

# Chuẩn hóa CSDL quan hệ

- Một số vấn đề lý thuyết thiết kế để có lược đồ tốt.
- Một lược đồ tốt được thể hiện qua 2 mức:
  - Mức khái niệm (hay logic): ngữ nghĩa rõ ràng, dễ hiểu, chính xác...
  - Mức cài đặt: các bộ được lưu trữ như thế nào..
- 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.

## Nội dung chính

- Sự dư thừa và dị thường dữ liệu
- Phụ thuộc hàm
- Hệ suy diễn Amstrong
- Thuật toán tìm bao đóng  $X_F^+$
- Tìm phủ tối thiểu
- Các dạng chuẩn

4.1 - Sự dư thừa và dị thường dữ liệu

Ví dụ 1: NHANVIEN\_PHONG

Manv	Ho	Dem	Ten	Donvi	Maql
11001	Trần	Văn	An	Nghiên cứu	11001
11002	Lê	Đình	Bắc	Đào tạo	11002
11003	Trần	Thị	Hảo	Đào tạo	11002
11004	Vũ	Đức	Lâm	Hành chính	11005
11005	Phạm	Hải	Ngọc	Hành chính	11005
11006	Trần	Văn	Cường	Nghiên cứu	11001
11007	Vũ	Vân	Long	Đào tạo	11002

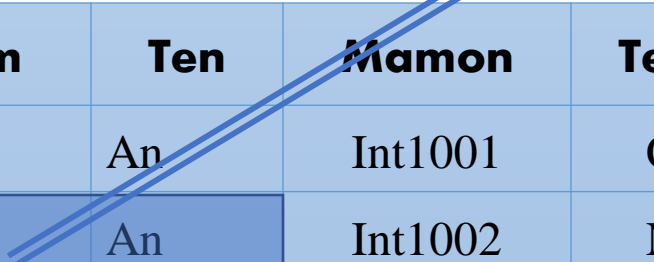


Dư thừa về Đơn vị / Người quản lý

4.1 - Sự dư thừa và dị thường dữ liệu

Ví dụ 2

Dư thừa



Masv	Ho	Dem	Ten	Mamon	Tenmon	Diem
T1	Trần	Văn	An	Int1001	CSDL	8
T1	Trần	Văn	An	Int1002	NNLT	9
C2	Lê	Đình	Bắc	Int1003	TRR	7
C2	Lê	Đình	Bắc	Int1002	NNLT	3
T3	Trần	Thị	Hảo	Int1003	TRR	10
T4	Vũ	Đức	Lâm	Int1002	NNLT	8
C2	Lê	Đình	Bắc	Int1001	CSDL	8
T4	Vũ	Đức	Lâm	Int1001	CSDL	7
C3	Phạm	Hải	Ngọc	Int1003	TRR	6

4.1 - Sự dư thừa và dị thường dữ liệu

Ví dụ 3

Masv	Ho	Dem	Ten	TRR	CSDL	NNLT	TB	Xeploai
T1	Trần	Văn	An	7	6	8	7.0	Khá
T2	Trần	Thị	Hảo	8	8	10	8.7	Giỏi
T3	Vũ	Đức	Lâm	5	9	8	7.3	Khá
T4	Phạm	Hải	Ngọc	6	5	6	5.7	Tbình



Dư thừa

## 4.1 - Sự dư thừa và dị thường dữ liệu

**Tại sao có sự dư thừa ??**

- Sự phụ thuộc lẫn nhau giữa các thuộc tính

*Ví dụ:*

Điểm các môn học → Điểm trung bình → xếp loại  
Mã phòng -> Tên phòng, người quản lý

