Phụ thuộc hàm và Chuẩn hóa CSDL

TS.Nguyễn Quốc Tuấn
Bm. Mạng & HTTT

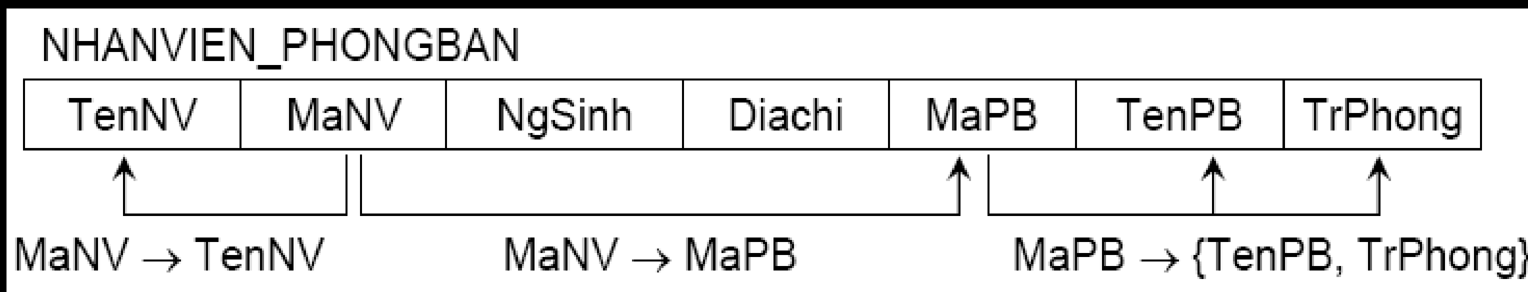
Nội dung

- Phụ thuộc hàm.
- Các dạng chuẩn.
- Một số thuật toán chuẩn hóa.

Phụ thuộc hàm (1)

- Phụ thuộc hàm(PTH) - Functional Dependencies
- Xét lược đồ quan hệ gồm n thuộc tính
  - R(U), U={A_1, A_2,…, A_n}
- PTH giữa hai tập thuộc tính X, Y ⊆ U
  - Ký hiệu: X → Y.
  - $\forall r \in R, \forall\, t_1, t_2 \in r$ nếu $t_1[X] = t_2[X]$ thì $t_1[Y] = t_2[Y]$.
- X là vế trái và Y là vế phải của PTH.
- X → Y được gọi là PTH hiển nhiên nếu Y ⊆ X
- X → Y được gọi là PTH nguyên tố (Y PTH đầy đủ vào X nếu nếu ∀X' ⊂ X thì X' không → Y

Phụ thuộc hàm (2)



- r ∈ R thỏa mãn các PTH gọi là trạng thái hợp lệ của R
- Nhận xét:
  - Các PTH xuất phát từ các ràng buộc trong thế giới thực.
  - ∀r ∈ R, ∀t ∈ r, t [X] là duy nhất thì X là một khóa của R.
  - Nếu K là một khóa của R thì K xác định hàm tất cả các tập thuộc tính của R.
  - PTH dùng để đánh giá một thiết kế CSDL

Bao đóng của tập PTH

- F là tập PTH trên R
  - F = {MaNV → TenNV, MaPB → {TenPB, TrPhong}, MaNV → MaPB}.
  - ∀r ∈ R thỏa F và MaNV → {TenPB, TrPhong} cũng đúng với r thì MaNV → {TenPB, TrPhong} gọi là được suy diễn từ F.
- Bao đóng của F, ký hiệu F^+, gồm
  - F
  - Tất cả các PTH được suy diễn từ F.
- F gọi là đầy đủ nếu $F = F^+$.

Luật suy diễn (1)

- Luật suy diễn dùng để suy diễn một PTH mới từ một tập PTH cho trước.
- Hệ luật suy diễn Armstrong
  - Phản xạ: $Y \subseteq X \Rightarrow X \rightarrow Y$.
  - Tăng trưởng: $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$, với $XZ = X \cup Z$.
  - Bắc cầu: $X \rightarrow Y, Y \rightarrow Z \Rightarrow X \rightarrow Z$.
- Các luật khác:
  - Phân rã: $X \rightarrow YZ \Rightarrow X \rightarrow Y, X \rightarrow Z$.
  - Hợp: $X \rightarrow Y, X \rightarrow Z \Rightarrow X \rightarrow YZ$.
  - Bắc cầu giả: $X \rightarrow Y, WY \rightarrow Z \Rightarrow WX \rightarrow Z$.

