

CHUẨN HÓA LỰC ĐỒ QUAN HỆ

CHUẨN HÓA DỰA TRÊN KHÓA CHÍNH

Xác định khóa cho một lược đồ quan hệ:

❖ **Định nghĩa:** Nếu R là một lược đồ quan hệ với các thuộc tính A_1, A_2, \dots, A_n và một tập các phụ thuộc hàm F trong đó $X \subseteq \{A_1, A_2, \dots, A_n\}$ thì X là một khóa của R khi:

1. $X \rightarrow A_1 A_2 \dots A_n \in F^+$, và

2. Không có tập con thực sự nào $Y \subseteq X$ mà $Y \rightarrow A_1 A_2 \dots A_n \in F^+$.

\Rightarrow Về cơ bản, định nghĩa này cho thấy phải tạo ra bao đóng của tất cả các tập con có thể có của R và quyết định xem tập nào suy diễn ra được tất cả các thuộc tính của lược đồ.

VÍ DỤ: XÁC ĐỊNH KHÓA

Cho $r = (C, T, H, R, S, G)$ với tập phụ thuộc hàm

$F = \{C \rightarrow T, HR \rightarrow C, HT \rightarrow R, CS \rightarrow G, HS \rightarrow R\}$

Bước 1: Tính $(A_i)^+$ với $1 \leq i \leq n$

$$\binom{6}{1} = \frac{6!}{1! \times (6-1)!} = \frac{720}{120} = 6$$

$C^+ = \{CT\}, \quad T^+ = \{T\}, \quad H^+ = \{H\}$

$R^+ = \{R\}, \quad S^+ = \{S\}, \quad G^+ = \{G\}$

\Rightarrow Không có thuộc tính đơn nào là khóa của r

Bước 2: Tính $(A_i A_j)^+$ với $1 \leq i \leq n, 1 \leq j \leq n$

$$\binom{6}{2} = \frac{6!}{2! \times (6-2)!} = \frac{720}{48} = 15$$

$(CT)^+ = \{C, T\}, \quad (CH)^+ = \{CHTR\},$

$(CR)^+ = \{CRT\}$

$(CS)^+ = \{CSGT\}, \quad (CG)^+ = \{CGT\},$

$(TH)^+ = \{THRC\}$

$(TR)^+ = \{TR\}, \quad (TS)^+ = \{TS\},$

$(TG)^+ = \{TG\}$

$(HR)^+ = \{HRCT\}, \quad (HS)^+ = \{HSRCTG\},$

$(HG)^+ = \{HG\}$

$(RS)^+ = \{RS\}, \quad (RG)^+ = \{RG\},$

$(SG)^+ = \{SG\}$

Tập thuộc tính (HS) là một khóa của r

VÍ DỤ: XÁC ĐỊNH KHÓA (cont.)

Bước 3: Tính $(A_i A_j A_k)^+$ với $1 \leq i \leq n$, $1 \leq j \leq n$, $1 \leq k \leq n$

$$(CTH)^+ = \{CTHR\},$$

$$(CTS)^+ = \{CTSG\},$$

$$(CHR)^+ = \{CHRT\},$$

$$(CHG)^+ = \{CHGTR\},$$

$$(CRG)^+ = \{CRGT\},$$

$$(THR)^+ = \{THRC\},$$

$$(THG)^+ = \{THGRC\},$$

$$(TRG)^+ = \{TRG\},$$

$$(HRS)^+ = \{HRSCTG\},$$

$$(HSG)^+ = \{HSGRCT\},$$

$$(CTR)^+ = \{CTR\}$$

$$(CTG)^+ = \{CTG\}$$

$$(CHS)^+ = \{CHSTRG\}$$

$$(CRS)^+ = \{CRSTG\}$$

$$(CSG)^+ = \{CSGT\}$$

$$(THS)^+ = \{THSRCG\}$$

$$(TRS)^+ = \{TRS\}$$

$$(TSG)^+ = \{TSG\}$$

$$(HRG)^+ = \{HRGCT\}$$

$$(RSG)^+ = \{RSG\}$$

$$\binom{6}{3} = \frac{6!}{3! \times (6-3)!} = \frac{720}{36} = 20$$

Các siêu khóa được đánh dấu màu đỏ.