- Thuộc tính đa trị trong lược đồ ER → nhiều bộ

*Ví dụ:*

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

4.1 - Sự dư thừa và dị thường dữ liệu



1. Lãng phí không gian nhớ

2. Dị thường cập nhật:

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

Manv	Ho	Dem	Ten	Donvi	Maql
11001	Trần	Văn	An	Nghiên cứu	11001
11002	Lê	Đình	Bắc	Đào tạo	11002
11003	Trần	Thị	Hảo	Đào tạo	11002
11004	Vũ	Đức	Lâm	Hành chính	11005
11005	Phạm	Hải	Ngọc	Hành chính	11005
11006	Trần	Văn	Cường	Nghiên cứu	11001
11007	Vũ	Vân	Long	Đào tạo	11002



## 4.2- Một số nguyên tắc thiết kế lược đồ

**Nguyên tắc 1:** Rõ ràng về ý nghĩa (quan hệ, thuộc tính), tránh các phụ thuộc (về ý nghĩa) giữa các thuộc tính với nhau

*Mỗi lược đồ quan hệ tương ứng với một kiểu thực thể hoặc một liên kết*

**Nguyên tắc 2:** Tránh các khả năng phát sinh dị thường cập nhật trong các quan hệ

*Tránh dư thừa, trùng lặp thông tin. Nếu có xuất hiện dị thường phải đảm bảo thao tác cập nhật thực hiện đúng đắn*

**Nguyên tắc 3:** Tránh đặt các thuộc tính có nhiều giá trị Null

**Nguyên tắc 4:** Các lược đồ quan hệ kết nối với điều kiện bằng trên các thuộc tính nên là khoá chính hoặc khoá ngoài theo cách đảm bảo không sinh ra các bộ “giả”

## 4.3 – Phụ thuộc hàm

### ***Định nghĩa:***

Cho lược đồ quan hệ  $R$ ;  $X, Y$  là các tập thuộc tính trên  $R$ .

Một phụ thuộc hàm giữa  $X$  và  $Y$  được kí hiệu  $X \rightarrow Y$  là một ràng buộc:

Với mỗi thể hiện  $r$  của lược đồ quan hệ  $R$ , với 2 bộ bất kỳ  $t_1$  và  $t_2$  trong  $r$   
nếu có  $t_1[X] = t_2[X]$  thì  $t_1[Y] = t_2[Y]$

*(tức là 2 bộ bất kỳ bằng nhau trên  $X$  thì cũng bằng nhau trên  $Y$ )*

*Ta nói  $Y$  phụ thuộc hàm vào  $X$  hay  $X$  xác định hàm  $Y$ ;  $X$  gọi là vế trái,  
 $Y$  là vế phải của phụ thuộc hàm*

Phụ thuộc hàm là tính chất ngữ nghĩa trên các thuộc tính của lược đồ, được xác định khi thiết kế chứ không suy đoán trên một thể hiện của lược đồ

## 4.3 – Phụ thuộc hàm

*Ví dụ:*

**SINHVIEN**(Masv, Ho, Dem,Ten, Ngaysinh, Noisinh, Lop)

Phụ thuộc hàm: **Masv → Ho, Dem,Ten, Ngaysinh, Noisinh, Lop**

**SINHVIEN\_DIEM**(Masv,Mamon, Ngaythi, Diem)

Phụ thuộc hàm **Masv,Mamon → Diem**

**MUON**(Sothe, Masach, Tennguoimuon, Tensach, Ngaymuon, Ngaytra)

Với các phụ thuộc hàm:

**Sothe → Tennguoimuon**

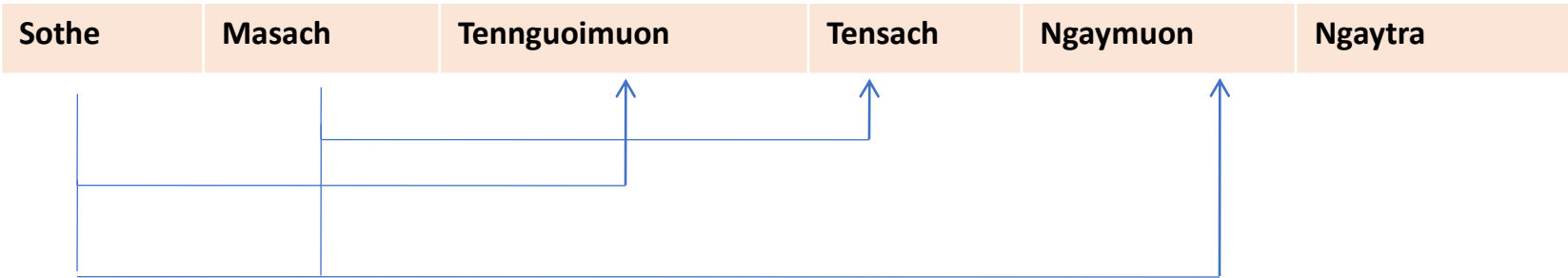
**Masach → Tensach**

**Sothe,Masach, Ngaymuon → Ngaytra**

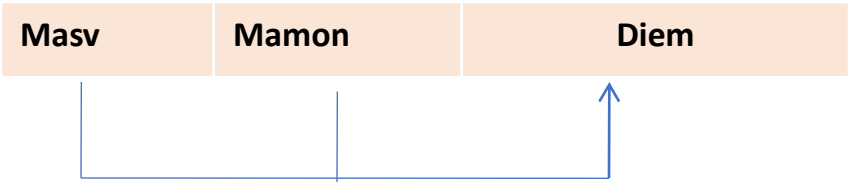
### 4.3 – Phụ thuộc hàm

*Thể hiện phụ thuộc hàm trên lược đồ*

MUON



SINHVIEN\_DIEM



#### 4.4 – Quy tắc suy diễn Phụ thuộc hàm

Cho lược đồ  $R(A)$ ,  $\mathcal{F}$  là tập các phụ thuộc hàm;

$X \rightarrow Y$  gọi là suy diễn được từ  $\mathcal{F}$ , nếu với mọi thể hiện  $r$  của  $R$  thỏa mãn các phụ thuộc hàm  $\mathcal{F}$  thì  $X \rightarrow Y$  cũng đúng trong  $r$

Kí hiệu  $\mathcal{F} \mid = X \rightarrow Y$

Ví dụ: MUON(Sothe, Masach, Tennguoimuon, Tensach, Ngaymuon, Ngaytra)

$\mathcal{F} = \{ \text{Sothe} \rightarrow \text{Tennguoimuon}; \text{Masach} \rightarrow \text{Tensach};$   
 $\text{Sothe, Masach} \rightarrow \text{Ngaymuon}, \text{Ngaytra} \}$

$\mathcal{F} \mid = \text{Sothe, masach} \rightarrow \text{Tensach}, \text{Ngaytra}$

## 4.4 – Quy tắc suy diễn Phụ thuộc hàm

*Quy tắc 1* (quy tắc phản xạ) : Nếu  $X \supset Y$  thì  $X \rightarrow Y$

*Quy tắc 2* (quy tắc tăng) :  $\{ X \rightarrow Y \} \models XZ \rightarrow YZ$

*Quy tắc 3* (quy tắc bắc cầu):  $\{ X \rightarrow Y, Y \rightarrow Z \} \models X \rightarrow Z$

*Quy tắc 4* (quy tắc chiếu) :  $\{ X \rightarrow YZ \} \models \{ X \rightarrow Y, X \rightarrow Z \}$

*Quy tắc 5* (quy tắc hợp) :  $\{ X \rightarrow Y, X \rightarrow Z \} \models X \rightarrow YZ$

*Quy tắc 6* (quy tắc tựa bắc cầu):  $\{ X \rightarrow Y, WY \rightarrow Z \} \models WX \rightarrow Z$