Luật suy diễn (2)

- Ví dụ 1:
  - Cho F={A → B, B → C, A → D}
  - Hãy chứng tỏ PTH A → CD suy diễn từ F nhờ luật dẫn Amstrong
  - Cách giải:
    - A → B, B → C ⇒ A → C (luật bắc cầu)
    - A → C, A → D ⇒ A → CD (luật hợp).
- Ví dụ 2: Cho F={AB→E,AG→I,BE→I,E→G,GI→H}
  - Hãy chứng tỏ PTH AB → GH suy diễn từ F nhờ luật dẫn Armstrong

Bao đóng của tập thuộc tính

- Làm thế nào để biết một PTH $X \rightarrow Y$ được suy diễn từ tập PTH F cho trước?
- Bao đóng của tập thuộc tính X đối với F, ký hiệu X^+ là
  - Tập các thuộc tính PTH vào X.
  - $X^+ = \{A \in U \mid X \rightarrow A \in F^+\}$
- Nhận xét:
  - $X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X^+$.
  - Nếu K là khóa của R thì $K^+ = U$.

Thuật toán tìm X^+

- Input: U, F và $X \subseteq U$
- Output: X^+
- Thuật toán
  - *B1*: $X^+ = X$;
  - *B2*: Nếu tồn tại $Y \rightarrow Z \in F$ và $Y \subseteq X^+$ thì
    - $X^+ = X^+ \cup Z$;
    - tiếp tục *B2*.
  - Ngược lại qua *B3*.
  - *B3*: output X^+

Ví dụ tìm X^+

- Input:
  - $F = \{AB \rightarrow C, BC \rightarrow D, D \rightarrow EG\}$
  - $X = BD$
- Output: X^+
- Thuật toán
  - $X^+ = BD$.
  - Lặp 1:
    - Tìm các PTH có vế trái là tập con của $X^+ = BD$
      - $D \rightarrow EG$, thêm EG vào X^+ ta được $X^+ = BDEG$.
  - Lặp 2:
    - Tìm các PTH có vế trái là tập con của $X^+ = BDEG$
      - Không có PTH nào.
  - Vậy $X^+ = BDEG$.

Ví dụ tìm X^+

- VD2: Cho lược đồ quan hệ Q(ABCDEGH) và tập PTH F
  - $F = \{ B \rightarrow A, DA \rightarrow CE, D \rightarrow H, GH \rightarrow C, AC \rightarrow D \}$
  - Tìm bao đóng của tập $X=\{AC\}$ dựa trên F
- VD3: Cho lược đồ quan hệ Q(ABCDEGH) và tập PTH F
  - $F = \{A \rightarrow C, A \rightarrow EG, B \rightarrow D, G \rightarrow E\}$
  - Xác định X^+
    - $X= \{AB\}$
    - $X=\{CGD\}$

Các tập PTH tương đương

- Tập PTH F được nói là phủ tập PTH G nếu $G \subset F^+$
- Hai tập PTH F và G là tương đương nếu
  - F phủ G và
  - G phủ F
- Nhận xét
  - $\forall X \rightarrow Y \in G$, nếu $Y \subseteq X_F^+$ thì F phủ G.
  - F và G tương đương nếu và chỉ nếu $F^+ = G^+$

Kiểm tra PTH suy diễn

- Cho F = {AB → C, A → D, D → E, AC → B}
  - Hai PTH AB → E và D → C có được suy diễn từ F hay không?

Tập PTH tối thiểu (1)

- Thừa PTH
  - {A → B, B → C, A → C}, vì A → C được suy diễn từ {A → B, B → C}
  - A → B, B → C ⇒ A → C (luật bắc cầu).
- Thừa thuộc tính
  - {A → B, B → C, A → CD}, vì A → CD được suy diễn từ {A → B, B → C, A → D}
    - A → B, B → C ⇒ A → C (luật bắc cầu)
    - A → C, A → D ⇒ A → CD (luật hợp).
  - {A → B, B → C, AC → D}, vì AC → D được suy diễn từ {A → B, B → C, A → D}
    - A → B, A → D ⇒ A → BD (luật hợp)
    - A → BD ⇒ AC → BCD (luật tăng trưởng)
    - AC → BCD ⇒ AC → D (luật phân rã).

