các PTH có vế trái là tập con của $X^+ = BDEG$
 - + Không có PTH nào.
 - Vậy $X^+ = BDEG$.
-

Kiểm tra PTH suy diễn

- Cho $F = \{AB \rightarrow C, A \rightarrow D, D \rightarrow E, AC \rightarrow B\}$
- Hai PTH $AB \rightarrow E$ và $D \rightarrow C$ có được suy diễn từ F hay không?

X	X_F^+
AB	ABCDE
D	DE



Được suy diễn từ F

Các tập PTH tương đương

- Tập PTH F được nói là phủ tập PTH G nếu $G \subset F^+$.
 - Hai tập PTH F và G là tương đương nếu
 - F phủ G và
 - G phủ F .
 - Nhận xét
 - $\forall \forall X \rightarrow Y \in G$, nếu $Y \subseteq X_F^+$ thì F phủ G .
 - F và G tương đương nếu và chỉ nếu $F^+ = G^+$.
-

Tập PTH tối thiểu (1)

▪ Thừa PTH

- $\{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$, vì $A \rightarrow C$ được suy diễn từ $\{A \rightarrow B, B \rightarrow C\}$
 $A \rightarrow B, B \rightarrow C \Rightarrow A \rightarrow C$ (luật bắc cầu).

▪ Thừa thuộc tính

- $\{A \rightarrow B, B \rightarrow C, A \rightarrow CD\}$, vì $A \rightarrow CD$ được suy diễn từ $\{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$
 $A \rightarrow B, B \rightarrow C \Rightarrow A \rightarrow C$ (luật bắc cầu)
 $A \rightarrow C, A \rightarrow D \Rightarrow A \rightarrow CD$ (luật hợp).
 - $\{A \rightarrow B, B \rightarrow C, AC \rightarrow D\}$, vì $AC \rightarrow D$ được suy diễn từ $\{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$
 $A \rightarrow B, A \rightarrow D \Rightarrow A \rightarrow BD$ (luật hợp)
 $A \rightarrow BD \Rightarrow AC \rightarrow BCD$ (luật tăng trưởng)
 $AC \rightarrow BCD \Rightarrow AC \rightarrow D$ (luật phân rã).
-

Tập PTH tối thiểu (2)

- Tập PTH F là tối thiểu nếu thỏa các điều kiện sau
 - Mọi PTH của F chỉ có một thuộc tính ở vế phải.
 - Không thể thay $X \rightarrow A$ thuộc F bằng $Y \rightarrow A$ với $Y \subset X$ mà tập mới tương đương với F .
 - Nếu bỏ đi một PTH bất kỳ trong F thì tập PTH còn lại không tương đương với F .
 - Phủ tối thiểu của tập PTH E là tập PTH tối thiểu F tương đương với E .
 - Nhận xét
 - Mọi tập PTH có ít nhất một phủ tối thiểu.
-

Thuật toán tìm phủ tối thiểu

- Nhập: tập PTH E.
 - Xuất: phủ tối thiểu F của E.
 - Thuật toán 7.2
 - B1: $F := \emptyset$.
 - B2: Với mọi $X \rightarrow Y \in E$, $Y = \{A_1, \dots, A_k\}$, $A_i \in U$
 $F := F \cup \{X \rightarrow \{A_i\}\}$.
 - B3: Với mỗi $X \rightarrow \{A\} \in F$, $X = \{B_1, \dots, B_l\}$, $B_i \in U$
Với mỗi B_i , nếu $A \in (X - \{B_i\})_F^+$ thì
 $F := (F - \{X \rightarrow \{A\}\}) \cup \{(X - \{B_i\}) \rightarrow \{A\}\}$.
 - B4: Với mỗi $X \rightarrow \{A\} \in F$
 $G := F - \{X \rightarrow \{A\}\}$
Nếu $A \in X_G^+$ thì $F := F - \{X \rightarrow \{A\}\}$.
-

Ví dụ tìm phủ tối thiểu

- Tìm phủ tối thiểu của $E = \{A \rightarrow BC, A \rightarrow B, B \rightarrow C, AB \rightarrow C\}$
 - $B1: F = \emptyset$.
 - $B2: F = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, AB \rightarrow C\}$.
 - $B3: \text{Xét } AB \rightarrow C$

$(B)_F^+ = BC$
 $F = \{A \rightarrow B, A \rightarrow C, B \rightarrow C\}$.
 - $B4: A \rightarrow C \text{ thừa.}$

$F = \{A \rightarrow B, B \rightarrow C\}$.
-

Siêu khóa và khóa

■ Cho $R(U)$

- $S \subseteq U$ là siêu khóa nếu $\forall r \in R, \forall t_1, t_2 \in r, t_1 \neq t_2$ thì $t_1[S] \neq t_2[S]$.
- $K \subseteq U$ là khóa nếu K là siêu khóa nhỏ nhất.
 - $A \in K$ được gọi là thuộc tính khóa.

■ Nhận xét

- S xác định hàm tất cả các thuộc tính của R .
 - R có thể có nhiều khóa.
-

Xác định khóa của lược đồ

- Nhập: tập PTH F xác định trên lược đồ $R(U)$.
 - Xuất: khóa K của R .
 - Thuật toán 7.3.1
 - $B1$:
 $K = U = \{A_1, \dots, A_n\};$
 $i = 1;$
 - $B2$:
Nếu $U \subseteq (K - \{A_i\})_F^+$ thì $K = K - \{A_i\}.$
 $i = i + 1;$
Nếu $i > n$ thì sang $B3$. Ngược lại, tiếp tục $B2$.
 - $B3$:
Xuất K .
-