**Quy tắc suy diễn Amstrong:** Tập các quy tắc 1 -3

#### 4.10— Các dạng chuẩn dựa trên khóa chính

- ***Chuẩn là gì?*** 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 dị thường trong cập nhật)
- ***Chuẩn hóa*** : quá trình phân tích lược đồ quan hệ dựa trên các FD và các khóa chính để đạt được:
  - Giảm tối đa sự dư thừa
  - Giảm tối đa các phép cập nhật dị thường

*Chuẩn và hóa do Codd đề xuất đầu tiên năm 1972 (1NF-3NF, sau đó Boyce-Codd, 4-5NF)*

#### 4.10– Các dạng chuẩn dựa trên khóa chính

##### *Thủ tục chuẩn hoá cung cấp*

- Một cơ cấu hình thức để phân tích các lược đồ quan hệ dựa trên *khoá* và *các phụ thuộc hàm*.
- 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.

##### *Chuẩn hóa cần đảm bảo tính chất:*

- Nối không mất mát (hoặc nối không phụ thêm- không thêm 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.



## a. Dạng chuẩn 1 (1NF)

*Một quan hệ gọi là 1NF nếu*

- Miền giá trị của mỗi thuộc tính chỉ chứa giá trị nguyên tử (đơn, ko phân chia được)
- Giá trị của mỗi thuộc tính trong các bộ là một giá trị đơn

*Ví dụ:*

SV\_DIEM(Masv, Mamon, Diem)

SV(Masv, **Hoten**, Gioitinh, Ngaysinh, Noisinh)

←  
Không thỏa mãn 1NF

a. Dạng chuẩn 1

Ví dụ:

NV\_DA(Mada,Tenda,Mavn,Sogio)

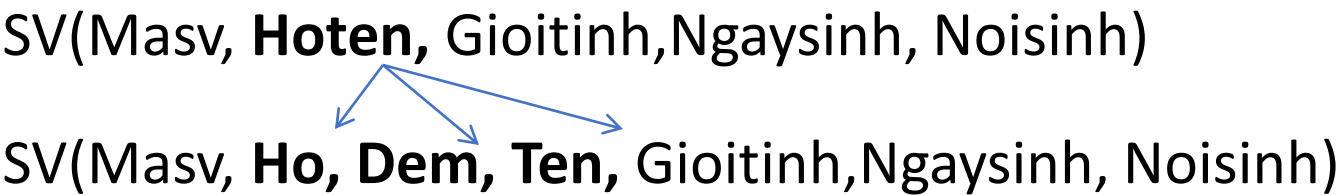
Mada	TenDa	Manv	Sogio
CO1	Cấp nước	001	20
		002	35
DO2	Cung cấp thiết bị điện..	002	20
		004	40

Không thỏa mãn 1NF

a. Dạng chuẩn 1

- Chuyển quan hệ không đạt chuẩn về dạng chuẩn 1

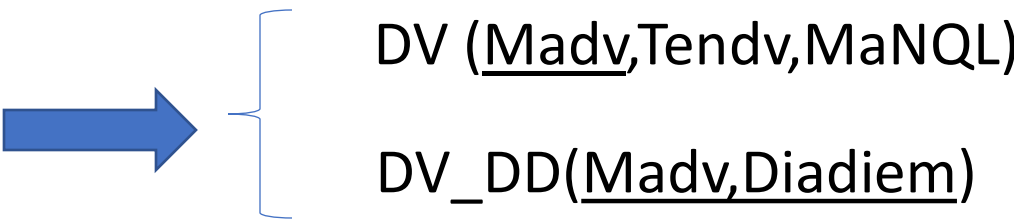
1. Thuộc tính phức hợp -> các thuộc tính đơn



2. Thuộc tính đa trị hoặc lặp ->tách quan hệ

DONVI(Madv,Tendv, MaNQL, **Diadiem**)

← Đa trị



a. Dạng chuẩn 1

NV\_DA(Mada,Tenda,Mavn,Sogio)

Lặp

Mada	TenDa	Manv	Sogio
CO1	Cấp nước	001	20
		002	35
DO2	Cung cấp thiết bị điện..	002	20
		004	40



