

# **Chương 8**

## **Phụ thuộc hàm và Dạng chuẩn trong CSDL quan hệ**

1

### **Nội dung trình bày**

1. Thiết kế CSDL là gì?
2. Một số nhận xét khi thiết kế CSDL.
3. Phụ thuộc hàm.
4. Các dạng chuẩn.

## Thiết kế CSDL

1. Định nghĩa: là tiến trình xây dựng nên mô hình dữ liệu cho một CSDL.
2. Quá trình thiết kế CSDL diễn ra ở nhiều cấp độ:
  - Thiết kế CSDL mức quan niệm (Conceptual database design): cho ra lược đồ CSDL ở mức quan niệm, tức là độc lập với DBMS.
    - Ta thường dùng mô hình dữ liệu cấp cao để thể hiện dữ liệu ở mức này. Ví dụ: ER.
  - Thiết kế CSDL mức logic (Logical database design): chuyển lược đồ CSDL mức quan niệm sang mô hình dữ liệu của một DBMS cụ thể.
  - Thiết kế CSDL mức vật lý (Physical database design): chỉ ra cách thức lưu trữ vật lý cho dữ liệu.


## Thiết kế CSDL

3. Mục đích của quá trình thiết kế CSDL:
  - Đáp ứng các yêu cầu thông tin cho các người dùng và các ứng dụng cụ thể.
  - Cung cấp kiến trúc lưu trữ dữ liệu đúng đắn, tự nhiên và dễ hiểu.
  - Hỗ trợ một số yêu cầu về hiệu năng: thời gian hồi đáp, thời gian xử lý, không gian lưu trữ, ...

## Ngữ nghĩa của các thuộc tính trong một quan hệ

- Khi gom nhóm các thuộc tính vào 1 lược đồ quan hệ thì ta ngụ ý rằng chúng cùng tạo nên một ngữ nghĩa thực tế và hợp lý. Vì vậy:
  - Thiết kế 1 lược đồ quan hệ sao cho dễ mô tả ngữ nghĩa. Đó là lược đồ gồm các thuộc tính của 1 loại thực thể hoặc 1 loại mối kết hợp.
  - Không kết hợp các thuộc tính của nhiều loại thực thể hoặc loại mối kết hợp trong một quan hệ, vì sẽ tạo ra quan hệ mang ý nghĩa mập mờ và khó hiểu.

## Dư thừa thông tin trên các bộ dữ liệu



MANV	TENNV	NGSINH	DCHI	PHG	TENPHG	TRPHG
009	Tiến	11/02/1960	119 Công Quỳnh, Tp HCM	5	Nghiên cứu	005
005	Tùng	20/08/1962	222 Nguyễn Văn Cừ, Tp HCM	5	Nghiên cứu	005
007	Hàng	11/3/1954	332 Nguyễn Thái Học, Tp HCM	4	Điều hành	008
001	Như	01/02/1967	291 Hồ Văn Huê, Tp HCM	4	Điều hành	008
004	Hùng	04/03/1967	95 Bà Rịa, Vũng Tàu	5	Nghiên cứu	005
003	Tâm	04/05/1957	34 Mai Thị Lữ, Tp HCM	5	Nghiên cứu	005
008	Quang	01/09/1967	80 Lê Hồng Phong, Tp HCM	4	Điều hành	008
006	Vinh	01/01/1965	45 Trưng Vương, Hà Nội	1	Quản lý	006

- Thông tin về tên phòng và người làm trưởng phòng lặp lại nhiều lần.

## Dị thường khi thêm bộ dữ liệu

MANV	TENNV	NGSINH	DCHI	PHG	TENPHG	TRPHG
009	Tiền	11/02/1960	119 Công Quỳnh, Tp HCM	5	Nghiên cứu	005
005	Tùng	20/08/1962	222 Nguyễn Văn Cừ, Tp HCM	5	Nghiên cứu	005
007	Hàng	11/3/1954	332 Nguyễn Thái Học, Tp HCM	4	Điều hành	008
001	Như	01/02/1967	291 Hồ Văn Huê, Tp HCM	4	Điều hành	008
004	Hùng	04/03/1967	95 Bà Rịa, Vũng Tàu	5	Nghiên cứu	005
003	Tâm	04/05/1957	34 Mai Thị Lư, Tp HCM	5	Nghiên cứu	005
008	Quang	01/09/1967	80 Lê Hồng Phong, Tp HCM	4	Điều hành	008
006	Vinh	01/01/1965	45 Trưng Vương, Hà Nội	1	Quản lý	006

- Dị thường khi thêm:
  - Thêm 1 nhân viên thì phải thêm thông tin phòng ban hoặc để các trường liên quan đến phòng ban là null. Việc thêm thông tin phòng lặp lại nhiều lần phải đảm bảo nhất quán.
  - Không thể thêm 1 phòng ban mới mà chưa có nhân viên nào thuộc phòng đó.

