

Chương 7

Chuẩn hóa

TS. Võ Đức Hoàng

Khoa Công nghệ thông tin

Trường Đại học Bách khoa - Đại học Đà Nẵng

Nội dung chi tiết



- Giới thiệu về chuẩn hóa
- Các dạng chuẩn dựa trên khóa chính

Giới thiệu về chuẩn hóa



- Quá trình chuẩn hóa (Codd đề nghị năm 1972): lấy một lược đồ quan hệ và thực hiện các phép kiểm tra để xác nhận lược đồ có thỏa mãn một số điều kiện nhất định (dạng chuẩn) hay không; nếu không các lược đồ quan hệ sẽ được tách ra
- Codd đề xuất ba dạng chuẩn 1, 2 và 3 (1NF, 2NF, 3NF)
- Sau đó Boyce và Codd đề nghị dạng chuẩn Boyce-Codd (BCNF) mạnh hơn 3NF
- Các dạng chuẩn này dựa trên các phụ thuộc hàm giữa các thuộc tính của một lược đồ quan hệ
- Sau đó dạng chuẩn 4 (4NF) và 5 (5NF) được đề xuất dựa trên phụ thuộc hàm đa trị và phụ thuộc hàm nối

- Chuẩn hóa dữ liệu có thể được xem là quá trình phân tích và tách các lược đồ quan hệ để đạt được hai mục tiêu:
 - Loại bỏ các bất thường của 1 quan hệ để có được các quan hệ có cấu trúc tốt hơn, nhỏ hơn
 - Giảm thiểu sự dư thừa dữ liệu ở mức thấp nhất và cho phép người dùng thêm, sửa, xóa mà không gây ra mâu thuẫn dữ liệu
- Phép tách các lược đồ quan hệ cần có hai tính chất:
 - Tính chất nối không mất mát (phải đạt được)
 - Tính chất bảo toàn phụ thuộc

Các dạng chuẩn dựa trên khóa chính

Dạng chuẩn 1 (1NF)



- Một lược đồ quan hệ $R(A_1, A_2, \dots, A_n)$ là ở **dạng chuẩn 1 (1NF - first normal form)** nếu $Dom(A_i)$ chỉ chứa các giá trị nguyên tố (không phân chia được) \forall và
- $\forall t = \langle t_1, t_2, \dots, t_n \rangle \in r(R), t_i \in Dom(A_i), t_i$ nhận giá trị đơn
- Như vậy 1NF không cho phép có các thuộc tính đa trị hoặc các nhóm thuộc tính lặp

Thuộc tính đa trị và nhóm thuộc tính lặp



MaDV	TenDV	MaNQL	Diadiem
5	Nghiên cứu	NV002	Nam Định, Hà Nội, Bắc Ninh
4	Hành chính	NV014	Hà Nội
1	Tài vụ	NV061	Hà Nội