DA(Mada,Tenda)

Mada	Tenda
CO1	Cấp nước
Do2	Cung cấp thiết bị điện

NV\_DA(Mada,Mavn,Sogio)

Mada	Manv	Sogio
CO1	001	20
CO1	002	35
DO2	002	20
DO2	004	40

## b. 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.

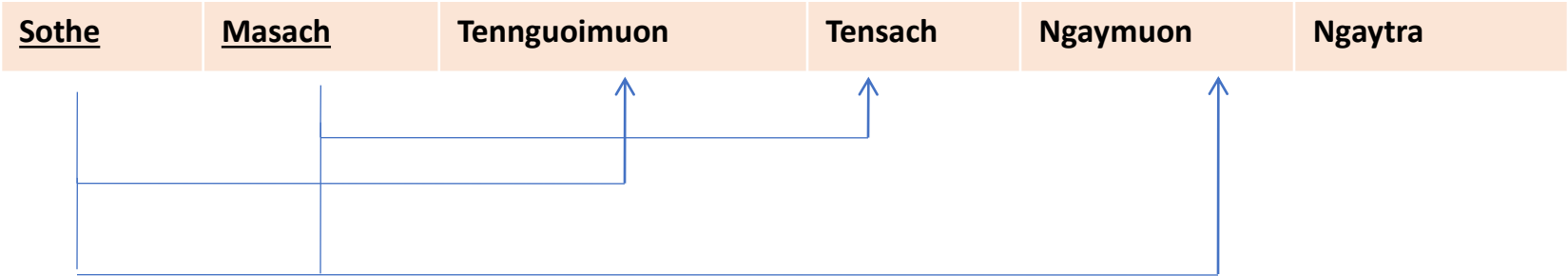
$\forall A, A \in X, (X - \{A\}) \rightarrow Y$  : là không đúng.

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

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

b. Dạng chuẩn 2

MUONTRA



- Sothe,Masach -> Ngaymuon                      ➡      *Phụ thuộc đầy đủ*
- Sothe,Masach -> Tensach                      ➡      *Phụ thuộc bộ phận*
- Sothe,Masach -> Nguoimuon                      ➡      *Phụ thuộc bộ phận*

## b. Dạng chuẩn 2

**Định nghĩa:** Một lược đồ quan hệ R ở dạng chuẩn 2 (2NF) nếu:

- R thỏa mãn chuẩn 1
- Mọi thuộc tính (*không khóa*) của R phụ thuộc hàm đầy đủ vào khóa chính

**Hay:** Mỗi thuộc tính không là thuộc tính khóa không phụ thuộc bộ phận vào khóa của R



Kiểm tra lược đồ thỏa mãn dạng chuẩn 2 ?

- Với các quan hệ có thuộc tính khóa đơn thì ko phải kt
- Chỉ kiểm tra các lược đồ có chứa phụ thuộc hàm bộ phận

