
Chương 4

4.1. Chuẩn hóa CSDL quan hệ

Nội dung chi tiết

- Giới hạn của ER
- Sự dư thừa
- Phụ thuộc hàm
- Hệ suy diễn Amstrong
- Thuật toán tìm bao đóng X^+_F
- Tìm phủ tối thiểu
- Các dạng chuẩn

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

- 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	...

Sự dư thừa (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

Sự dư thừa (tt)

- Sự dư thừa → sự 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_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	...

Sự dư thừa (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
 - Sử dụng các phép nối

Sự dư thừa (tt)

■ Một số nguyên tắc

- NT1: Rõ ràng về mặt ngữ nghĩa, tránh các 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 đặt 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”
 - Gây lỗi khi thực hiện các phép kết nối

Phụ thuộc hàm

- Lý thuyết về chuẩn hóa
 - Các phân tích để đưa ra lược đồ thực thể liên kết cần phải được sửa chữa ở các bước tiếp theo
 - Vấn đề nêu ở slide trên sẽ được giải quyết nếu có một phương pháp phân tích thích hợp
- ➔ lý thuyết chuẩn hóa (dựa trên phụ thuộc hàm, ...) sẽ là nền tảng cơ sở để thực hiện việc phân tích và chuẩn hóa lược đồ ER

Phụ thuộc hàm (tt)

- ĐN 1: Phụ thuộc hàm (FD-function dependancy) trên một lược đồ quan hệ R là một ràng buộc $X \rightarrow Y$, với X và Y là một tập các thuộc tính trong R
- ĐN 2: $(X \rightarrow Y)$ với mỗi thể hiện r của lược đồ quan hệ R: với 2 bộ bất kỳ t và s trong r nếu $t[X] = s[X]$ thì $t[Y] = s[Y]$
- Ví dụ: Ràng buộc dữ liệu là một trường hợp đặc biệt của phụ thuộc hàm
 - $MaNV \rightarrow TenNV, NS \dots$

Phụ thuộc hàm (tt)

- Ví dụ
 - Ngày sinh → Tuổi
 - Tuổi → Quyền lợi
 - MaNV → Tên NV
 - ???

Phụ thuộc hàm (tt)

■ Ví dụ:

- Ta có lược đồ quan hệ

MUON(SoTHE, MaSACH, NGUOIMuon, TenSACH, THOIGIAN)

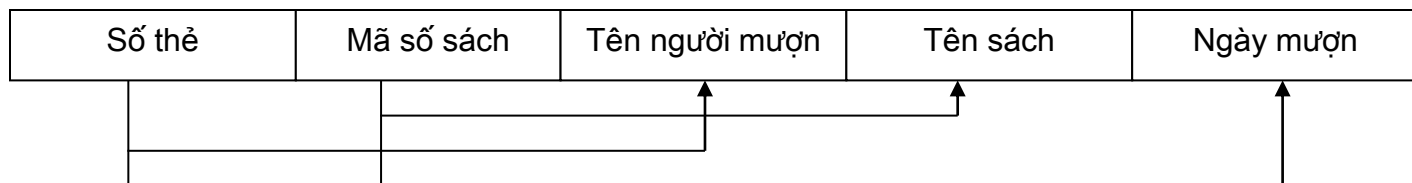
- Với các phụ thuộc hàm:

SoTHE → NGUOIMuon

MaSACH → TenSACH

SoTHE, MaSACH → THOIGIAN

- Có sơ đồ phụ thuộc hàm như sau:



Phụ thuộc hàm (tt)

- ĐN bao đóng: Nếu F là tập các FD trong lược đồ R và f là FD khác cũng trong R , thì F được coi là bao f nếu với mọi thể hiện r của R nếu thỏa mãn FD trong F thì cũng thỏa mãn f .
 - Ví dụ $F=\{A\rightarrow B, B\rightarrow C\}$ và $f=\{A\rightarrow C\}$
 - $F=\{ĐToan, DLy, DHoa\rightarrow DTB, DTB\rightarrow XepHang\}$,
 $f=\{DT, DL, DH\rightarrow XepHang\}$
- Bao đóng của tập F (Ký hiệu F^+) là tập các FD có thể suy diễn được từ F
- F và G được coi là tương đương nếu F bao G và G bao F

Phụ thuộc hàm (tt)