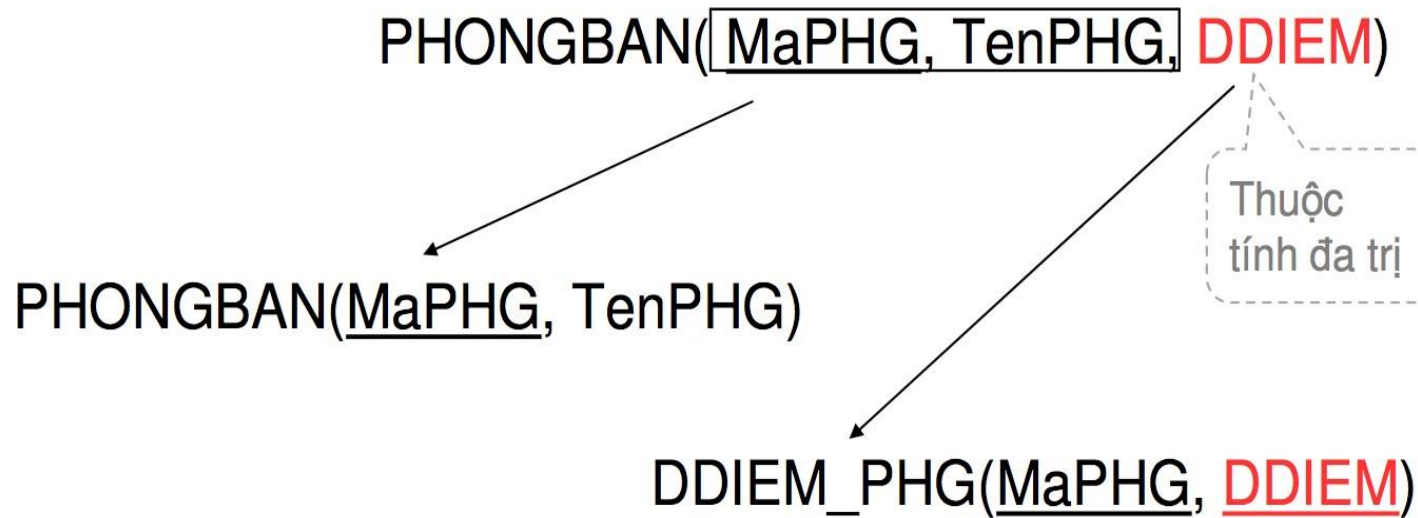
MaDA	TenDA	TenNV	Sogio
1	DA01	Vân, Nam	15, 20
2	DA02	Nam, Thanh, Bằng	10, 12, 28
3	DA03	Thanh	20

- Giả sử K là khóa của lược đồ quan hệ R và A là thuộc tính đa trị hoặc là nhóm thuộc tính lặp
- Để chuẩn hóa R về 1NF:
 - Loại A ra khỏi R
 - Tạo một lược đồ quan hệ mới $S(K, A)$ với khóa chính của S là:
 - $\{K, A\}$ nếu A là thuộc tính đa trị
 - $\{K, KA\}$ nếu A là nhóm thuộc tính lặp, KA là khóa bộ phận của A

Ví dụ chuẩn hóa 1NF

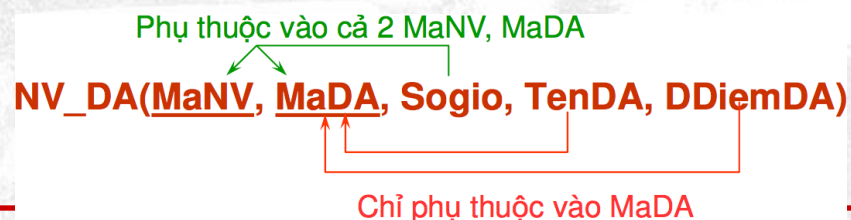
1. Lược đồ $DONVI(\underline{MaDV}, TenDV, MaNQL, Diadiem)$, với $Diadiem$ là thuộc tính đa trị:
 - Loại $Diadiem$ ra khỏi $DONVI$: $DONVI_1(\underline{MaDV}, TenDV, MaNQL)$
 - Tạo lược đồ quan hệ mới $DIADIEMDV(\underline{MaDV}, \underline{Diadiem})$
2. Lược đồ $NHANVIEN_LAMVIEC(\underline{MaDA}, TenDA, MaNV, Sogio)$, trong đó $(MaNV, Sogio)$ là nhóm thuộc tính lặp
 - Loại $(MaNV, Sogio)$ ra khỏi $NHANVIEN_LAMVIEC$:
 $NHANVIEN_LAMVIEC_1(\underline{MaDA}, TenDA)$
 - Tạo lược đồ quan hệ mới $NHANVIEN_SOGIO(\underline{MaDA}, \underline{MaNV}, Sogio)$

Ví dụ chuẩn hóa 1NF



Dạng chuẩn 2 (2NF)