Tập PTH tối thiểu (2)

- Tập PTH F là tối thiểu nếu thỏa các điều kiện sau:
  - Mọi PTH của F chỉ có một thuộc tính ở vế phải.
  - Không thể thay $X \rightarrow A$ thuộc F bằng $Y \rightarrow A$ với $Y \subset X$ mà tập mới tương đương với F.
  - Nếu bỏ đi một PTH bất kỳ trong F thì tập PTH còn lại không tương đương với F.
- Phủ tối thiểu của tập PTH E là tập PTH tối thiểu F tương đương với E.
- Nhận xét
  - Mọi tập PTH có ít nhất một phủ tối thiểu.

Thuật toán tìm tập PTH tối thiểu

- Input: tập PTH E.
- Output: phủ tối thiểu F của E.
- Thuật toán:
  - *B1*: $F = \varnothing$
  - *B2*: Với mọi $X \rightarrow Y \in E$, $Y = \{A_1, \ldots, A_k\}$, $A_i \in U$
    - $F = F \cup \{X \rightarrow \{A_i\}\}$
  - *B3*: Với mỗi $X \rightarrow \{A\} \in F$, $X = \{B_1, \ldots, B_l\}$, $B_i \in U$
    - Với mỗi B_i, nếu $A \in (X - \{B_i\})_F^+$ thì
      - $F = (F - \{X \rightarrow \{A\}\}) \cup \{(X - \{B_i\}) \rightarrow \{A\}\}$
  - *B4*: Với mỗi $X \rightarrow \{A\} \in F$
    - $G = F - \{X \rightarrow \{A\}\}$
    - Nếu $A \in X_G+$ thì $F = F - \{X \rightarrow \{A\}\}$.

Ví dụ tìm tập PTH tối thiểu

- Tìm phủ tối thiểu của

  $E = \{A \rightarrow BC, A \rightarrow B, B \rightarrow C, AB \rightarrow C\}$

  - *B1*: $F = \varnothing$.
  - *B2*: $F = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, AB \rightarrow C\}$.
  - *B3*: Xét $AB \rightarrow C$
    - $(B)_F+ = C$
    - $F = \{A \rightarrow B, A \rightarrow C, B \rightarrow C\}$.
  - *B4*: $A \rightarrow C$ thừa.
  - $F = \{A \rightarrow B, B \rightarrow C\}$.

Siêu khóa và Khóa

- Cho R(U)
  - $S \subseteq U$ là siêu khóa nếu $\forall r \in R$, $\forall t_1, t_2 \in r$, $t_1 \neq t_2$ thì $t_1[S] \neq t_2[S]$.
  - $K \subseteq U$ là khóa nếu K là siêu khóa nhỏ nhất.
    - $A \in K$ được gọi là thuộc tính khóa.
- Nhận xét
  - S xác định hàm tất cả các thuộc tính của R.
  - R có thể có nhiều khóa.

Xác định khóa của lược đồ

- Input: tập PTH F xác định trên lược đồ R(U).
- Output : khóa K của R.
- Thuật toán
  - *B1*:
    - $K = U = \{A_1, \ldots, A_n\}$
    - $i = 1;$
  - *B2*:
    - Nếu $U \subseteq (K - \{A_i\})_F^+$ thì $K = K - \{A_i\}$.
    - $i = i + 1;$
    - Nếu $i > n$ thì sang *B3*. Ngược lại, tiếp tục *B2*.
  - *B3*:
    - Output K.