VÍ DỤ: XÁC ĐỊNH KHÓA (cont.)

Bước 4: Tính $(A_i A_j A_k A_r)^+$ với $1 \leq i \leq n, 1 \leq j \leq n, 1 \leq k \leq n, 1 \leq r \leq n$

$$(CTHR)^+ = \{CTHR\},$$

$$(CTHS)^+ = \{CTHSRG\}$$

$$(CTHG)^+ = \{CTHGR\},$$

$$(CHRS)^+ = \{CHRSTG\}$$

$$(CHRG)^+ = \{CHRGTR\},$$

$$(CRSG)^+ = \{CRSGT\}$$

$$(THRS)^+ = \{THRSCG\},$$

$$(THRG)^+ = \{THRGC\}$$

$$(TRSG)^+ = \{TRSG\},$$

$$(HRSG)^+ = \{HRSGCT\}$$

$$(CTRS)^+ = \{CTRS\},$$

$$(CTSG)^+ = \{CTSG\}$$

$$(CSHG)^+ = \{CSHGTR\},$$

$$(THSG)^+ = \{THSGRC\}$$

$$(CTRG)^+ = \{CTRG\}$$

Các siêu khóa được đánh dấu màu đỏ.

$$\binom{6}{4} = \frac{6!}{4! \times (6-4)!} = \frac{720}{48} = 15$$

VÍ DỤ: XÁC ĐỊNH KHÓA (cont.)

Bước 5: Tính $(A_i A_j A_k A_r A_s)^+$ với $1 \leq i \leq n, 1 \leq j \leq n, 1 \leq k \leq n,$
 $1 \leq r \leq n, 1 \leq s \leq n$

$$(CTHRS)^+ = \{CTHSRG\}$$

$$(CTHRG)^+ = \{CTHGR\}$$

$$(CTHSG)^+ = \{CTHSGR\}$$

$$(CHRSR)^+ = \{CHRSRG\}$$

$$(CTRSG)^+ = \{CTRSG\}$$

$$(THRSG)^+ = \{THRSGC\}$$

Các siêu khóa được đánh dấu màu đỏ

$$\binom{6}{5} = \frac{6!}{5! \times (6-5)!} = \frac{720}{120} = 6$$

VÍ DỤ: XÁC ĐỊNH KHÓA (cont.)

Bước 6: Tính $(A_i A_j A_k A_r A_s A_t)^+$ với $1 \leq i \leq n, 1 \leq j \leq n, 1 \leq k \leq n,$
 $1 \leq r \leq n, 1 \leq s \leq n, 1 \leq t \leq n$

$$(CTHRSG)^+ = \{CTHSRG\}$$

Siêu khóa được đánh dấu màu đỏ.

$$\binom{6}{6} = \frac{6!}{6! \times (6-6)!} = \frac{720}{720} = 1$$

Với 6 thuộc tính, số trường hợp phải xét là:

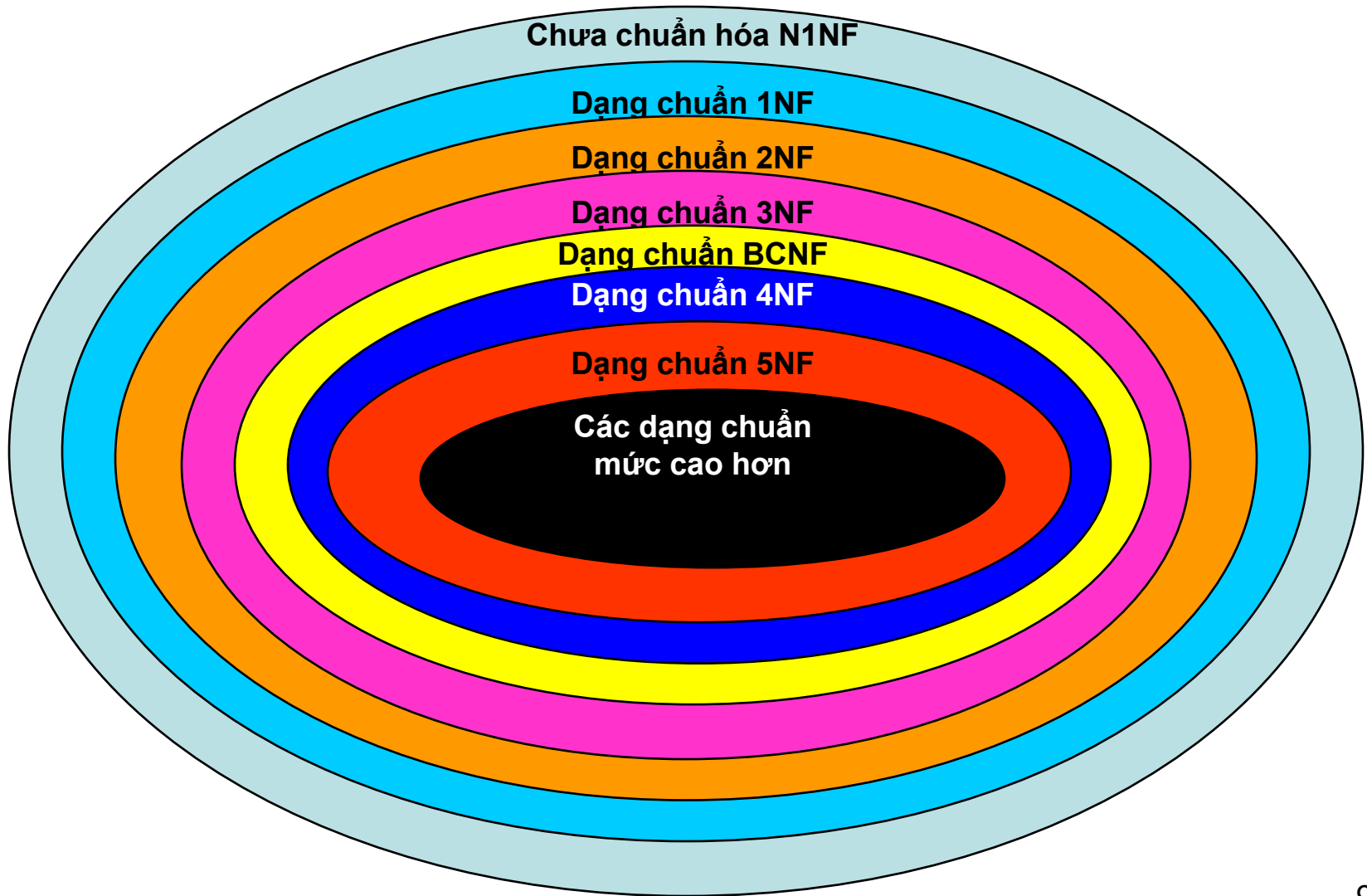
$$\binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 6 + 15 + 20 + 15 + 1 = 63$$

Bài tập: Tìm tất cả các khóa của $R = (A, B, C, D)$ với tập phụ thuộc hàm $F = \{A \rightarrow B, B \rightarrow C\}$

CHUẨN HÓA DỰA TRÊN KHÓA CHÍNH (Cont.)

- ❖ **Chuẩn hóa** là một kỹ thuật chính thức dùng cho việc phân tích các quan hệ dựa trên khóa chính (hoặc khóa dự bị), các thuộc tính và các phụ thuộc hàm.
- ❖ Kỹ thuật chuẩn hóa liên quan đến tập các luật được sử dụng để kiểm tra các quan hệ sao cho CSDL có thể đạt chuẩn hóa tới một mức nào đó.
- ❖ Khi quan hệ vi phạm một luật, cần phải tách nó ra thành một số các quan hệ khác nhỏ hơn.
- ❖ Việc chuẩn hóa được thực hiện bao gồm một số các bước, mỗi bước liên quan đến một dạng chuẩn cụ thể với các tính chất rõ ràng.