- Một phụ thuộc hàm $X \rightarrow Y$ là một **phụ thuộc hàm đầy đủ** nếu bỏ đi bất kỳ một thuộc tính $A \in X$ thì phụ thuộc hàm đó không còn đúng nữa
- Nếu tồn tại một thuộc tính B sao cho $(X - B) \rightarrow Y$, chúng ta nói rằng Y **phụ thuộc hàm bộ phận** vào X
- Một lược đồ quan hệ R là ở **dạng chuẩn 2 (2NF - second normal form)** nếu:
 - R là 1NF, và
 - Mỗi thuộc tính không khóa A trong R phụ thuộc hàm đầy đủ vào khóa chính của R

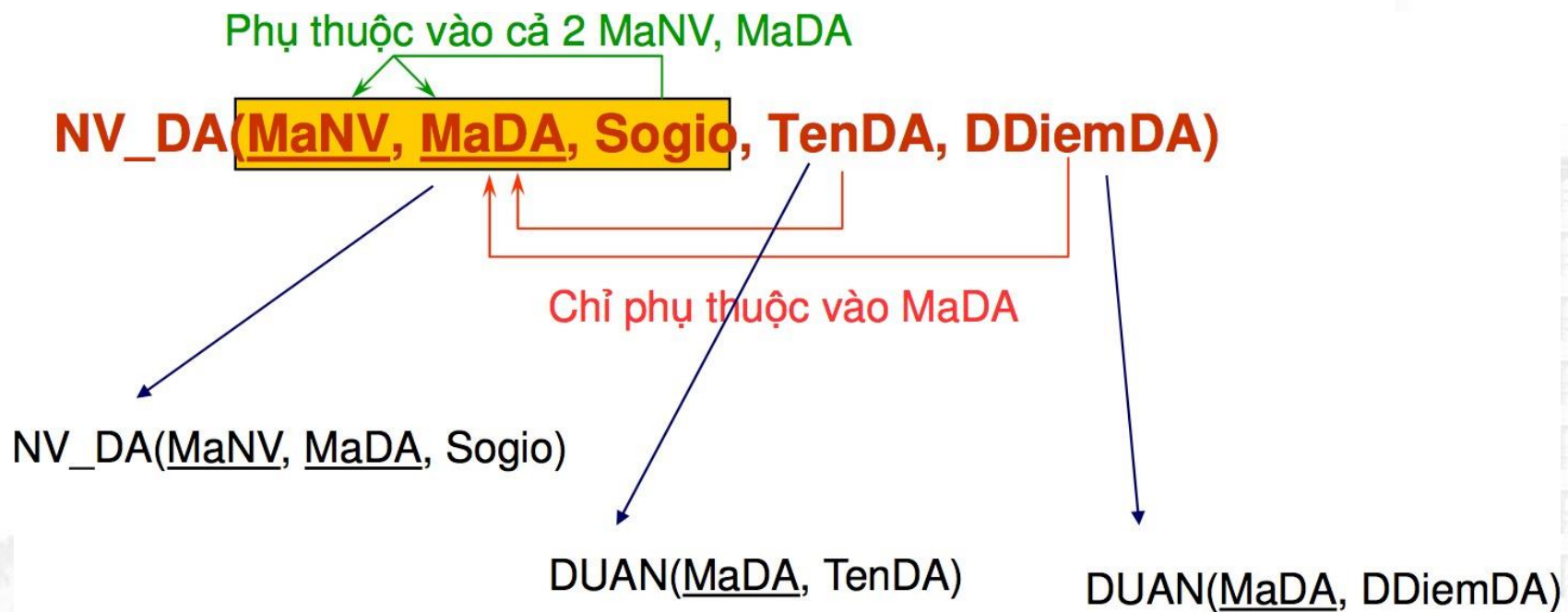


- Giả sử R là lược đồ quan hệ ở dạng chuẩn 1NF và không là 2NF, có khóa K
- Để chuẩn hóa R về 2NF, đối với mỗi nhóm thuộc tính không khóa X phụ thuộc hàm bộ phận vào K :
 - Loại X ra khỏi R
 - Gọi khóa bộ phận của K xác định hàm X là $K_X : K_X \rightarrow X$
 - Tạo lược đồ quan hệ mới $S(\underline{K_X}, X)$ với K_X là khóa chính