Ví dụ tìm khóa của lược đồ

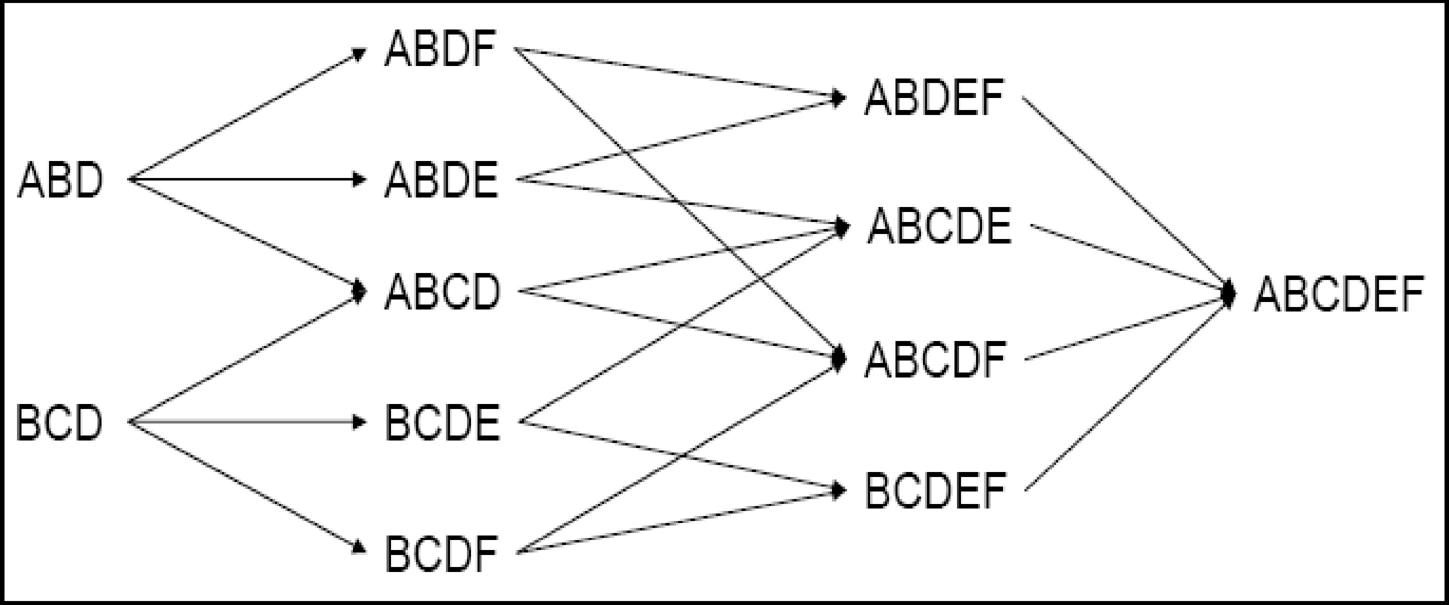
- Cho R(U), U = {A, B, C, D, E, F, G}.
  - F = {B → A, D → C, D → BE, DF → G}.
- Tìm khóa của R
  - B1:
    - K = ABCDEFG.
  - B2:
    - Lặp 1: $(BCDEFG)_F^+$ = BCDEFGA ⇒ K = BCDEFG.
    - Lặp 2: $(CDEFG)_F^+$ = CDEFGBA ⇒ K = CDEFG.
    - Lặp 3: $(DEFG)_F^+$ = DEFGCBA ⇒ K = DEFG.
    - Lặp 4: $(EFG)_F^+$ = EFG.
    - Lặp 5: $(DFG)_F^+$ = DFGCBEA ⇒ K = DFG.
    - Lặp 6: $(DG)_F^+$ = DGCBEA.
    - Lặp 7: $(DF)_F^+$ = DFCBEAG ⇒ K = DF.
  - B3:
    - Khóa là K = DF.

Xác định tất cả khóa của lược đồ

- Input: tập PTH F xác định trên lược đồ R(U).
- Output: tất cả khóa của R.
- Thuật toán
  - *B1*:
    - Xây dựng 2^n tập con của $U = \{A_1, \ldots, A_n\}$
    - $S = \{\}$;
  - *B2*:
    - Với mỗi tập con $X \subseteq U$
    - Nếu $U \subseteq X_F^+$ thì $S = S \cup \{X\}$
  - *B3*:
    - $\forall X, Y \in S$, nếu $X \subset Y$ thì $S = S - \{Y\}$
  - *B4*:
    - S là tập các khóa của R

Ví dụ tìm tất cả khóa của lược đồ

- Cho R(U), U = {A, B, C, D, E, F}.
  - F = {AE → C, CF → A, BD → F, AF → E}.
- Tìm tất cả khóa của R
  - Tập siêu khóa
    - S = {ABD, BCD, ABCD, ABDE, BCDE, ABCDE, ABDF, BCDF, ABCDF, ABDEF, BCDEF, ABCDEF}.