QUAN HỆ GIỮA CÁC DẠNG CHUẨN



CÁC YÊU CẦU CHUẨN HÓA

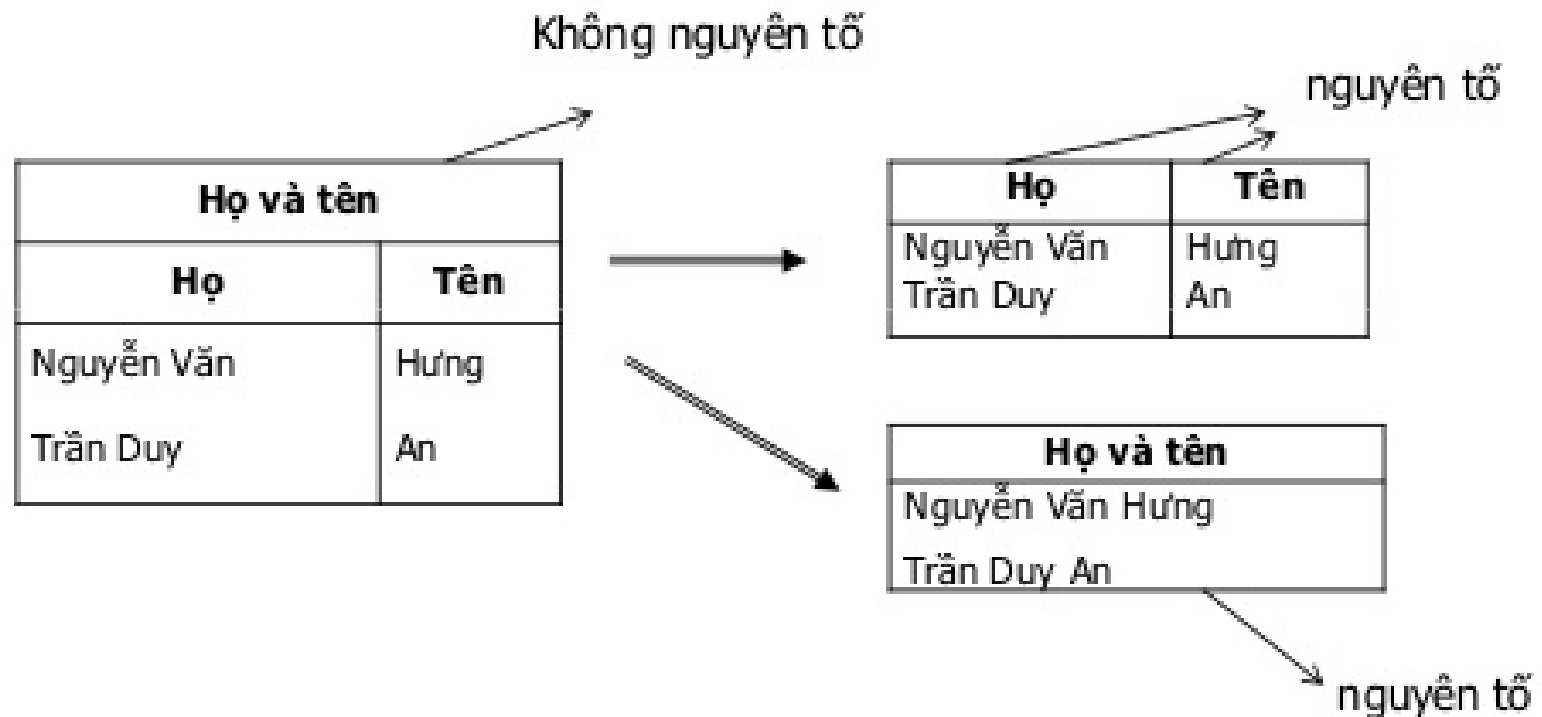
- ❖ Đối với mô hình quan hệ, một vấn đề rất quan trọng và thiết yếu là phải nhận ra được một quan hệ vừa tạo ra đã ở dạng chuẩn 1 (1NF) hay chưa. Tất cả các dạng chuẩn ở mức cao hơn sau đó là tùy theo từng trường hợp, có thể có hoặc không.
- ❖ Tuy nhiên, để tránh trường hợp dị thường thông tin khi cập nhật dữ liệu, người thiết kế CSDL thường được khuyến nghị là phải đưa toàn bộ các quan hệ trong CSDL về ít nhất là dạng chuẩn 3 (3NF).
- ❖ Việc chuẩn hóa sẽ được thực hiện từ dạng chưa chuẩn hóa, đưa về dạng chuẩn 1, sau đó đưa về dạng chuẩn 2, dạng chuẩn 3, ... (đến các dạng chuẩn ở các mức cao hơn).