Ví dụ chuẩn hóa 2NF

- Lược đồ quan hệ ở dạng chuẩn 1NF $NHANVIEN_DUAN(\underline{MaNV}, \underline{MaDA}, Sogio, Hoten, TenDA, Diadiem)$ với các phụ thuộc hàm:
 - $\{MaNV, MaDA\} \rightarrow \{Sogio, Hoten, TenDA, Diadiem\}$
 - $MaNV \rightarrow Hoten$ (phụ thuộc hàm bộ phận)
 - $MaDA \rightarrow \{TenDA, Diadiem\}$ (phụ thuộc hàm bộ phận)
- $NHANVIEN_DUAN$ không là 2NF do $Hoten, TenDA$ và $Diadiem$ phụ thuộc hàm bộ phận vào khóa chính $\{MaNV, MaDA\}$
- Chuẩn hóa về 2NF:
 - Loại $Hoten, TenDA, Diadiem$ ra khỏi $NHANVIEN_DUAN$:
 $NHANVIEN_DUAN_1(\underline{MaNV}, \underline{MaDA}, Sogio)$
 - Tạo lược đồ quan hệ mới 1: $N_1(\underline{MaNV}, Hoten)$
 - Tạo lược đồ quan hệ mới 2: $N_2(\underline{MaDA}, TenDA, Diadiem)$

Ví dụ chuẩn hóa 2NF



Dạng chuẩn 3 (3NF)

- Một phụ thuộc hàm $X \rightarrow Y$ trong một lược đồ quan hệ R là một **phụ thuộc hàm bắc cầu** nếu tồn tại một tập hợp thuộc tính Z của R không phải là khóa và không phải là tập con của khóa sao cho $X \rightarrow Z$ và $Z \rightarrow Y$ đều đúng
- Một lược đồ quan hệ R là ở **dạng chuẩn 3 (3NF - third normal form)** nếu
 - R là 2NF, và
 - Không có thuộc tính không khóa nào của R phụ thuộc hàm bắc cầu vào khóa chính

- Giả sử R là lược đồ quan hệ có khóa chính K ; R ở dạng chuẩn 2NF và không ở dạng chuẩn 3NF
- Để chuẩn hóa R về 3NF, với mỗi thuộc tính X phụ thuộc bắc cầu vào K :
 - Loại bỏ X ra khỏi R
 - Gọi Y là thuộc tính bắc cầu, ta có: $K \rightarrow Y$ và $Y \rightarrow X$
 - Tạo lược đồ quan hệ mới $S(\underline{Y}, X)$ với Y là khóa chính

Ví dụ chuẩn hóa 3NF

- Lược đồ *LAMVIEC* (*MaNV*, *Hoten*, *Ngaysinh*, *MaDV*, *TenDV*, *MaDA*, *TenDA*) với các thuộc tính đơn, ở dạng chuẩn 2NF vì khóa chính chỉ có một thuộc tính (→ không có phụ thuộc hàm bộ phận), có các phụ thuộc hàm:
 - $MaNV \rightarrow \{Hoten, Ngaysinh, MaDV, TenDV, MaDA, TenDA\}$
 - $MaDV \rightarrow TenDV$
 - $MaDA \rightarrow TenDA$
- *LAMVIEC* không ở dạng chuẩn 3NF vì các phụ thuộc hàm bắc cầu $MaNV \rightarrow MaDV \rightarrow TenDV$ và $MaNV \rightarrow MaDA \rightarrow TenDA$
- Chuẩn hóa *LAMVIEC* về 3NF:
 - Loại $\{TenDV, TenDA\}$ phụ thuộc bắc cầu vào khóa chính: $LAMVIEC_1(\underline{MaNV}, Hoten, Ngaysinh, MaDV, MaDA)$
 - Tạo lược đồ quan hệ mới $N_1(\underline{MaDV}, TenDV)$
 - Tạo lược đồ quan hệ mới $N_2(\underline{MaDA}, TenDA)$