Chuẩn hóa dữ liệu

- Giới thiệu về chuẩn hóa?
- Các dạng chuẩn
  - Dạng 1 (1 Normal Form - 1NF)
  - Dạng 2 (2 Normal Form - 2NF)
  - Dạng 3 (3 Normal Form - 3NF).
  - Dạng Boyce - Codd (Boyce - Codd Normal Form - BCNF)

Các dạng chuẩn

- Dạng 1 (1 Normal Form - 1NF)
- Dạng 2 (2 Normal Form - 2NF)
- Dạng 3 (3 Normal Form - 3NF).
- Dạng Boyce - Codd (Boyce - Codd Normal Form - BCNF)

Dạng chuẩn 1 (1)

- Định nghĩa
  - Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 1 khi và chỉ khi mọi thuộc tính của R là thuộc tính đơn.
- Ví dụ

PHONGBAN

TENPB	<u>MAPB</u>	TrPhong	CacTruso
Hành chính	5	22221	Đống Đa, Hoàng Mai
Nghiên cứu	2	21113	Ba Đình

Không thuộc dạng chuẩn 1

PHONGBAN

TENPB	<u>MAPB</u>	TrPhong	CacTruso
Hành chính	5	22221	Đống Đa
Hành chính	5	22221	Hoàng Mai
Nghiên cứu	2	21113	Ba Đình

Thuộc dạng chuẩn 1

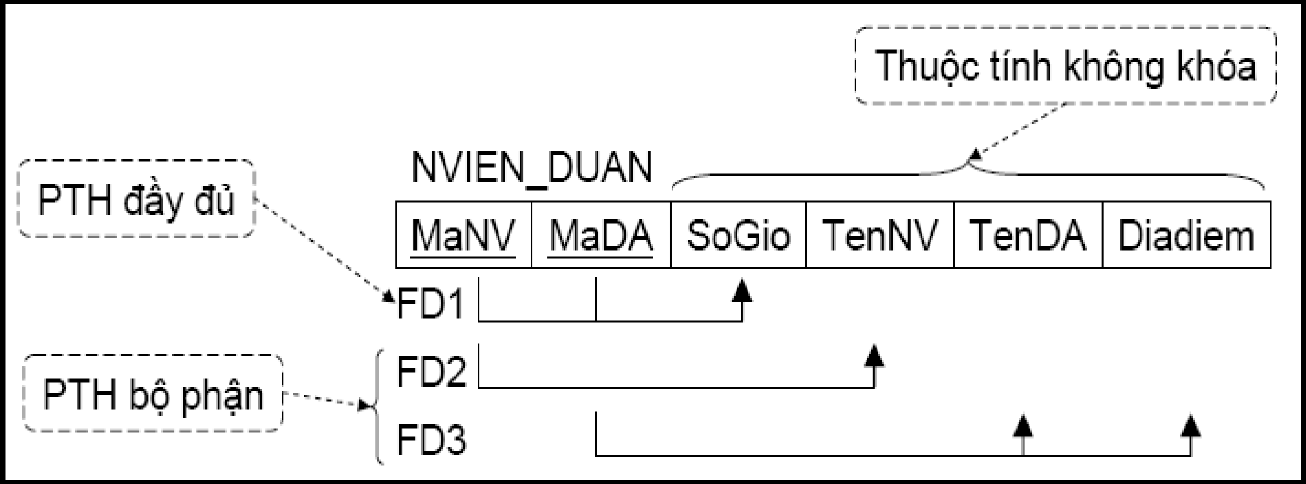
Dạng chuẩn 2 (1)

- Định nghĩa
  - Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 2 khi và chỉ khi:
    - R ở dạng chuẩn 1
    - Mọi thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào khóa chính.
  - R(U), K ⊆ U là khóa chính của R
    - $A \in U$ là thuộc tính không khóa nếu $A \notin K$.
    - $X \rightarrow Y$ là PTH đầy đủ nếu $\forall A \in X$ thì $(X - \{A\}) \rightarrow Y$ không đúng trên R.

      Ngược lại $X \rightarrow Y$ là PTH bộ phận.