DẠNG CHƯA CHUẨN HÓA (Non-first Normal Form - N1NF)

- ❖ Các quan hệ ở dạng chưa chuẩn hóa đồng nghĩa với việc chúng chưa ở dạng chuẩn 1.
- ❖ **Các quan hệ chưa ở dạng chuẩn 1** chứa một hoặc một số thuộc tính không nguyên tố, các thuộc tính lặp, và các thuộc tính dẫn xuất.
- ❖ **Thuộc tính chứa giá trị nguyên tố:** là những thuộc tính chứa giá trị đơn và không thể phân rã được nữa.

VÍ DỤ

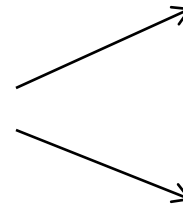
- Chuyển đổi các thuộc tính ghép về các thuộc tính đơn.



VÍ DỤ

- Chuyển đổi các thuộc tính đa trị thành các thuộc tính đơn trị.

<u>SN</u>	Tên	Tuổi	Môn học
0012	Hùng	19	C++, Java
0123	Toản	20	C++
0032	Hồng	18	CSDL
0133	Anh	20	Toán, Java



<u>SN</u>	Tên	Tuổi
0012	Hùng	19
0123	Toản	20
0032	Hồng	18
0133	Anh	20

<u>SN</u>	<u>Môn học</u>
0012	C++
0012	Java
0123	C++
0032	CSDL
0133	Toán
0133	Java

DẠNG CHUẨN 1

(First Normal Form - 1NF)

❖ ***Một quan hệ ở dạng chuẩn 1:***

- ❖ Mọi giá trị thuộc tính của quan hệ đều ở dạng nguyên tố.
- ❖ Không có thuộc tính đa trị.
- ❖ Không có thuộc tính dẫn xuất.

Một số khái niệm về khóa

- **Siêu khóa:** Là một tập các thuộc tính xác định duy nhất thực thể trong quan hệ.
- **Khóa:** là một siêu khóa mà khi loại bỏ bất kỳ thuộc tính nào từ khóa này thì nó không còn là một siêu khóa nữa. Nghĩa là, khóa có số thuộc tính là nhỏ nhất.
- **Khóa dự bị:** là một tập các thuộc tính khóa nhỏ nhất của lược đồ quan hệ.
- **Khóa chính:** là một khóa dự bị được chọn ra. Tất cả các khóa dự bị còn lại trở thành khóa phụ hay khóa thứ cấp.

Một số khái niệm về khóa (tiếp)

- ***Thuộc tính khóa:*** là thuộc tính của quan hệ R và là thành viên của một khóa dự bị nào đó.
- ***Thuộc tính không khóa:*** là thuộc tính của quan hệ R mà không phải là thành viên của một khóa dự bị nào.

DẠNG CHUẨN 2 (Second Normal Form - 2NF)

- ❖ **Dạng chuẩn 2** dựa trên khái niệm **phụ thuộc hàm đầy đủ**.
- ❖ **Một phụ thuộc hàm $X \rightarrow Y$ được gọi là phụ thuộc hàm đầy đủ** nếu loại bỏ bất kỳ thuộc tính A nào của X thì sẽ làm phụ thuộc hàm này không còn đúng nữa. Nói cách khác $X \rightarrow Y$ tối giản trái.

=> Không có bất kỳ thuộc tính $A \in X$ nào mà $X - \{A\} \rightarrow Y$
- ❖ **Một phụ thuộc hàm $X \rightarrow Y$ là phụ thuộc hàm không đầy đủ** nếu tồn tại một thuộc tính A nào đó của X bị loại bỏ mà phụ thuộc hàm đó vẫn đúng.

=> Tồn tại một thuộc tính $A \in X$ mà $X - \{A\} \rightarrow Y$

DẠNG CHUẨN 2 (Second Normal Form - 2NF) (Cont.)

- ❖ **Định nghĩa dạng chuẩn 2:** Một lược đồ quan hệ R với tập phụ thuộc hàm F ở dạng chuẩn 2NF khi thỏa mãn:
 - Là dạng chuẩn 1NF
 - Mọi thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào mọi khóa dự bị của R.