Ví dụ tìm khóa của lược đồ

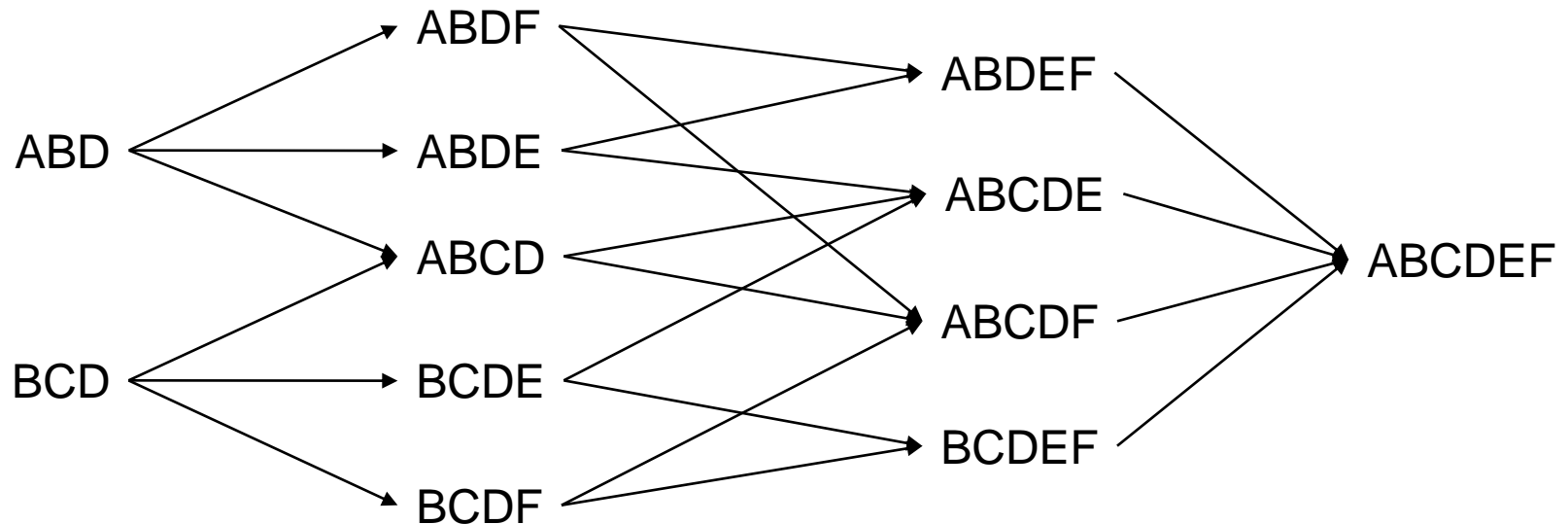
- Cho $R(U)$, $U = \{A, B, C, D, E, F, G\}$.
 - $F = \{B \rightarrow A, D \rightarrow C, D \rightarrow BE, DF \rightarrow G\}$.
 - Tìm khóa của R
 - B1:
 $K = ABCDEFG$.
 - B2:
 - Lặp 1: $(BCDEFG)_F^+ = BCDEFGA \Rightarrow K = BCDEFG$.
 - Lặp 2: $(CDEFG)_F^+ = CDEFGBA \Rightarrow K = CDEFG$.
 - Lặp 3: $(DEFG)_F^+ = DEFGCBA \Rightarrow K = DEFG$.
 - Lặp 4: $(EFG)_F^+ = EFG$.
 - Lặp 5: $(DFG)_F^+ = DFGCBEA \Rightarrow K = DFG$.
 - Lặp 6: $(DG)_F^+ = DGCBEA$.
 - Lặp 7: $(DF)_F^+ = DFCBEAG \Rightarrow K = DF$.
 - B3:
Khóa là $K = DF$.
-

Xác định tất cả khóa của lược đồ

- Nhập: tập PTH F xác định trên lược đồ $R(U)$.
 - Xuất: tất cả khóa của R .
 - Thuật toán 7.3.2
 - $B1$:
Xây dựng 2^n tập con của $U = \{A_1, \dots, A_n\}$;
 $S = \{\}$;
 - $B2$:
Với mỗi tập con $X \subseteq U$
Nếu $U \subseteq X_F^+$ thì $S = S \cup \{X\}$.
 - $B3$:
 $\forall X, Y \in S$, nếu $X \subset Y$ thì $S = S - \{Y\}$.
 - $B4$:
 S là tập các khóa của R .
-

Ví dụ tìm tất cả khóa của lược đồ

- Cho $R(U)$, $U = \{A, B, C, D, E, F\}$.
 - $F = \{AE \rightarrow C, CF \rightarrow A, BD \rightarrow F, AF \rightarrow E\}$.
- Tìm tất cả khóa của R
 - Tập siêu khóa
 $S = \{ABD, BCD, ABCD, ABDE, BCDE, ABCDE, ABDF, BCDF, ABCDF, ABDEF, BCDEF, ABCDEF\}$.



Xác định tất cả các khóa của lược đồ

- Tập hợp các thuộc tính ở vế trái các PTH của F : VT_F
 - Tập hợp các thuộc tính ở vế phải các PTH của F : VP_F
 - Tập nguồn là tập các thuộc tính không xuất hiện ở vế phải các PTH của F : $TN = U - VP_F$
 - Tập đích là tập các thuộc tính có xuất hiện ở vế phải và không xuất hiện ở vế trái các PTH của F : $TD = U - (TN \cup VT_F)$
 - Tập trung gian là tập các thuộc tính vừa xuất hiện trong vế trái, vừa xuất hiện trong vế phải của các PTH của F : $L = U - (TN \cup TD)$
 - Nếu K là khóa của R thì tồn tại $L_i \subseteq L$ sao cho $K = TN \cup L_i$.
-

Nội dung trình bày

- Nguyên tắc thiết kế các lược đồ quan hệ.
 - Phụ thuộc hàm.
 - Các dạng chuẩn.
 - Một số thuật toán chuẩn hóa.
-

Chuẩn hóa lược đồ CSDL

- Chuẩn hóa là gì?
 - Các dạng chuẩn là gì?
 - Các dạng chuẩn
 - Dạng 1 (1 Normal Form - 1NF).
 - Dạng 2 (2 Normal Form - 2NF).
 - Dạng 3 (3 Normal Form - 3NF).
 - Dạng Boyce - Codd (Boyce - Codd Normal Form - BCNF).
-

Dạng chuẩn 1 (1)

- Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 1 nếu và chỉ nếu mọi thuộc tính của R là thuộc tính đơn.