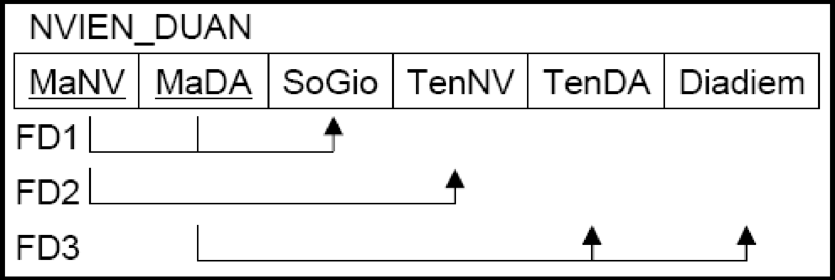
Dạng chuẩn 2 (2)

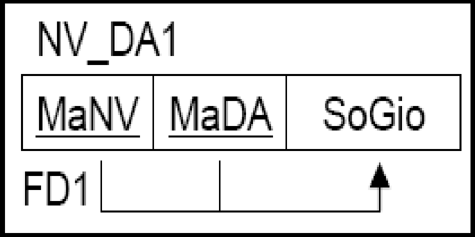
- Ví dụ 1:

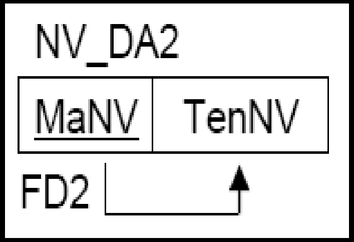


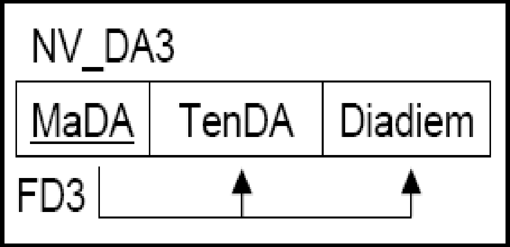
Dạng chuẩn 2 (3)

- Ví dụ 2:







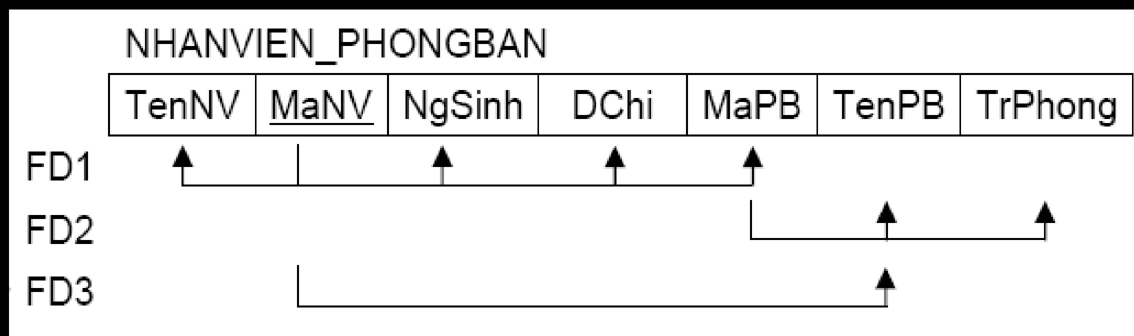


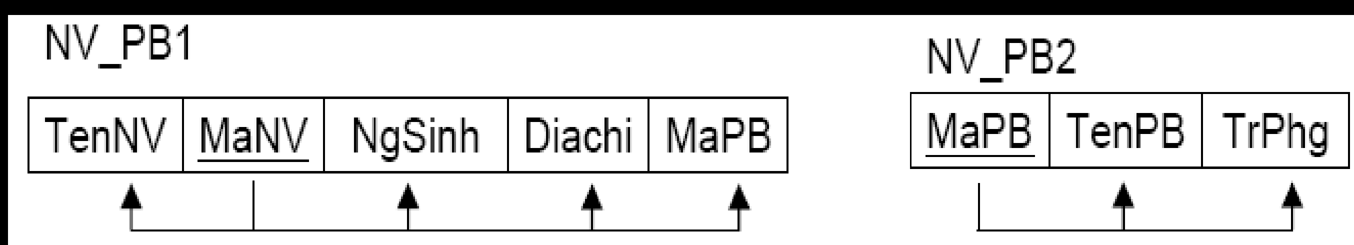
Dạng chuẩn 3 (1)

- Định nghĩa
  - Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 3 khi và chỉ khi:
    - R ở dạng chuẩn 2
    - Mọi thuộc tính không khóa đều không phụ thuộc hàm bắc cầu vào khóa chính.
  - R(U)
    - X → Y là PTH bắc cầu nếu ∃Z ⊆ U, Z không là khóa và cũng không là tập con của khóa của R mà X → Z và Z → Y đúng trên R.