b. Dạng chuẩn 2

Chuẩn hóa về dạng chuẩn 2

MUONTRA

<u>Sothe</u>	<u>Masach</u>	Tennguoimuon	Tensach	Ngaymuon	Ngaytra

Sothe,Masach→Tennguoimuon

Sothe,Masach→Tensach

Sothe,Masach→Ngaymuon

Sothe,Masach→Ngaytra

Sothe→Tennguoimuon

Masach→Tensach

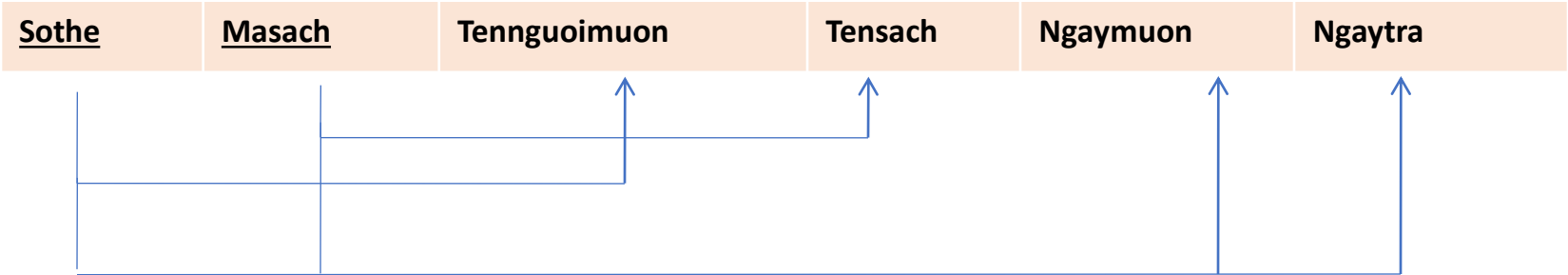
Phụ thuộc bộ phận vào khóa



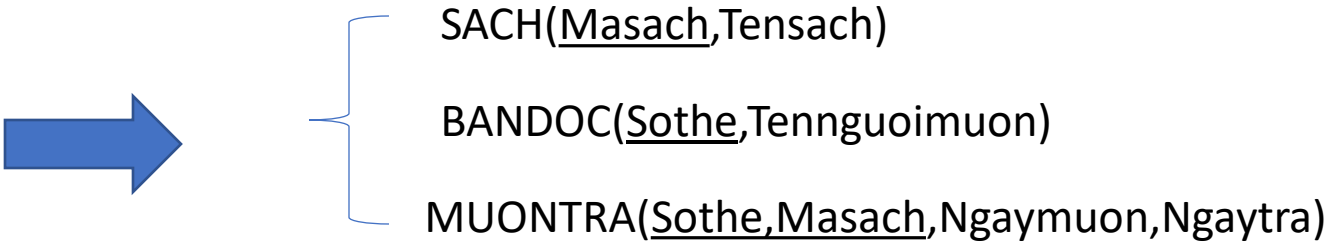
b. Dạng chuẩn 2

Chuẩn hóa về dạng chuẩn 2

MUONTRA



Tách các thuộc tính không khóa phụ thuộc bộ phận vào khóa chính thành quan hệ riêng; khóa của quan hệ mới là khóa bộ phận tương ứng



## b. Dạng chuẩn 2

*Ví dụ1: Chuẩn hóa quan hệ R thành dạng chuẩn 2*

**$R(\underline{A}, \underline{B}, C, D, E)$**

**$F = \{ AB \twoheadrightarrow C, AB \twoheadrightarrow D, AB \twoheadrightarrow E, B \rightarrow C, A \rightarrow E \}$**



**$R1(\underline{A}, E)$**

**$R2(\underline{B}, C)$**

**$R(\underline{A}, \underline{B}, D)$**

## b. Dạng chuẩn 2

*Ví dụ 2: Chuẩn hóa quan hệ R thành dạng chuẩn 2*

**R(A,B,C,D,E,F,G,H)**

**F = {  $AB \rightarrow C$ ,  $AB \rightarrow D$ ,  $AB \rightarrow E$ ,  $AB \rightarrow F$ ,  $AB \rightarrow G$ ,  $AB \rightarrow H$**

**$B \rightarrow C$ ,  $A \rightarrow E$ ,  $B \rightarrow G$ }**



**R1 (A,E)**

**R2(B,C, G)**

**R(A,B,D,F,H)**

## b. Dạng chuẩn 2

*Bài tập: Cho quan hệ*

**$R(\underline{A}, \underline{B}, C, D, E, F, G, H, I, J)$  với tập phụ thuộc hàm**

**$F = \{ AB \twoheadrightarrow C, A \twoheadrightarrow DE, B \twoheadrightarrow F, F \twoheadrightarrow GH, D \twoheadrightarrow IJ \}$**