- Ký hiệu $F \models X \rightarrow Y$: phụ thuộc hàm $X \rightarrow Y$ được suy diễn từ tập các phụ thuộc hàm F
- QT1 (quy tắc phản xạ) : Nếu $X \supseteq Y$ thì $X \rightarrow Y$
- QT2 (quy tắc tăng) : $\{ X \rightarrow Y \} \models XZ \rightarrow YZ$
- QT3 (quy tắc bắc cầu): $\{ X \rightarrow Y, Y \rightarrow Z \} \models X \rightarrow Z$
- QT4 (quy tắc chiếu) : $\{ X \rightarrow YZ \} \models X \rightarrow Y$ và $X \rightarrow Z$
- QT5 (quy tắc hợp) : $\{ X \rightarrow Y, X \rightarrow Z \} \models X \rightarrow YZ$
- QT6 (quy tắc tựa bắc cầu): $\{ X \rightarrow Y, WY \rightarrow Z \} \models WX \rightarrow Z$

Hệ suy diễn Amstrong

- Quy tắc suy diễn Amstrong đưa ra cách thức để tính toán và kiểm tra các thuộc tính trong tập FD
- Bao gồm 3 quy tắc 1-3 (phản xạ, tăng, bắc cầu)
 - QT1 (quy tắc phản xạ) :
 - $TenNV, DChi \rightarrow TenNV$
 - QT2 (quy tắc tăng) :
 - $MaNV \rightarrow TenNV$ thì $MaNV, NS \rightarrow TenNV, NS$
 - QT3 (quy tắc bắc cầu) : $\{ X \rightarrow Y, Y \rightarrow Z \} \models X \rightarrow Z$
 - Nếu $DT, DL, DH \rightarrow DTB, DTB \rightarrow XepL$ thì $DT, DL, DH \rightarrow XepL$

Hệ suy diễn Armstrong (tt)

- Hệ Ams là đúng: nếu FD $f:X \rightarrow Y$ có thể được suy diễn từ tập các FD F sử dụng các quy tắc suy diễn thì f nằm trong các quan hệ mà thỏa mãn tất cả các FD trong F
- Ví dụ Cho biết $X \rightarrow Y$ và $X \rightarrow Z$ thì
 - $X \rightarrow XY$ (quy tắc tăng theo X)
 - $YX \rightarrow YZ$ (quy tắc tăng theo Y)
 - $X \rightarrow YZ$ (bắc cầu)
- Vậy $X \rightarrow YZ$ thỏa mãn tất cả các quan hệ mà thỏa mãn FD $X \rightarrow Y$ và $X \rightarrow Z$

Hệ suy diễn Armstrong (tt)

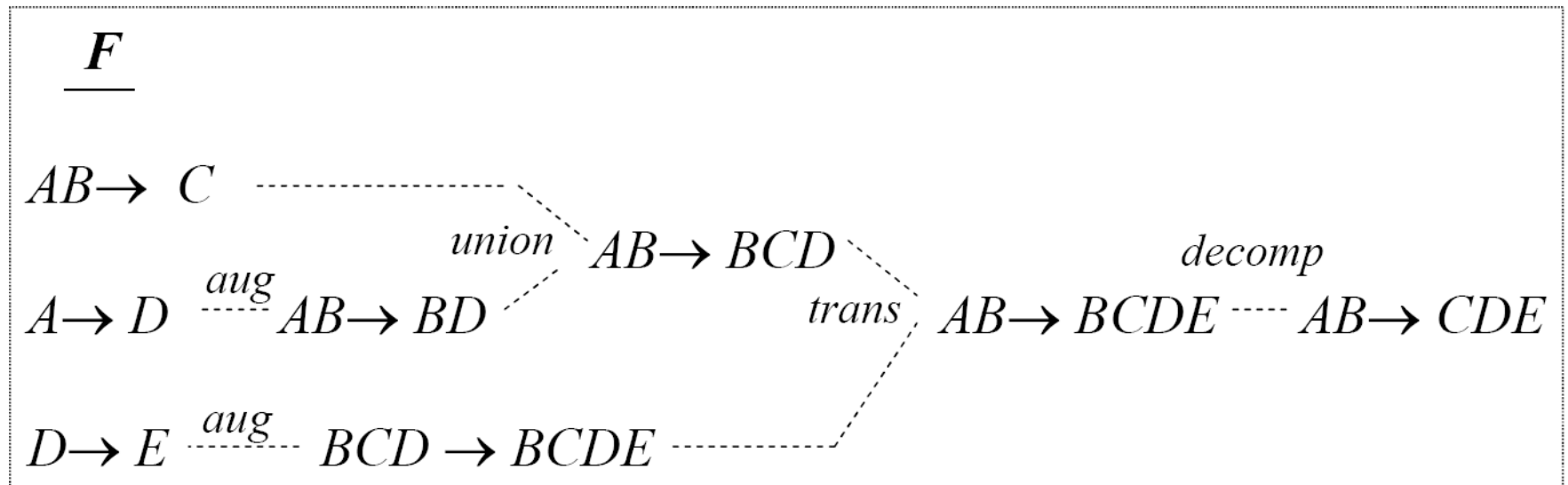
- Hệ Ams là đầy đủ: Nếu F bao f , thì f có thể suy diễn được từ F sử dụng hệ các quy tắc suy diễn
- Kết quả rút ra được từ tính đầy đủ này là chúng ta có thuật toán để xác định xem F có bao f hay không
 - Bản chất thuật toán là sử dụng hệ suy diễn theo tất cả các cách có thể nhằm tìm F^+ , sau đó kiểm tra xem f có nằm trong F^+ hay không