## Dị thường khi xóa dữ liệu

MANV	TENNV	NGSINH	DCHI	PHG	TENPHG	TRPHG
009	Tiền	11/02/1960	119 Công Quỳnh, Tp HCM	5	Nghiên cứu	005
005	Tùng	20/08/1962	222 Nguyễn Văn Cừ, Tp HCM	5	Nghiên cứu	005
007	Hàng	11/3/1954	332 Nguyễn Thái Học, Tp HCM	4	Điều hành	008
001	Như	01/02/1967	291 Hồ Văn Huê, Tp HCM	4	Điều hành	008
004	Hùng	04/03/1967	95 Bà Rịa, Vũng Tàu	5	Nghiên cứu	005
003	Tâm	04/05/1957	34 Mai Thị Lư, Tp HCM	5	Nghiên cứu	005
008	Quang	01/09/1967	80 Lê Hồng Phong, Tp HCM	4	Điều hành	008
006	Vinh	01/01/1965	45 Trưng Vương, Hà Nội	1	Quản lý	006

- Dị thường khi xóa: Khi xóa một nhân viên cuối cùng trong 1 phòng ban thì thông tin phòng ban kéo theo cũng bị xóa.

## Dị thường khi cập nhật

MANV	TENNV	NGSINH	DCHI	PHG	TENPHG	TRPHG
009	Tiền	11/02/1960	119 Công Quỳnh, Tp HCM	5	Nghiên cứu	005
005	Tùng	20/08/1962	222 Nguyễn Văn Cừ, Tp HCM	5	Nghiên cứu	005
007	Hằng	11/3/1954	332 Nguyễn Thái Học, Tp HCM	4	Điều hành	008
001	Như	01/02/1967	291 Hồ Văn Huê, Tp HCM	4	Điều hành	008
004	Hùng	04/03/1967	95 Bà Rịa, Vũng Tàu	5	Nghiên cứu	005
003	Tâm	04/05/1957	34 Mai Thị Lữ, Tp HCM	5	Nghiên cứu	005
008	Quang	01/09/1967	80 Lê Hồng Phong, Tp HCM	4	Điều hành	008
006	Vinh	01/01/1965	45 Trưng Vương, Hà Nội	1	Quản lý	006

- Dị thường khi cập nhật: Khi thay đổi người quản lý của 1 phòng ban thì phải cập nhật đồng loạt, nếu không sẽ xảy ra tình trạng không nhất quán.

## Về giá trị null

- Đối với một bộ cụ thể, giá trị null tại một cột có nhiều ý nghĩa:
  - Không thể áp dụng thuộc tính này đối với bộ đang xét.
  - Giá trị đó là không biết, hoặc chưa biết.
- Một quan hệ có nhiều giá trị null:
  - Lãng phí không gian lưu trữ.
  - Không có lợi về mặt ngữ nghĩa trong thực tế.
- *Vì vậy, trong quá trình thiết kế, tránh tối đa việc cho ra quan hệ thường xuyên xuất hiện giá trị null.*

## Phụ thuộc hàm (PTH)

- Cho  $R(A_1, A_2, \dots, A_n)$ ,  $r(R)$ , ký hiệu  $R^+ = \{A_1, A_2, \dots, A_n\}$
- Phụ thuộc hàm giữa 2 tập thuộc tính  $X, Y \subseteq R^+$ .
  - Ký hiệu :  $X \rightarrow Y$ ,  $X$  là vế trái,  $Y$  là vế phải.
  - Ý nghĩa:  $\forall r \in R, t_1, t_2 \in r$ , nếu  $t_1[X] = t_2[X]$  thì  $t_1[Y] = t_2[Y]$ .
  - Ví dụ:

NV\_PB(MANV, TENNV, NGSINH, DCHI, MAPB, TENPB, TRPHG)

Có các PTH:

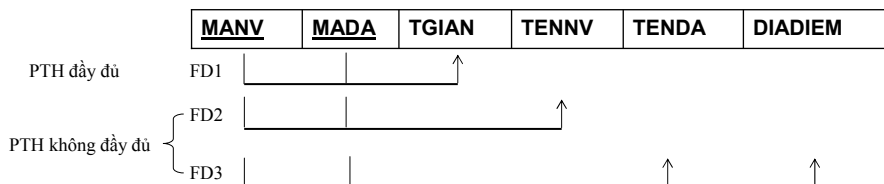
$MANV \rightarrow TENNV$ ,  
 $MANV \rightarrow MAPB$ ,  
 $MAPB \rightarrow \{TENPB, TRPHG\}$

## Phụ thuộc hàm

- Nhận xét:
  - PTH thể hiện các ràng buộc từ thế giới thực.
  - $\forall r \in R, \forall t \in r$ , nếu  $t[X]$  là duy nhất thì  $X$  là khóa của  $R$ .
  - Nếu  $K$  là khóa của  $R$  thì  $K$  xác định hàm tất cả các tập thuộc tính con của  $R^+$ .
  - PTH dùng để đánh giá một kết quả thiết kế CSDL.