PHONGBAN

TenPB	<u>MaPB</u>	TrPhg	CacTruso
Nghien cuu	5	333445555	Tan Binh, Thu Duc
Hanh chinh	4	987654321	Go Vap

← Không thuộc dạng chuẩn 1

PHONGBAN

TenPB	<u>MaPB</u>	TrPhg	<u>Truso</u>
Nghien cuu	5	333445555	Tan Binh
Nghien cuu	5	333445555	Thu Duc
Hanh chinh	4	987654321	Go Vap

← Thuộc dạng chuẩn 1

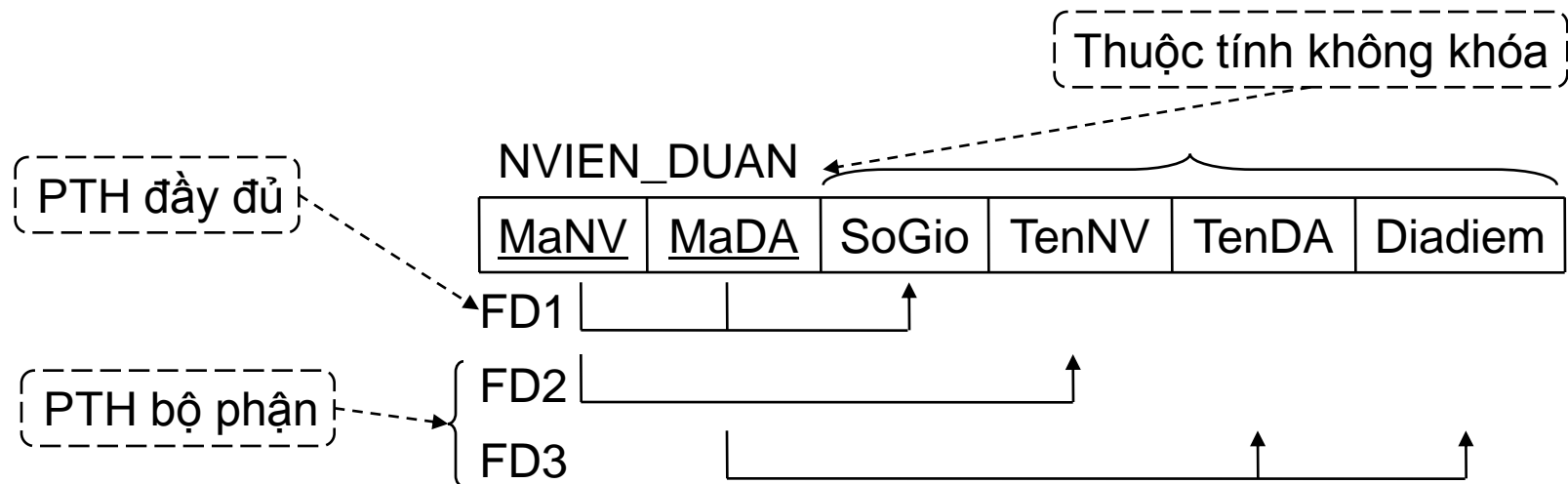
Dạng chuẩn 1 (2)

- Nhận xét

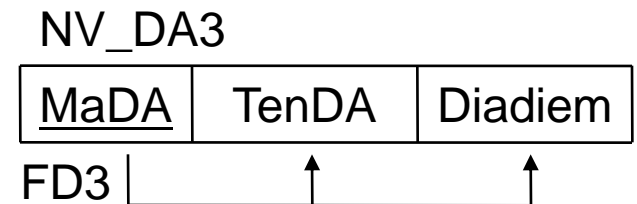
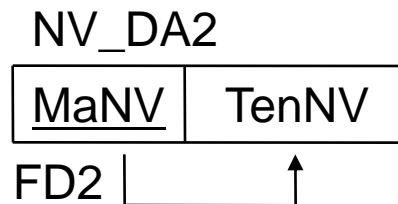
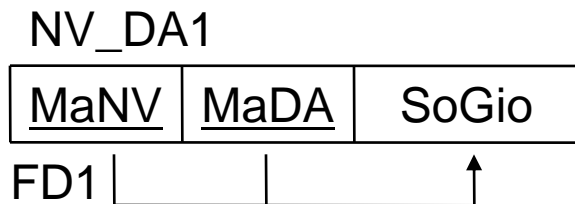
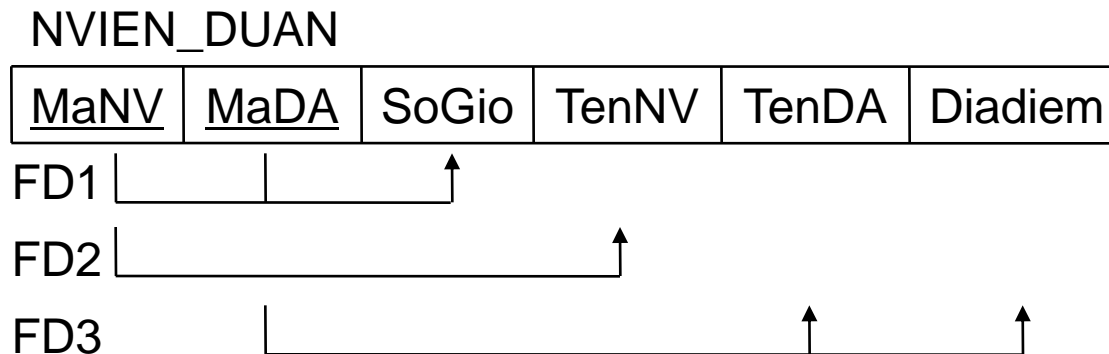
- Mọi lược đồ quan hệ đều thuộc dạng chuẩn 1.
- Dạng chuẩn 1 có thể dẫn đến sự trùng lặp dữ liệu. Do đó gây ra các dị thường về cập nhật dữ liệu.