Dạng chuẩn 3 (2)

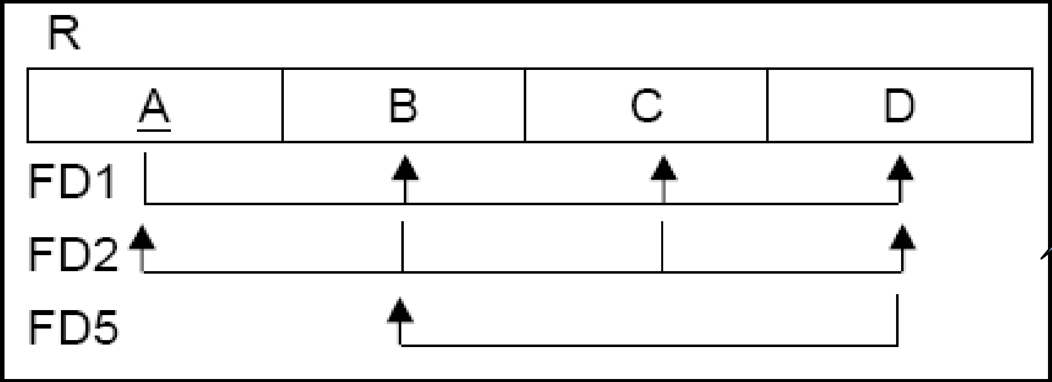
- Ví dụ:
  - FD3 là PTH bắc cầu





Dạng chuẩn Boyce Codd (1)

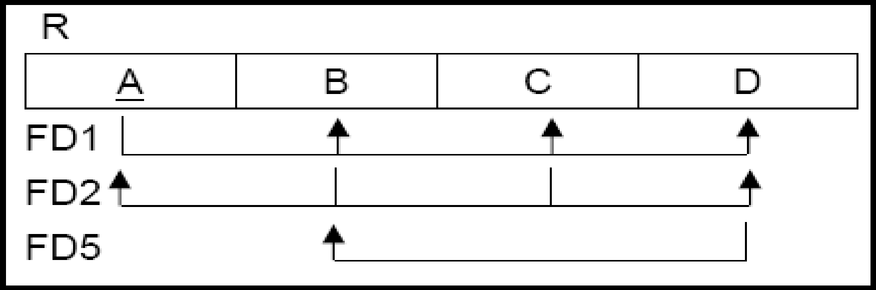
- Định nghĩa
  - Lược đồ quan hệ R được gọi là thuộc dạng chuẩn BCNF khi và chỉ khi:
    - PTH không hiển nhiên X → Y đúng trên R thì X là siêu khóa của R.
- Ví dụ
  - Cho lược đồ quan hệ R(ABCD)



R ở dạng chuẩn nào?

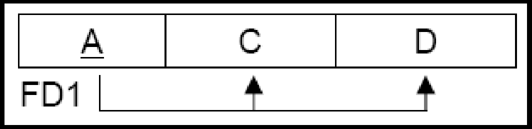
Dạng chuẩn Boyce Codd (2)

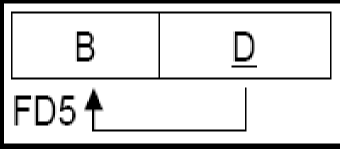
A	B	C	D
1	a	a	1
2	a	b	1
3	b	a	2
4	b	b	2



A	C	D
1	a	1
2	b	1
3	a	2
4	b	2

D	B
1	a
2	b





Dạng chuẩn Boyce Codd (3)

- Nhận xét:
  - Mọi quan hệ thuộc dạng chuẩn BCNF cũng thuộc dạng chuẩn 3
  - Dạng chuẩn BCNF đơn giản và chặt chẽ hơn chuẩn 3
  - Mục tiêu của quá trình chuẩn hóa là đưa lược đồ quan hệ về dạng chuẩn 3 hoặc chuẩn BCNF.