Hệ suy diễn Armstrong (tt)

- Hệ Ams là chính xác: Khái niệm đúng và đầy đủ đã liên kết thành một chuỗi ý nghĩa đầy đủ về tính chính xác của hệ suy diễn Armstrong (định nghĩa này chỉ đúng trong các thể hiện của quan hệ)
- Điều này đồng thời cho biết một cách chính xác rằng thuật toán tìm bao dựa trên hệ suy diễn là chính xác

Hệ suy diễn Armstrong (tt)

- Tìm F^+



- Tất cả các FD bao gồm $AB \rightarrow BD$, $AB \rightarrow BCD$, $BCD \rightarrow BCDE$, $AB \rightarrow CDE$ là các phần tử của F^+

Thuật toán tìm bao đóng X^+_F

- Xác định thuộc tính đóng là cách hiệu quả nhất để tìm bao đóng
- Tập các thuộc tính đóng của tập các thuộc tính (X) với điều kiện thỏa mãn tập các FD (F) (ký hiệu X^+_F) là tập tất cả các thuộc tính (A) sao cho $X \rightarrow A$
- ➔ Gọi là tập các thuộc tính phụ thuộc hàm vào X trên F
- $X^+_{F_1}$ không nhất thiết phải bằng $X^+_{F_2}$ nếu $F_1 \leftrightarrow F_2$
- Tập các thuộc tính đóng và suy diễn
 - Thuật toán: Cho biết tập các FD F ta có $X \rightarrow Y$ nếu và chỉ nếu $X^+_F \supset Y$

Ví dụ

$F: AB \rightarrow C$

$A \rightarrow D$

$D \rightarrow E$

$AC \rightarrow B$

X	X_F^+
A	$\{A, D, E\}$
AB	$\{A, B, C, D, E\}$ (Hence AB is a key)
B	$\{B\}$
D	$\{D, E\}$

- $AB \rightarrow E$ có suy diễn được từ F không?
- $D \rightarrow C$ có suy diễn được từ F không?

Thuật toán tìm bao đóng $X^+_F(tt)$

- $X^+ = X$;
- **Repeat**
 - Old $X^+ = X^+$;
 - Với mỗi phụ thuộc hàm $Y \rightarrow Z$ trong F thực hiện
 - nếu $X^+ \supset Y$ thì $X^+ = X^+ \cup Z$;
- **until** ($X^+ = \text{Old } X^+$);
- Nếu T thuộc X^+ thì $X \rightarrow T$ là suy diễn được từ F

Ví dụ

- Xác định bao đóng
- Bài toán: Tìm bao đóng của AB với các phụ thuộc hàm sau

$$AB \rightarrow C \quad (a)$$

$$A \rightarrow D \quad (b)$$

$$D \rightarrow E \quad (c)$$

$$AC \rightarrow B \quad (d)$$

- Giải
 - Khởi tạo: $X^+ = \{AB\}$
 - Dùng (a): $X^+ = \{ABC\}$
 - Dùng (b): $X^+ = \{ABCD\}$
 - Dùng (c): $X^+ = \{ABCDE\}$

Phụ thuộc hàm tương đương

- Một tập hợp các phụ thuộc hàm E được phủ bởi một tập các phụ thuộc hàm F (hoặc F phủ E) nếu
 - mỗi một phụ thuộc hàm trong E đều ở trong F^+
 - mỗi phụ thuộc hàm trong E có thể suy diễn được từ F .
- Hai tập phụ thuộc hàm E và F là tương đương nếu $E^+ = F^+$.
 - Như vậy tương đương có nghĩa là mỗi phụ thuộc hàm trong E có thể suy diễn được từ F và mỗi phụ thuộc hàm trong F có thể suy diễn được từ E .

Ví dụ

- Xét hai tập phụ thuộc hàm
 - $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow AD, E \rightarrow H\}$
 - $E = \{A \rightarrow CD, E \rightarrow AH\}$
- Kiểm tra xem E tương đương F ?