DẠNG CHUẨN 2

(Second Normal Form - 2NF) (Cont.)

- ❖ Một lược đồ quan hệ R với tập phụ thuộc hàm F ở dạng chuẩn 2NF khi thỏa mãn một trong các điều kiện sau:
 - Tất cả các khóa dự bị đều có một thuộc tính
 - Không có thuộc tính không khóa nào.
 - Không tồn tại thuộc tính không khóa mà phụ thuộc hàm một phần vào một khóa dự bị của R.

VÍ DỤ

Ví dụ: Cho $R = (A, D, P, G)$, với tập phụ thuộc hàm $F = \{AD \rightarrow P, A \rightarrow G\}$ và một khóa $K = \{AD\}$

R không ở dạng chuẩn 2NF vì G phụ thuộc một phần vào khóa AD do $A \rightarrow G$.

Phân rã R thành:

$$R_1 = (A, D, P)$$

$$R_2 = (A, G)$$

$$K_1 = \{AD\}$$

$$K_2 = \{A\}$$

$$F_1 = \{AD \rightarrow P\}$$

$$F_2 = \{A \rightarrow G\}$$

CHUẨN HÓA TỪ 1NF SANG 2NF

- ❖ Loại bỏ các thuộc tính không khóa mà phụ thuộc một phần vào khóa chính để tách thành bảng riêng. Khóa chính của bảng riêng là bộ phận khóa mà chúng phụ thuộc vào.
- ❖ Các thuộc tính còn lại lập thành một bảng với khóa chính là khóa ban đầu.

DẠNG CHUẨN 3 (Third Normal Form - 3NF)

- ❖ **Dạng chuẩn 3** dựa trên khái niệm phụ thuộc hàm bậc cầu.
- ❖ **Định nghĩa phụ thuộc hàm bậc cầu:** Cho một lược đồ quan hệ R và một tập các phụ thuộc hàm F của R , một tập con $X \subseteq R$ và một thuộc tính $A \in R$. A được gọi là phụ thuộc hàm bậc cầu vào X nếu tồn tại $Y \subseteq R$ mà $X \rightarrow Y$, Y không $\rightarrow X$ và $Y \rightarrow A$, và $A \notin X \cup Y$.
- ❖ **Một cách định nghĩa khác cho phụ thuộc hàm bậc cầu:** Một phụ thuộc hàm $X \rightarrow Y$ trong lược đồ quan hệ R là phụ thuộc bậc cầu nếu có một tập thuộc tính $Z \subseteq R$ mà Z không phải là tập con của bất kỳ khóa nào của R nhưng $X \rightarrow Z$ và $Z \rightarrow Y$.

DẠNG CHUẨN 3

(Third Normal Form - 3NF) (Cont.)

- ***Định nghĩa dạng chuẩn 3NF:*** Một lược đồ quan hệ R ở dạng 3NF với một tập phụ thuộc hàm F nếu thỏa mãn:
 - Với bất kỳ phụ thuộc hàm $X \rightarrow A$ trong F thì hoặc X là một siêu khóa của R, hoặc A là một thuộc tính khóa.
- ***Một cách định nghĩa khác về dạng chuẩn 3NF:*** Một lược đồ quan hệ R ở dạng 3NF với một tập phụ thuộc hàm F nếu thỏa mãn:
 - R ở dạng 2NF
 - Không có thuộc tính không khóa nào phụ thuộc hàm bắc cầu vào khóa của R.

DẠNG CHUẨN 3

(Third Normal Form - 3NF) (Cont.)

❖ ***Chuẩn hóa về 3NF (cách 1):***

- Tìm phủ tối thiểu của tập phụ thuộc hàm của lược đồ ban đầu
- Loại bỏ các thuộc tính phụ thuộc bắc cầu ra khỏi quan hệ, lập thành bảng mới với khóa chính là thuộc tính bắc cầu
- Các thuộc tính còn lại lập thành một bảng với khóa chính là khóa ban đầu.

DẠNG CHUẨN 3

(Third Normal Form - 3NF) (Cont.)