Dạng chuẩn 2 theo khóa chính (1)

- Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 2 nếu mọi thuộc tính không khóa của R phụ thuộc đầy đủ vào khóa chính của R.
- $R(U)$, $K \subseteq U$ là khóa của R
 - $A \in U$ là thuộc tính không khóa nếu $A \notin K$.
 - $X \rightarrow Y$ là PTH đầy đủ nếu $\forall A \in X$ thì $(X - \{A\}) \rightarrow Y$ không đúng trên R.
Ngược lại $X \rightarrow Y$ là PTH bộ phận.
- Ví dụ

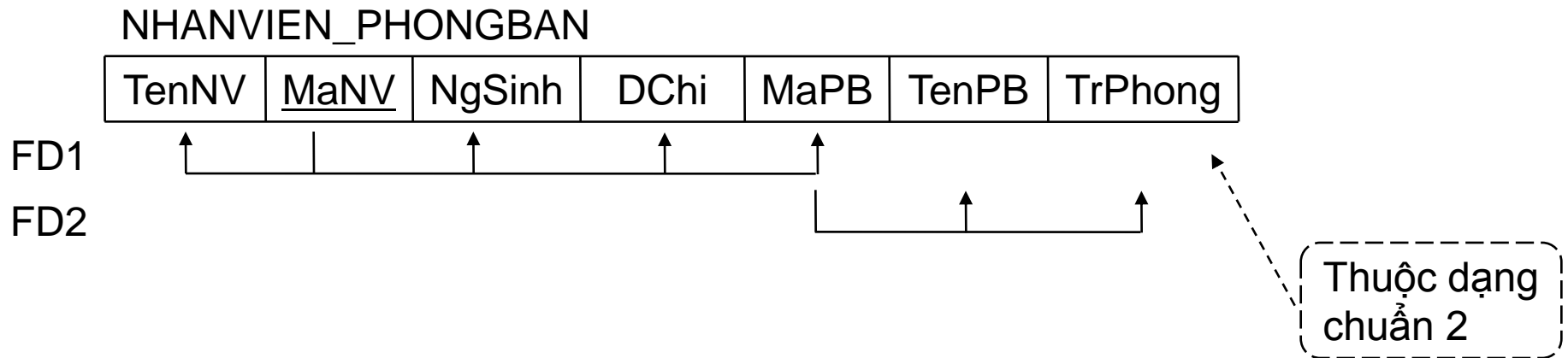


Dạng chuẩn 2 theo khóa chính (2)



3 lược đồ NV_DA1, NV_DA2, NV_DA3 thuộc dạng chuẩn 2

Dạng chuẩn 2 theo khóa chính (3)