Phụ thuộc hàm tối thiểu

- Định nghĩa: 1 tập FD gọi là tối thiểu nó thỏa mãn các điều kiện sau
 - Vế phải của các FD trong F chỉ có 1 thuộc tính
 - Không thể thay thế $X \rightarrow A$ bằng $Y \rightarrow A$ với điều kiện Y là tập con của X và vẫn giữ được tập các phụ thuộc mà tương đương với F
 - Không thể bớt được bất kỳ phụ thuộc hàm nào sao cho bảo toàn được tập các phụ thuộc hàm trong F

Phụ thuộc hàm tối thiểu (tt)

■ Thuật toán tìm phủ tối thiểu

1. $G := F$;

2. Thay thế $X \rightarrow \{A_1, A_2, \dots, A_n\}$ trong G bằng n phụ thuộc hàm $X \rightarrow A_1, X \rightarrow A_2, \dots, X \rightarrow A_n$.

3. Với mỗi $X \rightarrow A$ trong G

1. Với mỗi thuộc tính B là một phần tử của X

1. Nếu $((G - (X \rightarrow A)) \cup ((X - \{B\}) \rightarrow A))$ là tương đương với G

2. thì thay thế $X \rightarrow A$ bằng $(X - \{B\}) \rightarrow A$ ở trong G

4. Với mỗi phụ thuộc hàm $X \rightarrow A$ còn lại trong G

1. Nếu $(G - \{X \rightarrow A\})$ là tương đương với G

2. thì loại bỏ $X \rightarrow A$ ra khỏi G

Ví dụ

Tìm phủ tối thiểu cho tập phụ thuộc hàm:

$$\{A \rightarrow BC, B \rightarrow AC, C \rightarrow AB\}$$

Áp dụng thuật toán trên, chúng ta có

- B1: $G = \{A \rightarrow BC, B \rightarrow AC, C \rightarrow AB\}$
- B2: $G = \{A \rightarrow B, A \rightarrow C, B \rightarrow A, B \rightarrow C, C \rightarrow A, C \rightarrow B\}$
- B3:
 - Do $A \rightarrow B$ và $B \rightarrow C$ nên $A \rightarrow C$ là thừa. Do $C \rightarrow B$ và $B \rightarrow A$ nên $C \rightarrow A$ là thừa. Bỏ những phụ thuộc hàm thừa đi, ta có $\{A \rightarrow B, B \rightarrow A, B \rightarrow C, C \rightarrow B\}$ là một phủ tối thiểu.
 - Do $A \rightarrow B$ và $B \rightarrow C$ nên $A \rightarrow C$ là thừa. Do có $B \rightarrow C$ và $C \rightarrow A$ nên $B \rightarrow A$ là thừa. Do có $C \rightarrow A$ và $A \rightarrow B$ nên $C \rightarrow B$ là thừa. Bỏ những phụ thuộc hàm thừa đi, ta nhận được một phủ tối thiểu khác là $\{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$

Thuật toán tìm khóa

- Tìm khóa K của R dựa trên tập phụ thuộc hàm F
 - Đặt $K := R$;
 - Với mỗi thuộc tính A trong K
 - {tính $(K-A)^+$ đối với F ;
 - Nếu $(K-A)^+$ chứa tất cả các thuộc tính trong R thì đặt $K := K - \{A\}$;
- Cách 2
 - Tìm siêu khóa S :
 - Tìm R_F là tập các thuộc tính ở về trái
 - Tìm bao đóng của các tập con trong R_F
 - Áp dụng bước 2 ở trên cho siêu khóa S

Các 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

Các dạng chuẩn (tt)

- Thủ tục chuẩn hoá cung cấp
 - Một cơ cấu hình thức để phân tích các lược đồ quan hệ dựa trên **các khoá** của nó và **các phụ thuộc hàm giữa các thuộc tính** của nó.
 - Một loạt các **kiểm tra dạng chuẩn** có thể thực hiện trên các lược đồ quan hệ riêng rẽ sao cho cơ sở dữ liệu quan hệ có thể được chuẩn hoá đến một mức cần thiết.
- Tính chất
 - Nói không mất mát (hoặc nói không phụ thêm)- vd:bộ giả
 - Bảo toàn sự phụ thuộc
 - nó đảm bảo rằng từng phụ thuộc hàm sẽ được biểu hiện trong các quan hệ riêng rẽ nhận được sau khi tách.