❖ ***Chuẩn hóa về 3NF (cách 2):***

- Tìm phủ tối thiểu (G) của tập phụ thuộc hàm (F) của lược đồ (R) ban đầu
- Ứng với mỗi phụ thuộc hàm trong G, lập một bảng quan hệ mới
- Nếu trong tất cả các bảng mới tạo ra tại bước 2 không có bảng nào chứa một khóa dự bị nào của R thì tạo ra thêm một bảng mới chứa một khóa dự bị của R.

VÍ DỤ 1

Cho $R = (A, B, C, D)$, khóa $K = \{A, B\}$, và tập phụ thuộc hàm $F = \{AB \rightarrow C, C \rightarrow D\}$

R không là dạng chuẩn 3NF vì D là một thuộc tính không khóa mà lại phụ thuộc bắc cầu vào khóa AB .

CHUẨN HÓA

- Chuẩn hóa về dạng 1NF: loại bỏ dữ liệu dư thừa
- Chuẩn hóa về 2NF: loại bỏ các phụ thuộc hàm bộ phận
- Chuẩn hóa về 3NF: loại bỏ các phụ thuộc hàm bắc cầu

DẠNG CHUẨN BOYCE-CODD

(Boyce-Codd Nomal Form - BCNF)

- ❖ Dạng chuẩn Boyce-Codd (BCNF) là một chuẩn nghiêm ngặt hơn chuẩn 3NF.
- ❖ **Định nghĩa:** Một lược đồ quan hệ R được coi là ở dạng chuẩn Boyce-Codd với tập phụ thuộc hàm F nếu với bất kỳ phụ thuộc hàm $X \rightarrow A$ nào và $A \not\subseteq X$, thì X là một siêu khóa của R .
- ❖ **Ví dụ:** Cho lược đồ quan hệ $R = (A, B, C)$, tập phụ thuộc hàm $F = \{AB \rightarrow C, C \rightarrow A\}$ và khóa $K = \{AB\}$

R không ở dạng chuẩn BCNF vì có $C \rightarrow A$ và C không phải là một siêu khóa của R .

DẠNG CHUẨN BOYCE-CODD **(Boyce-Codd Nomal Form - BCNF) (Cont.)**

- ❖ Sự khác nhau giữa dạng chuẩn 3NF và BCNF là BCNF cho phép bỏ đi luật với phụ thuộc hàm $X \rightarrow A$ thì X phải là thuộc tính khóa.
- ❖ Trong thực tế, hầu hết các lược đồ quan hệ ở dạng chuẩn 3NF thì cũng ở dạng chuẩn BCNF. Chỉ khi trong lược đồ có $X \rightarrow A$ mà X không phải là một siêu khóa hoặc A là một thuộc tính khóa thì lược đồ này ở dạng chuẩn 3NF mà không ở dạng chuẩn BCNF.

PHÂN TÁCH LƯỢC ĐỒ VỀ CÁC DẠNG CHUẨN

- ❖ Mục đích cơ bản của thiết kế CSDL quan hệ là đảm bảo mọi quan hệ trong CSDL đều ở dạng chuẩn 3NF hoặc BCNF. Điều này giúp loại bỏ được hầu hết các dị thường thông tin khi cập nhật.
- ❖ Để đưa một lược đồ CSDL về dạng chuẩn 3NF hoặc BCNF, người thiết kế phải đảm bảo thỏa mãn 2 tính chất:
 1. Kết nối không mất mát thông tin
 2. Các phụ thuộc hàm được bảo toàn sau khi phân tách.

PHÂN TÁCH LƯỢC ĐỒ VỀ CÁC DẠNG CHUẨN

(Cont.)

- ❖ Hiện tại, có một số thuật toán phân tách về 3NF đảm bảo cả 2 tính chất (1) và (2).
 - ❖ Tuy nhiên, không có thuật toán nào phân tách về BCNF mà đảm bảo cả tính chất (1) và (2).
 - Chỉ có một thuật toán phân tách về BCNF đảm bảo tính chất (1) nhưng không đảm bảo tính chất (2).
- => *Chuẩn 3NF được coi là dạng chuẩn mạnh trong khả năng có thể để phân tách các lược đồ quan hệ. Nếu cố gắng đưa về BCNF thì có thể dẫn tới việc không bảo toàn các phụ thuộc hàm.*