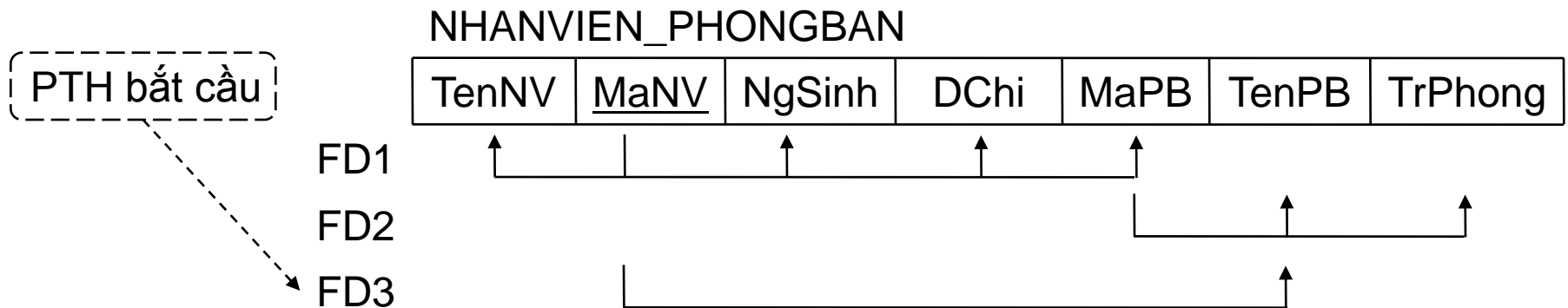


■ Nhận xét

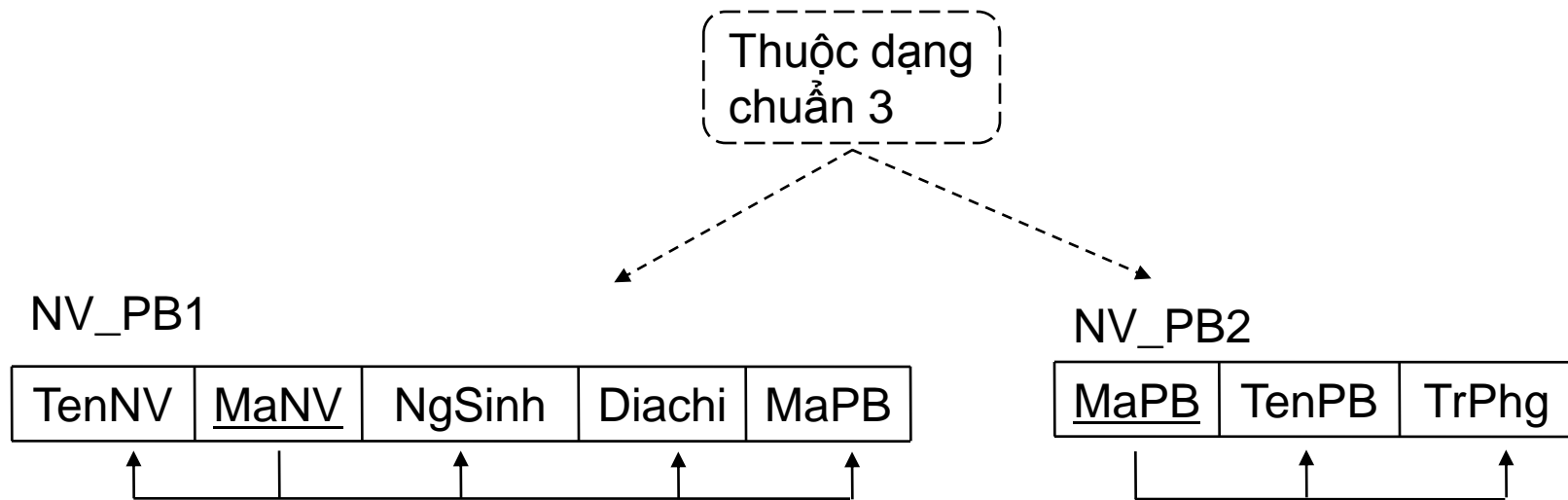
- Mọi lược đồ quan hệ thuộc dạng chuẩn 2 cũng thuộc dạng chuẩn 1.
 - Nếu R chỉ có một khóa K và $\text{card}(K) = 1$ thì R thuộc dạng chuẩn 2.
 - Còn xuất hiện sự trùng lặp dữ liệu. Do đó gây ra các dị thường về cập nhật dữ liệu.
-

Dạng chuẩn 3 theo khóa chính (1)

- Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 3 nếu
 - R thuộc dạng chuẩn 2.
 - Mọi thuộc tính không khóa của R không phụ thuộc bắt cầu vào khóa chính của R.
- Cho R(U)
 - $X \rightarrow Y$ là PTH bắt cầu nếu $\exists Z \subseteq U$, Z không là khóa và cũng không là tập con của khóa của R mà $X \rightarrow Z$ và $Z \rightarrow Y$ đúng trên R.
- Ví dụ



Dạng chuẩn 3 theo khóa chính (2)

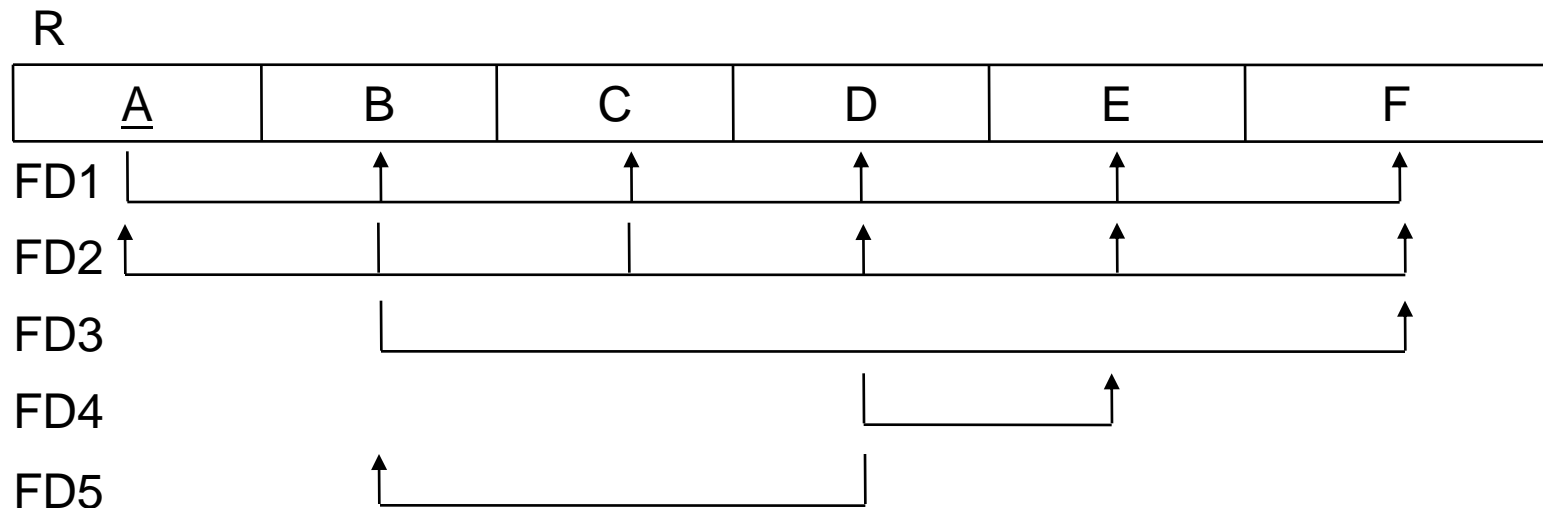


■ Nhận xét

- Mọi lược đồ quan hệ thuộc dạng chuẩn 3 cũng thuộc dạng chuẩn 2.
 - PTH bắt cầu là nguyên nhân dẫn đến trùng lặp dữ liệu.
 - Dạng chuẩn 3 là dạng chuẩn tối thiểu trong thiết kế CSDL.
-