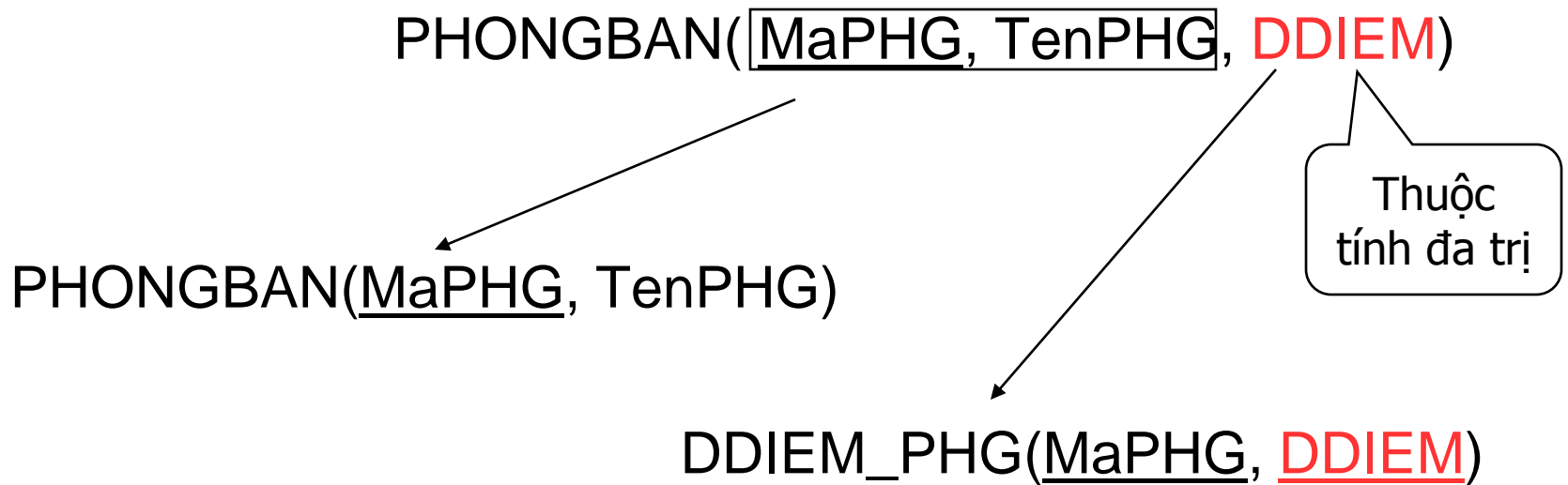
Các dạng chuẩn (tt)

■ Phân loại

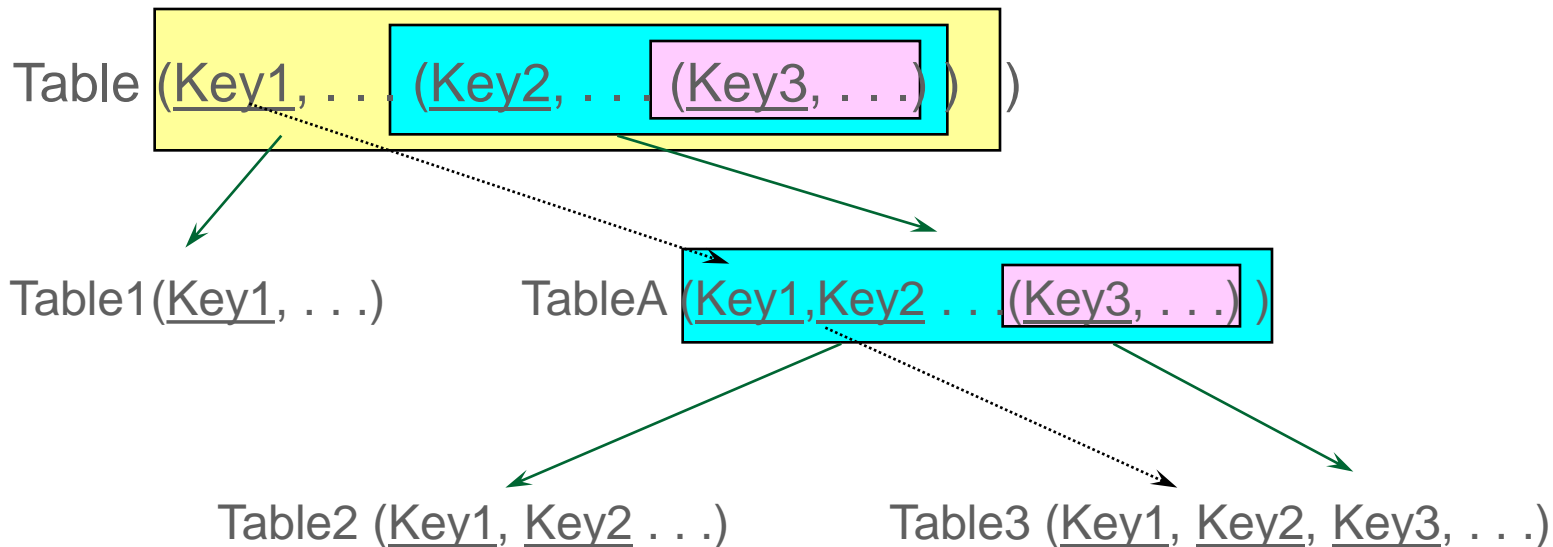
- Boyce Codd đề nghị 3 dạng
 - 1NF (first normal form): tương đương với định nghĩa của lược đồ quan hệ (quan hệ và bộ)
 - 2NF: ko có giá trị trong thực tiễn
 - 3NF → BCNF: thường sử dụng nhiều nhất
- 4NF, 5NF do tính đa trị và phụ thuộc hàm nối