*Khóa của quan hệ R ?      AB*

*Chuyển về dạng chuẩn 2?*

**$R_1(\underline{A}, \underline{D}, \underline{E}, \underline{I}, \underline{J}) \quad F_1 = \{ A \rightarrow D, A \rightarrow E, D \rightarrow I, D \rightarrow J \}$**

**$R_2(\underline{B}, \underline{F}, \underline{G}, \underline{H}) \quad F_2 = \{ B \rightarrow F, F \rightarrow G, F \rightarrow H \}$**

**$R(A, B, C) \quad F = \{ AB \twoheadrightarrow C \}$**

### c. Dạng chuẩn 3

#### ■ *Phụ thuộc bắc cầu:*

Phụ thuộc hàm  $\mathbf{X \rightarrow Y}$  được gọi bắc cầu nếu trong R có  $\mathbf{X \rightarrow Z}$  và  $\mathbf{Z \rightarrow Y}$ ; với Z là tập thuộc tính không thuộc khóa.

*Ta nói Y phụ thuộc bắc cầu vào X*

Ví dụ:

$$\mathbf{F = \{AB \rightarrow C, AB \rightarrow D, D \rightarrow F, E \rightarrow F, D \rightarrow E\}}$$

*Phụ thuộc hàm bắc cầu:*       $\mathbf{D \rightarrow F}$

c. **Dạng chuẩn 3**

*Lược đồ **R** là dạng chuẩn 3 nếu:*

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

*Tức là: mỗi phụ thuộc hàm **X → Y** thì*

- Hoặc X siêu khóa
- Hoặc Y là thuộc tính khóa.

Ví dụ: **R(A,B,C,D,E,F)**

với **F1 = {AB → C, AB → D, AB → E, AB → F, E → B}**

**S(A,B,C,D,E,F)**

với **F2 = {AB → C, AB → D, E → D}**

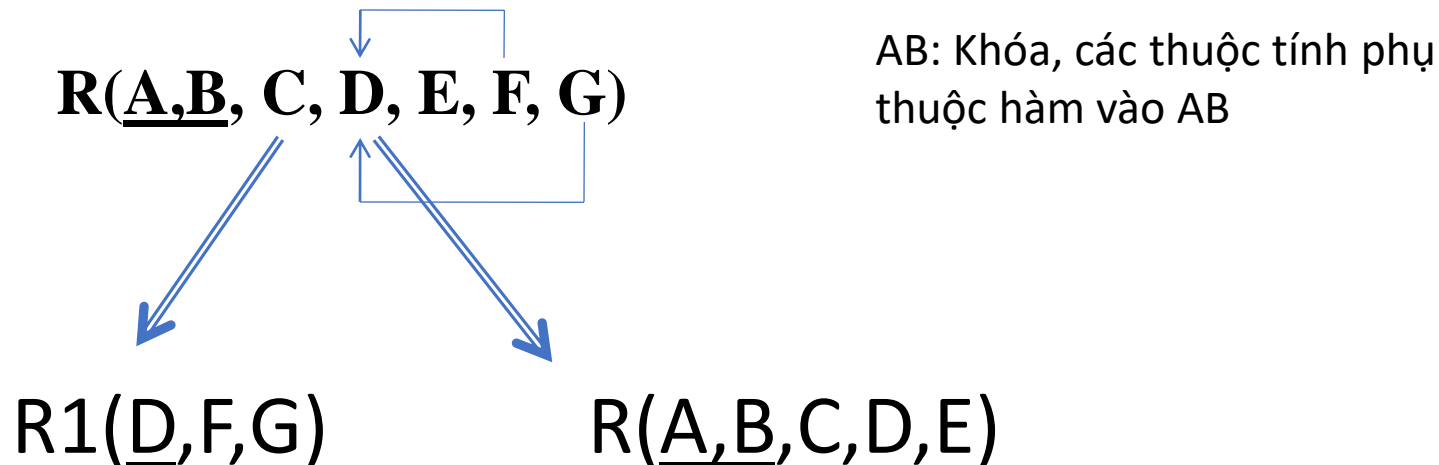
3NF

không thỏa 3NF

### c. Dạng chuẩn 3

*Chuẩn hóa lược đồ  $R$  :*

- Tách quan hệ mới gồm các thuộc tính phụ thuộc bậc cầu.
- Loại các thuộc tính phụ thuộc bậc cầu vào thuộc tính khóa trong quan hệ ban đầu;