Dạng chuẩn 2 tổng quát

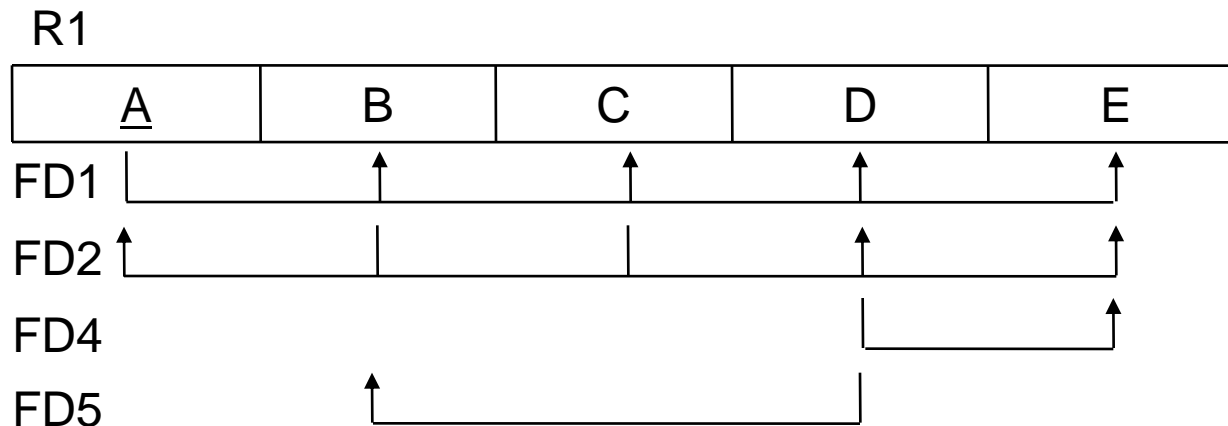
- Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 2 nếu mọi thuộc tính không khóa của R phụ thuộc đầy đủ vào các khóa của R.
- Cho R(ABCDEF) có 2 khóa là A và BC.



Lược đồ R không thuộc dạng chuẩn 2

Dạng chuẩn 3 tổng quát

- Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 3 nếu PTH không hiển nhiên $X \rightarrow A$ đúng trên R thì
 - X là siêu khóa của R, hoặc
 - A là thuộc tính khóa của R.
- R1(ABCDE) có 2 khóa là A và BC.

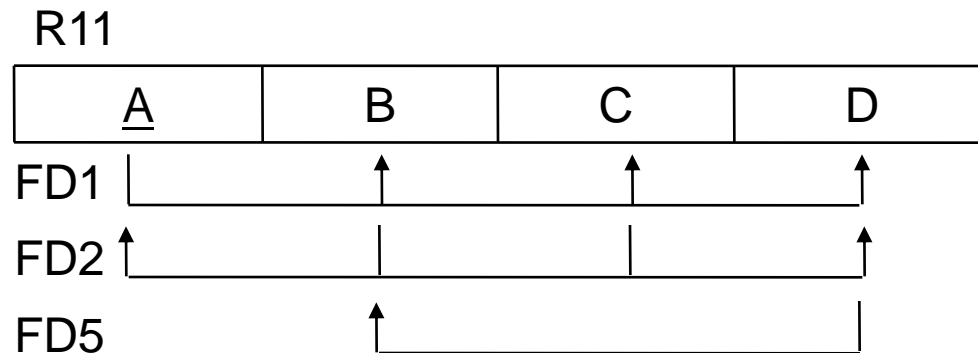


Lược đồ bên
thuộc dạng
chuẩn 2,
nhưng không
thuộc dạng
chuẩn 3

- Nhận xét
 - Định nghĩa tổng quát cho phép kiểm tra dạng chuẩn 3 mà không cần kiểm tra dạng chuẩn 2.

Dạng chuẩn Boyce - Codd (1)

- Lược đồ quan hệ R được gọi là thuộc dạng chuẩn BC nếu PTH không hiển nhiên $X \rightarrow Y$ đúng trên R thì X là siêu khóa của R.
- R11(ABCD)



[Lược đồ R11 thuộc dạng chuẩn 3, nhưng không thuộc dạng chuẩn BC]

Dạng chuẩn Boyce - Codd (2)

R11

<u>A</u>	B	C	D
1	a	a	1
2	a	b	1
3	b	a	2
4	b	b	2

Trùng lặp dữ liệu



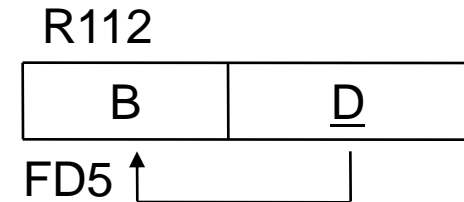
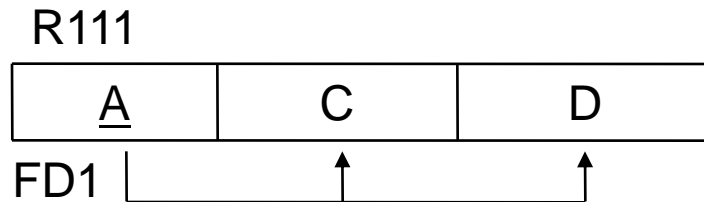
R111

<u>A</u>	C	D
1	a	1
2	b	1
3	a	2
4	b	2

R112

<u>D</u>	B
1	a
2	b

Dạng chuẩn Boyce - Codd (3)



2 lược đồ trên thuộc dạng chuẩn BC

■ Nhận xét

- Mọi lược đồ quan hệ thuộc dạng chuẩn BC cũng thuộc dạng chuẩn 3.
 - Dạng chuẩn BC đơn giản và chặt chẽ hơn dạng chuẩn 3.
 - Mục tiêu của quá trình chuẩn hóa là đưa các lược đồ quan hệ về dạng chuẩn 3 hoặc BC.
-