Dạng chuẩn 1

- Đn: gọi là 1NF nếu miền giá trị của một thuộc tính chỉ chứa giá trị nguyên tử (đơn, ko 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



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 1 (tt)

- Vấn đề còn tồn tại trong 1NF

- Xét lược đồ

DDIEM_PHG(MaPHG, DDIEM)

- Vẫn bị lặp lại
- Ẩn chứa các phụ thuộc hàm bộ phận
-

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

Dạng chuẩn 2

- 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.

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

- 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)

- 2NF:
 - Thỏa mãn 1NF
 - Phụ thuộc hàm đầy đủ vào khóa chính
- Với các quan hệ có thuộc tính khóa đơn thì ko phải xét
- Chỉ kiểm tra các lược đồ có chứa phụ thuộc hàm bộ phận

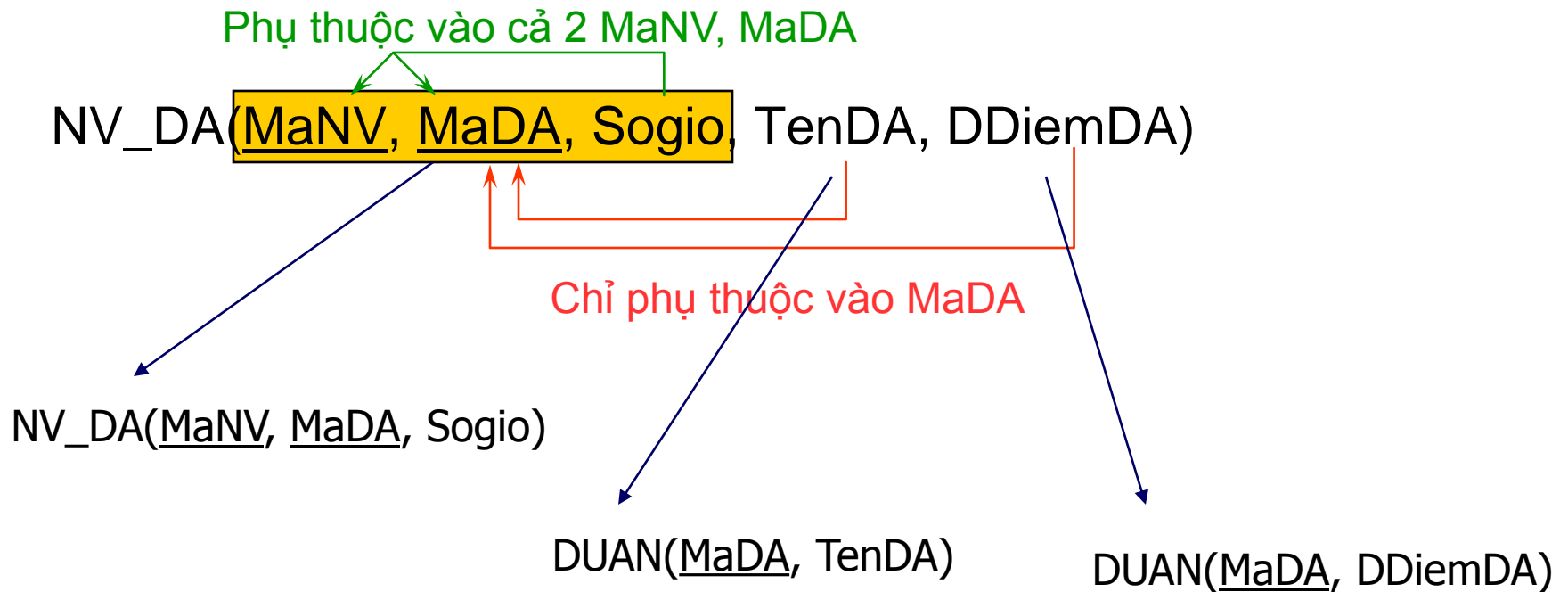
Dạng chuẩn 2 (tt)