### c. Dạng chuẩn 3

*Ví dụ* **NV\_DV**(Manv, Hoten, Ngaysinh, Madv, Tendv, MaQl)

*Với các phụ thuộc hàm*

{ **Manv**→Hoten, **Manv**→Ngaysinh, **Manv**→Madv,  
**Manv**→MaQl, **Manv**→Tendv,  
**MaDv**→Tendv, **MaDv**→MaQl }

*Các thuộc tính phụ thuộc hàm* **Tendv, MaQl** phụ thuộc bắc cầu vào khóa chính

**DV**(Madv, Tendv, MaQl)

**NV**(Manv, Hoten, Ngaysinh, Madv)



Chuẩn hóa(1-3)

NF	Nhận biết (chưa đạt chuẩn)	Cách chuẩn hóa
1	Quan hệ có thuộc tính đa trị /(quan hệ) lặp	Tách tất cả thuộc tính lặp hoặc đa trị thành 1 quan hệ mới
2	Có thuộc tính phụ thuộc 1 phần (bộ phận) 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 bắc cầu, tồn tại phụ thuộc hàm giữa các thuộc tính ko phải là khóa	Tách các thuộc tính đó thành lược đồ mới

### d. **Dạng chuẩn Boyce-Codd (BCNF)**

Một lược đồ quan hệ R được gọi là ở dạng chuẩn **Boyce-Codd (BCNF)** nếu:

- Thỏa mãn dạng chuẩn 3NF
- Không có thuộc tính khóa phụ thuộc hàm vào thuộc tính không khóa.

Ví dụ



**NV(Manv, Hoten, Ngaysinh, Madv)**

Với pth: {**Manv**→Hoten, **Manv**→Ngaysinh, **Manv**→Madv}

d. **Dạng chuẩn Boyce-Codd (BCNF)**

*Chuẩn hóa lược đồ về dạng BCNF*

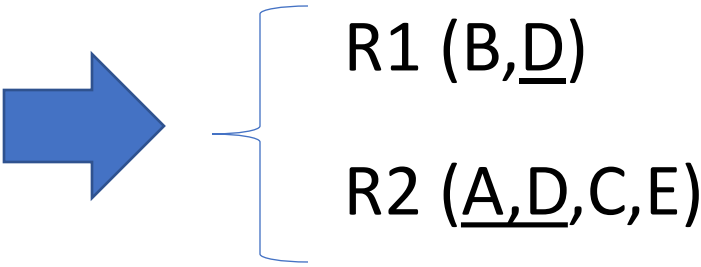
Ví dụ:

Cho R (A,B,C,D,E)

Với các phụ thuộc hàm:

$F = \{AB \rightarrow C, AB \rightarrow D, AB \rightarrow E, \mathbf{D \rightarrow B}\}$

*không thỏa mãn BCNF*



## d. **Dạng chuẩn Boyce-Codd (BCNF)**

*Chuẩn hóa lược đồ về dạng BCNF*

- b1: Tách các thuộc tính không khóa và thuộc tính khóa phụ thuộc hàm vào thuộc tính không khóa thành quan hệ mới, thuộc tính không khóa trở thành khóa trong quan hệ mới.
- b2: Loại các thuộc tính khóa ở bước 1 khỏi lược đồ gốc
- b3: Bổ sung thuộc tính không khóa xác định hàm thuộc tính khóa đã loại bỏ (bước 2) vào khóa của quan hệ gốc