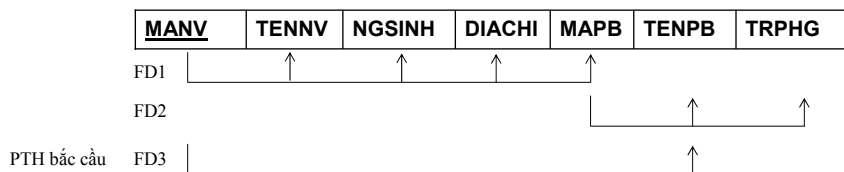
## Phụ thuộc đầy đủ

- $X \rightarrow Y$  là PTH đầy đủ nếu  $\forall A \in X$  thì  $(X - \{A\}) \rightarrow Y$  không đúng trên R.



## Phụ thuộc bắc cầu

- Cho R,  $X \rightarrow Y$  là PTH bắc cầu nếu  $\exists Z \subseteq R^+$ ,  $X \rightarrow Z$  và  $Z \rightarrow Y$  đúng trên R.
- Ví dụ:



- $MANV \rightarrow TENPB$  là PTH bắc cầu vì  $\exists MAPB \subseteq R^+$ ,  $MANV \rightarrow MAPB$  và  $MAPB \rightarrow TENPB$

## Thuộc tính khóa

- Một quan hệ có thể có nhiều khóa. Mỗi khóa được gọi là khóa ứng viên.
- Ta chọn ra trong các khóa ứng viên một khóa để làm khóa chính cho R.
- Thuộc tính khóa: Một thuộc tính được gọi là thuộc tính khóa của R nếu đó là một thuộc tính của một khóa ứng viên nào đó của R.
- Thuộc tính không khóa: Một thuộc tính được gọi là thuộc tính không khóa của R nếu nó không phải là thành viên của một khóa ứng viên nào của R.

## Một số vấn đề liên quan PTH

- Bao đóng của tập thuộc tính.
- Bao đóng của tập PTH  $F$  gồm  $F$  và các PTH được suy diễn từ  $F$ .
- Luật suy diễn: dùng để suy diễn ra PTH mới từ một tập PTH cho trước.
- Kiểm tra 1 PTH có phải được suy diễn từ tập PTH cho trước hay không.
- Tập PTH tương đương.
- Tập PTH tối tiểu.
- Xác định khóa của 1 lược đồ quan hệ.



## Dạng chuẩn

- Là tiêu chuẩn để đánh giá một lược đồ quan hệ lưu trữ dữ liệu có tốt hay không.
  - Có các dạng chuẩn sau:
    - Dạng chuẩn 1.
    - Dạng chuẩn 2.
    - Dạng chuẩn 3.
    - Dạng chuẩn BCK.
    - Dạng chuẩn 4.
    - Dạng chuẩn 5.
- Dựa trên PTH (functional dependencies)
- Dựa trên phụ thuộc đa trị (multivalued dependencies)
- Dựa trên chuỗi kết hàm (join dependencies)
- Ta quan tâm 4 dạng chuẩn dựa trên PTH.

## Dạng chuẩn 1

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

<u>MAPB</u>	TENPB	TRPHG	CHINHANH
5	Nghien cuu	NV01	Tan Binh Thu Duc
6	Hanh chinh	NV02	Go Vap Q3

- Lược đồ không đạt DC1, do CHINHANH không là thuộc tính đơn trị.

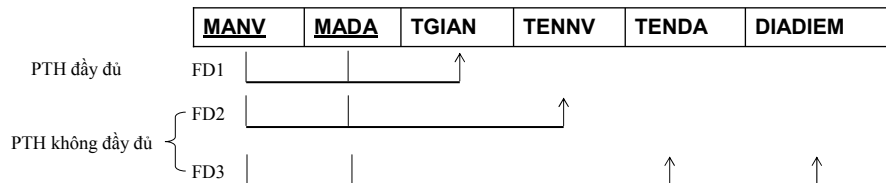
## Dạng chuẩn 1

– Nhận xét:

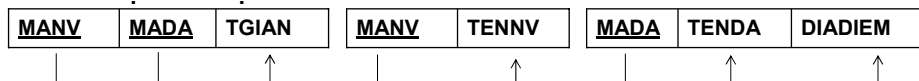
- Mọi lược đồ quan hệ đều thuộc dạng chuẩn 1.
- Dạng chuẩn 1 vẫn còn tình trạng trùng lặp dữ liệu, nên gây ra các bất thường về cập nhật dữ liệu.