Ví dụ chuẩn hóa 3NF

Phụ thuộc vào MaNV

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

Phụ thuộc vào MaDV

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

DONVI(MaDV, TenDV, TruongPHG)

Dạng chuẩn Boyce-Codd (BCNF)



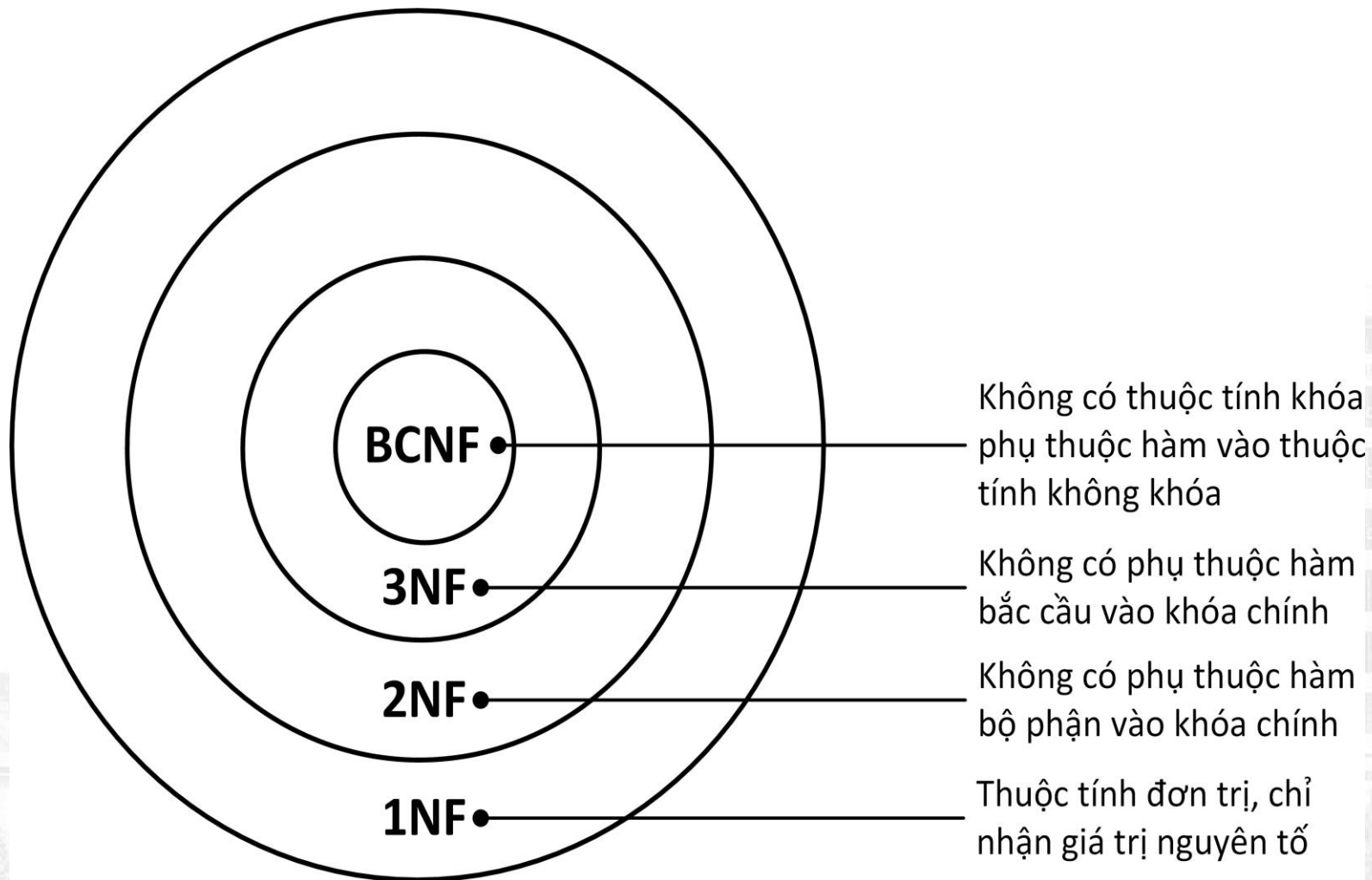
- Một lược đồ quan hệ R là ở **dạng chuẩn Boyce-Codd (BCNF - Boyce-Codd normal form)** nếu
 - R là 3NF, và
 - Không có thuộc tính khóa phụ thuộc hàm vào thuộc tính không khóa trong R