- Ví dụ



Dạng chuẩn 2 (tt)

- Ví dụ

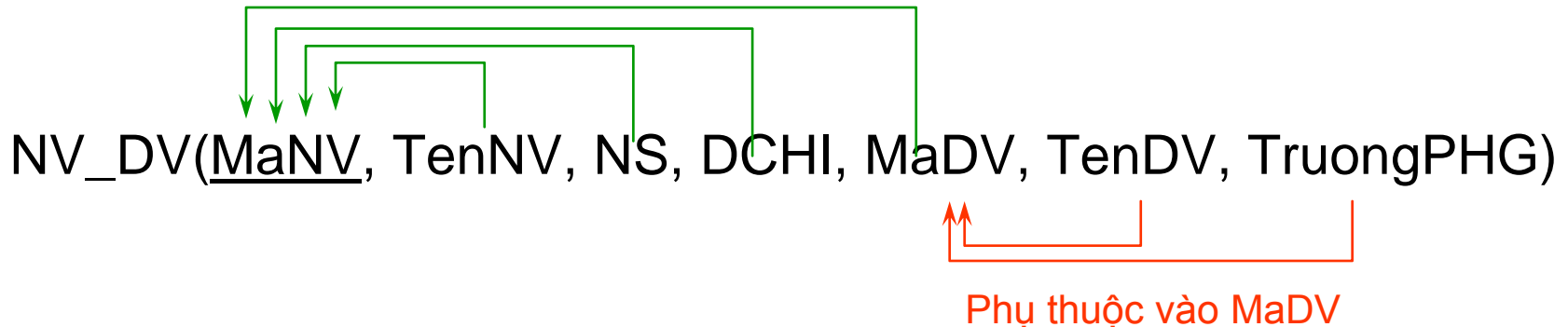


Dạng chuẩn 3

- 3NF dựa trên khái niệm phụ thuộc bắc cầu.
- ĐN: Một lược đồ quan hệ R là ở 3NF nếu nó thoả mãn (theo Codd)
 - Thỏa mãn 2NF
 - Không có thuộc tính không khoá nào của R là phụ thuộc bắc cầu vào khoá chính.

Dạng chuẩn 3 (tt)

Phụ thuộc vào MaNV



- Tất cả các thuộc tính phải phụ thuộc vào thuộc tính khóa
 - Một vài thuộc tính phụ thuộc vào thuộc tính ko phải là khóa
 - Chuẩn hóa → Tách nhóm các thuộc tính đó thành quan hệ mới

Dạng chuẩn 3 (tt)

Phụ thuộc vào MaNV

NV_DV(MaNV, TenNV, NS, DCHI, MaDV, TenDV, TruongPHG)

```
graph TD; MaNV --> TenNV; MaNV --> NS; MaNV --> DCHI; MaNV --> MaDV; MaDV --> TenDV; MaDV --> TruongPHG;
```

Phụ thuộc vào MaDV

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

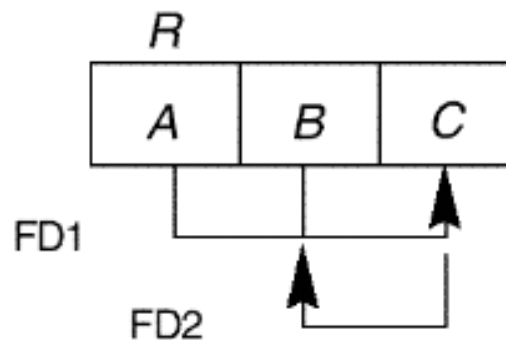
DONVI(MaDV, TenDV, TruongPHG)

Tóm tắt 3 dạng chuẩn 1-3

NF	Nhận biết	Cách chuẩn hóa
1	Quan hệ ko có thuộc tính đa trị và quan hệ lặp	Chuyển tất cả quan hệ lặp hoặc đa trị thành 1 quan hệ mới
2	Phụ thuộc 1 phần vào thuộc tính khóa	Tách thuộc tính phụ thuộc 1 phần thành lược đồ mới, đảm bảo quan hệ với lược đồ liên quan
3	Phụ thuộc ẩn, 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 chuẩn 3NF
 - Không có các thuộc tính khóa phụ thuộc hàm và thuộc tính không khóa.
- Ví dụ



Dạng chuẩn Boyce-Codd(tt)

- Ví dụ:

$R(\underline{A1}, \underline{A2}, A3, A4, A5)$

Với các phụ thuộc hàm:

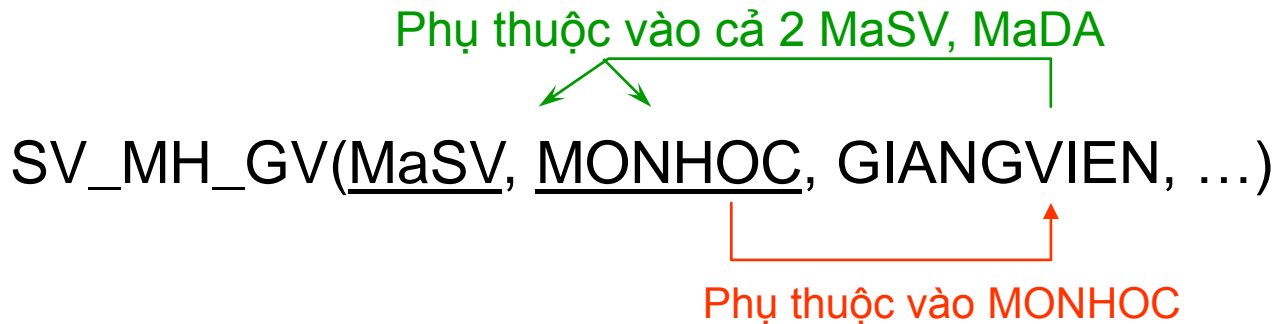
- $A1, A2 \rightarrow A3, A4, A5$
- $A4 \rightarrow A2$

Dạng chuẩn Boyce-Codd(tt)

- Nếu một lược đồ quan hệ không thoả 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)

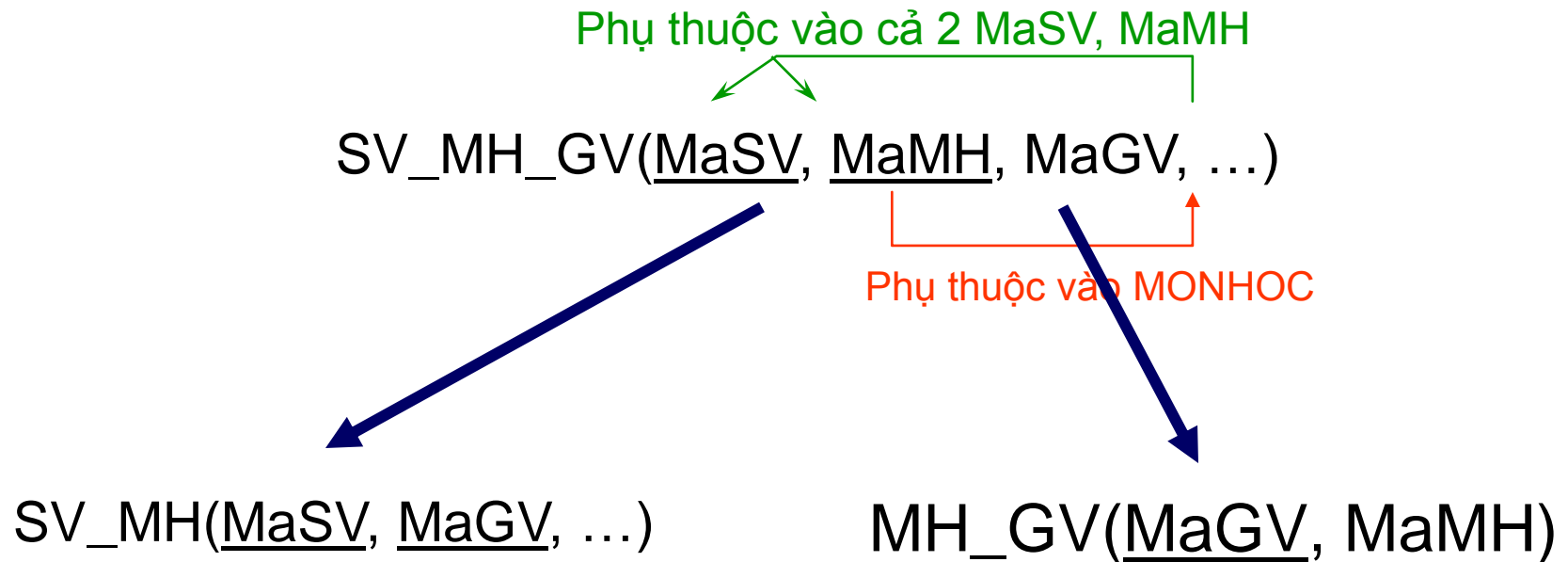
Dạng chuẩn Boyce-Codd(tt)

- Ví dụ



Dạng chuẩn Boyce-Codd(tt)

- Ví dụ



Tài liệu tham khảo

- ***Giáo trình CSDL***
 - Chương 4

- ***Database management system***
 - Chapter 15

- ***Fundamentals of Database Systems***
 - Chapter 14