Thiết kế Top-Down

- Các bước thực hiện

- Thiết kế lược đồ mức khái niệm với mô hình dữ liệu cấp cao (EER).
 - Chuyển lược đồ khái niệm thành tập hợp các quan hệ.
 - Với mỗi quan hệ xác định tập PTH.
 - Áp dụng các quy tắc chuẩn hóa để loại bỏ các PTH bộ phận và bắt cầu trong các quan hệ.
-

Nội dung trình bày

- Nguyên tắc thiết kế các lược đồ quan hệ.
 - Phụ thuộc hàm.
 - Các dạng chuẩn.
 - Một số thuật toán chuẩn hóa.
-

Phân rã lược đồ quan hệ

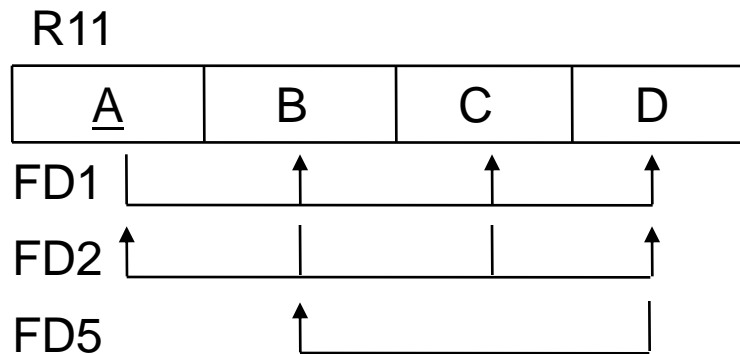
- Lược đồ quan hệ chung $R(A_1, \dots, A_n)$
 - Tập hợp tất cả các thuộc tính của các thực thể.
 - Xác định tập PTH F trên R .
 - Phân rã
 - Sử dụng các thuật toán chuẩn hóa để tách R thành tập các lược đồ $D = \{R_1, \dots, R_m\}$.
 - Yêu cầu
 - Bảo toàn thuộc tính.
 - Các lược đồ R_i phải ở dạng chuẩn 3 hoặc Boyce-Codd.
-

Phân rã bảo toàn PTH (1)

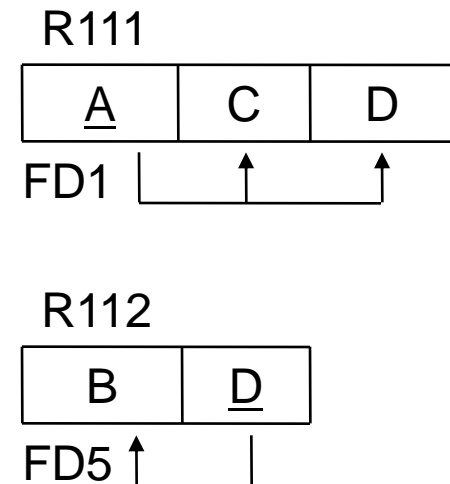
■ Tính chất bảo toàn PTH

- Xét lược đồ R và tập PTH F . Giả sử R được phân rã thành $D = \{R_1, \dots, R_m\}$.
 - Đặt $\pi_{R_i}(F) = \{X \rightarrow Y \in F^+ : X \cup Y \subset R_i\}$.
 - D được gọi là phân rã bảo toàn phụ thuộc hàm đối với F nếu $(\pi_{R_1}(F) \cup \dots \cup \pi_{R_m}(F))^+ = F^+$.

■ Ví dụ



----->



Phân rã bảo toàn PTH (2)

R11	<u>A</u>	B	C	D
	1	α	β	2
	2	β	γ	3
	3	α	δ	2

R111	<u>A</u>	C	D
	1	β	2
	2	γ	3
	3	δ	2
	4	β	4

R112	<u>D</u>	B
	2	α
	3	β
	4	α

<u>A</u>	B	C	D
1	α	β	2
...
4	α	β	4

Thêm bộ (4, β , 4) vào R111
và (4, α) vào R112
thì trạng thái csdl sẽ không
thỏa PTH FD2

Phân rã bảo toàn PTH (3)

▪ Định lý 7.1

- Tồn tại một phân rã bảo toàn PTH $D = \{R_1, \dots, R_m\}$ của lược đồ R đối với tập PTH F sao cho các R_i ở dạng chuẩn 3.

▪ Thuật toán 7.4

- 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 3.
 - *B1*:
 - Tìm phủ tối thiểu G của F .
 - *B2*:
 - Với mỗi $X \rightarrow A_i \in G$, xây dựng lược đồ $R_i(U_i)$, $U_i = X \cup \{A_i\}$. Khóa chính của R_i là X .
-

Phân rã bảo toàn PTH (4)

- *B3*:
 - Giả sử xong B2 ta có các lược đồ R_1, \dots, R_m . Nếu $U_1 \cup \dots \cup U_m \neq U$ thì xây dựng thêm lược đồ $R_{m+1}(U_{m+1})$, $U_{m+1} = U - (U_1 \cup \dots \cup U_m)$. Khóa chính của R_{m+1} là U_{m+1} .
- *B4*:
 - Xuất các lược đồ R_i .