- Giả sử R là lược đồ quan hệ có khóa K , R là 3NF và không là BCNF
- Để chuẩn hóa R về BCNF, đối với mỗi thuộc tính không khóa A xác định hàm thuộc tính khóa $A \rightarrow K_S$, với $K_S \subset K$:
 - Loại K_S ra khỏi R và bổ sung A vào khóa chính của R . Khóa chính mới của R là: $K - K_S \cup \{A\}$
 - Tạo lược đồ mới $S(\underline{A}, K_S)$ với A là khóa chính

Ví dụ chuẩn hóa BCNF

- Lược đồ $R(\underline{A_1}, \underline{A_2}, A_3, A_4, A_5, A_6)$ có các phụ thuộc hàm:
 - $\{A_1, A_2\} \rightarrow \{A_3, A_4, A_5, A_6\}$: $\{A_1, A_2\}$ là khóa chính
 - $A_4 \rightarrow A_2$
 - $A_6 \rightarrow A_1$
- R là 2NF (không có phụ thuộc hàm bộ phận) và 3NF (không có phụ thuộc hàm bắc cầu từ khóa chính)
- R không là BCNF vì có thuộc tính khóa phụ thuộc hàm vào thuộc tính không khóa
- Chuẩn hóa về BCNF:
 - Loại A_2 ra khỏi R và bổ sung A_4 vào khóa chính: $R(\underline{A_1}, \underline{A_4}, A_3, A_5, A_6)$
 - Tạo quan hệ mới $S_1(\underline{A_4}, A_2)$ với A_4 là khóa chính Loại A_1 ra khỏi R và bổ sung A_6 vào khóa chính: $R(\underline{A_6}, \underline{A_4}, A_3, A_5)$
 - Tạo quan hệ mới $S_2(\underline{A_6}, A_1)$ với A_6 là khóa chính

Minh họa và "từ khóa" của các



Định nghĩa tổng quát các dạng chuẩn

Dạng chuẩn 1 (1NF)

- Một lược đồ quan hệ $R(A_1, A_2, \dots, A_n)$ là ở **dạng chuẩn 1 (1NF)** nếu $Dom(A_i)$ chỉ chứa các giá trị nguyên tố (không phân chia được) \forall và
- $\forall t = \langle t_1, t_2, \dots, t_n \rangle \in r(R), t_i \in Dom(A_i), t_i$ nhận giá trị đơn

Dạng chuẩn 2 (2NF)

- Một lược đồ quan hệ R là ở **dạng chuẩn 2 (2NF)** nếu mỗi thuộc tính không khóa trong R không phụ thuộc hàm bộ phận vào một khóa bất kỳ của R

Dạng chuẩn 3 (3NF)

■ Một lược đồ quan hệ R là ở **dạng chuẩn 3 (3NF)** nếu với mọi phụ thuộc hàm $X \rightarrow A$ thỏa mãn trong R thì ta đều có:

- X là một siêu khóa của R , hoặc
- A là một thuộc tính khóa của R

Dạng chuẩn Boyce-Codd (BCNF)



- Một lược đồ quan hệ R là ở **dạng chuẩn Boyce-Codd (BCNF)** nếu với mọi phụ thuộc hàm $X \rightarrow A$ thỏa mãn trong R thì ta đều có X là một siêu khóa của R