## Dạng chuẩn 2

– Lược đồ quan hệ được gọi là đạt dạng chuẩn 2 nếu mọi thuộc tính không khóa của R phụ thuộc đầy đủ vào khóa chính của R.



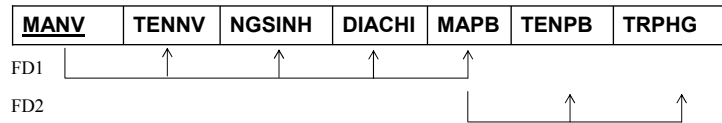
- Lược đồ quan hệ đã cho không đạt DC2. Sau đây là các lược đồ đạt DC2.



## Dạng chuẩn 2

– Nhận xét:

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



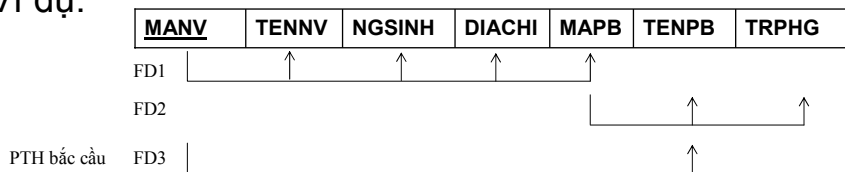
- Ví dụ: Lược đồ trên đạt DC2 nhưng vẫn còn trùng lặp dữ liệu.

### Dạng chuẩn 3

- Lực đồ quan hệ R được gọi là đạt dạng chuẩn 3 nếu:

- R đạt dạng chuẩn 2.
- Mọi thuộc tính không khóa của R không phụ thuộc bắc cầu vào khóa chính của R.

- Ví dụ:



- Lược đồ quan hệ trên không đạt DC3 vì  $MANV \rightarrow TENPB$  là PTH bắc cầu (vì  $\exists MAPB \subseteq R^+$ ,  $MANV \rightarrow MAPB$  và  $MAPB \rightarrow TENPB$ ).

## Dạng chuẩn 3



– Hai lược đồ quan hệ trên đạt DC3.

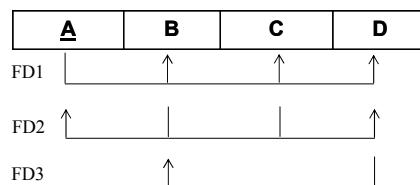
– Nhận xét:

- Mọi lược đồ quan hệ thuộc dạng chuẩn 3 cũng thuộc dạng chuẩn 2.
- Phụ thuộc hàm bậc cầu là nguyên nhân dẫn đến trùng lặp dữ liệu.
- DC3 là dạng chuẩn tối thiểu trong thiết kế CSDL.

## Dạng chuẩn Boyce-Codd-Kent (BCK)

– Lược đồ quan hệ R được gọi là thuộc dạng chuẩn BCK nếu PTH không hiển nhiên  $X \rightarrow Y$  đúng trên R thì X là siêu khóa của R.

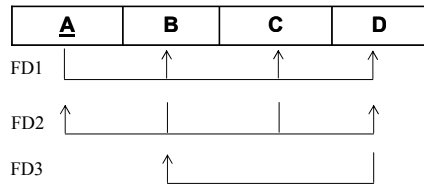
– Ví dụ:



- Lược đồ trên đạt DC3 nhưng không đạt DC BCK

## Dạng chuẩn BCK

– Lược đồ đạt DC3 vẫn còn trùng lặp dữ liệu:



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

<u>A</u>	C	D
----------	---	---



B	<u>D</u>
---	----------



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

B	D
a	1
b	2

## Dạng chuẩn BCK

– Nhận xét:

- Mọi lược đồ quan hệ đạt DC BCK cũng thuộc DC3.
- Mục tiêu của quá trình thiết kế CSDL là đưa các lược đồ quan hệ về DC3 hoặc BCK.

- Dạng chuẩn của lược đồ CSDL là dạng chuẩn thấp nhất mà các lược đồ quan hệ trong CSDL đó đều thỏa.

## Thiết kế CSDL

- Một chiến lược TK CSDL:
  - Thiết kế mô hình DL ở mức quan niệm (ER).
  - Chuyển mô hình DL mức quan niệm sang mức logic.
  - Với mỗi quan hệ, xác định tập PTH.
  - Nâng chuẩn của 1 lược đồ quan hệ bằng cách tách lược đồ để loại bỏ PTH không đầy đủ hoặc PTH bắc cầu.