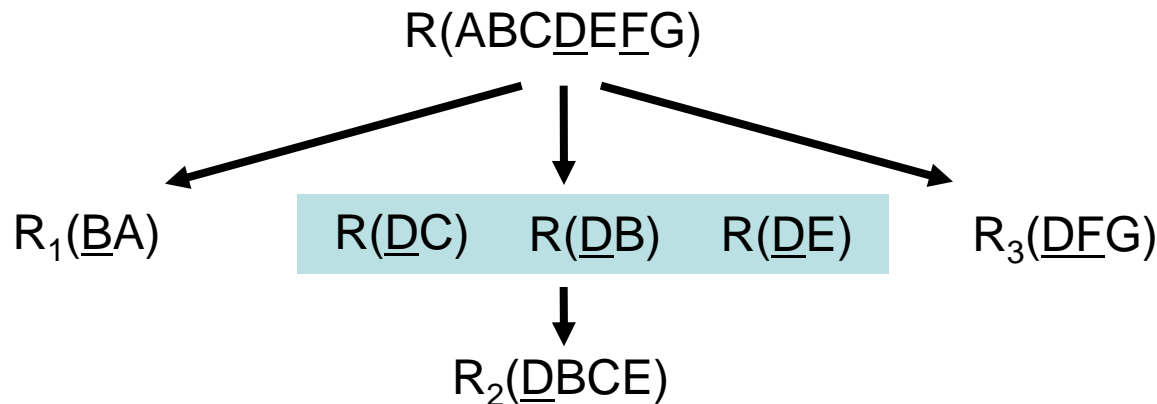
Ví dụ phân rã bảo toàn PTH (1)

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

$$\begin{array}{ccc} & R(ABC\underline{D}EFGHI) & \\ & \swarrow \downarrow \searrow & \\ R_1(\underline{B}A) & R_2(\underline{D}BCE) & R_3(\underline{D}FG) \end{array}$$
 - $B3$:
 - Vì $U_1 \cup U_2 \cup U_3 = \{ABCDEFG\}$ nên đặt $R_4(\underline{H}I)$.
 - $B4$:
 - $D = \{R_1, R_2, R_3, R_4\}$.
-

Phân rã không mất thông tin (1)

▪ Tính chất không mất thông tin

- Xét lược đồ R và tập PTH F . Giả sử R được phân rã thành $D = \{R_1, \dots, R_m\}$.
 - D được gọi là phân rã không mất thông tin đối với F nếu với mọi trạng thái $r \in R$ thì $(\pi_{R_1}(r) * \dots * \pi_{R_m}(r)) = r$.

▪ Định lý 7.2

- Phân rã $D = \{R_1(U_1), R_2(U_2)\}$ của $R(U)$ không mất thông tin đối với tập PTH F nếu và chỉ nếu:
 - $(U_1 \cap U_2) \rightarrow (U_1 - U_2) \in F^+$, hoặc
 - $(U_1 \cap U_2) \rightarrow (U_2 - U_1) \in F^+$.

▪ Định lý 7.3

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

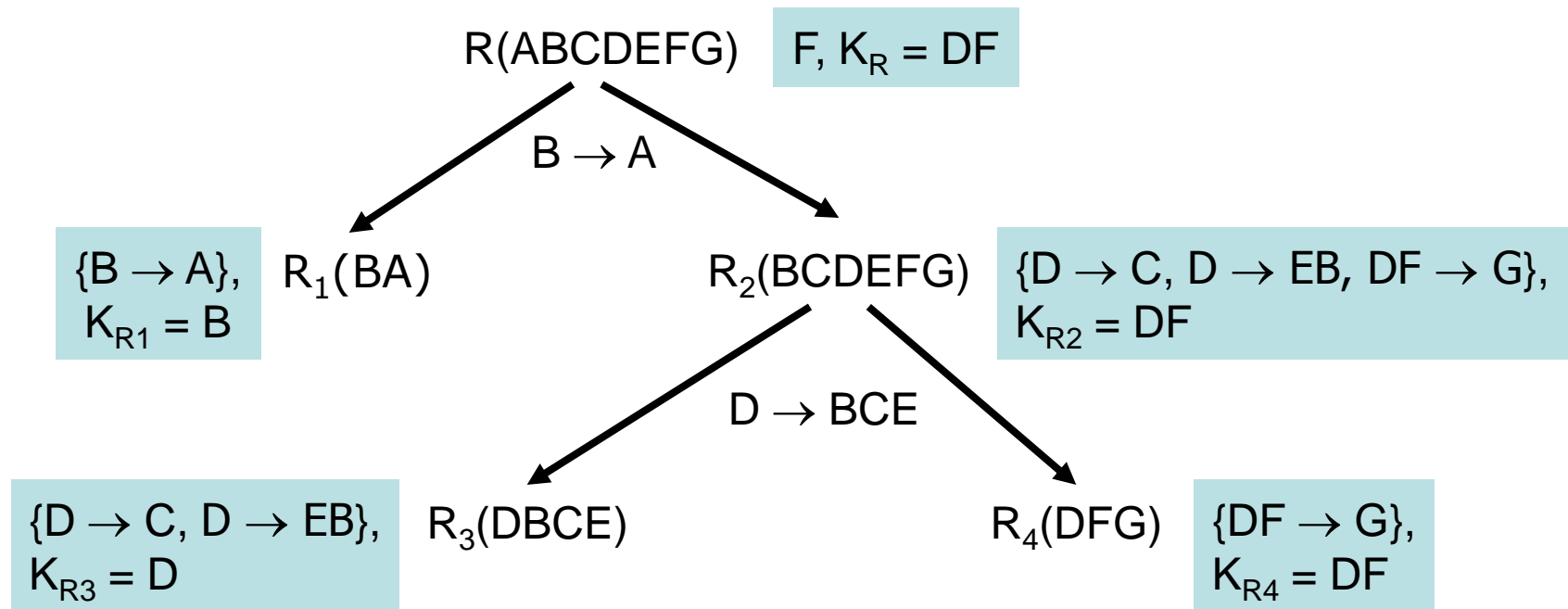
Phân rã không mất thông tin (2)

▪ Thuật toán 7.5

- 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 không mất thông tin (1)

- Cho:
 - $R(ABCDEFG)$
 - $F = \{B \rightarrow A, D \rightarrow C, D \rightarrow EB, DF \rightarrow G\}$
- Tách về dạng chuẩn BC, không mất thông tin.



Ví dụ phân rã không mất thông tin (2)

