PHỤ THUỘC HÀM VÀ CHUẨN HÓA CSDL

1

NỘI DUNG

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

2

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 $\subseteq$ U
  - Ký hiệu: X $\rightarrow$ 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 $\rightarrow$ Y được gọi là PTH hiển nhiên nếu Y $\subseteq$ X
- X $\rightarrow$ Y được gọi là PTH nguyên tố (Y PTH đầy đủ vào X) nếu $\forall$X'$\subset$ X thì X' không $\rightarrow$ Y

3

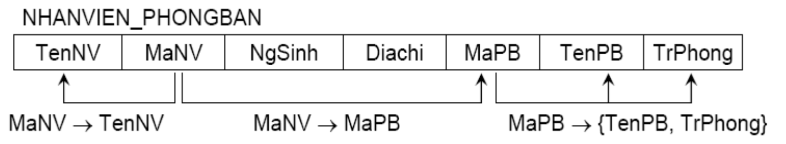
- Hóa đơn (<u>số hóa đơn</u>, mã nhân viên, Mã khách hang, ngày hóa đơn)
- Chi tiết hóa đơn (<u>số hóa đơn, mã hàng</u>, số lượng)

- Số hóa đơn-> mã nhân viên
- Số hóa đơn -> khách hang
- (số hóa đơn,mã hàng)-> số lượng

Số hóa đơn	Mã hàng	Số lượng	Đơn giá
01	H1	3	10000
01	H2	1	15000
01	H3	2	20000
02	H2	6	15000
02	H3	2	20000

4

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: $X \rightarrow Y, WY \rightarrow Z \Rightarrow WX \rightarrow Z$.

7

- 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: $X \rightarrow Y, WY \rightarrow Z \Rightarrow WX \rightarrow Z$.
- Ví dụ 2: Cho $F=\{AB \rightarrow E, AG \rightarrow I, BE \rightarrow I, E \rightarrow G, GI \rightarrow H\}$
  - Hãy chứng tỏ PTH $AB \rightarrow GH$ suy diễn từ F nhờ luật dẫn Armstrong

8

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 $K^+ = U$.

9

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^+

10

VÍ DỤ TÌM X^+

- Input:
  - F = {AB → C, BC → D, D → EG}
  - X = BD
- Output: X^+
- Thuật toán
  - X^+ = BD.
  - Lặp 1:
    - Tìm các PTH có vế trái là tập con của X^+ = BD
      - D → 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.

11

VÍ DỤ TÌM X^+

X^+ = AC
Ta có AC → D nên X^+ = ACD
Lại có D → H nên X^+ = ACDH
Lại có DA → CE nên X^+ = ACDHE

- VD2: Cho lược đồ quan hệ Q(ABCDEGH) và tập PTH F
  - F = { B → A, DA → CE, D → H, GH → C, AC → 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 → C,A → EG, B → D, G → E}
  - Xác định X^+
    - X= {AB}
    - X={CGD}

1. X^+ = AB
Ta có A → C nên X^+ = ABC
Lại có A → EG nên X^+ = ABCEG
Lại có B → D nên X^+ = ABCEGD

2. X^+ = CGD
Ta có G → E nên X^+ = CGDE

12

VÍ DỤ

- VD4: Cho lược đồ quan hệ R(U,F), U = ABCDEG
  F = { AB → C; C → A; BC → D; D → EG; CG → BD; ACD → B; CD → AG}.
  a) Tính $(CD)^+$
- b) Tính $(AB)^+$
- c) Tính $(BC)^+$
- VD5: F = { AC → D; ABD → C; D → A; D → EG; DG → BC; CD → B; CE → D; DE → AG}
  a) Tính (AD)+
  b) Tính (DG)+
- c) Tính (CD)+

13

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 $F^+ = G^+$

14

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?

1. X^+ = AB
Ta có AB → C nên X^+ = ABC
Lại có A → D nên X^+ = ABCD
Lại có D → E nên X^+ = ABCDE

2. X^+ = D
Ta có D → E nên X^+ = DE

15

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).

16

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.

17

THUẬT TOÁN TÌM TẬP PTH TỐI THIỂU

- Input: tập PTH E.
- Output: phủ tối thiểu F của E.
- Thuật toán:
  - *B1*: $F = \emptyset$
  - *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 $F = F - \{X \rightarrow \{A\}\}$.

18

VÍ DỤ TÌM TẬP PTH TỐI THIỂU

- Tìm phủ tối thiểu của
  E = {A → BC, A → B, B → C, AB → C}

F=rỗng
Tách vế phải:
E = {A → B, A → C, B → C, AB → C}
Loại bỏ dư thừa thuộc tính:
  Xét AB → C: loại A: tính B+=BC chứa C, loại được A
    E = {A → B, A → C, B → C}
Loại bỏ dư thừa phụ thuộc hàm:
  Xét A → B: A+=AC không chứa B, không loại được
  Xét A → C: A+=ABC chứa C, loại được
    E = {A → B, B → C}
  Xét B→ C: B+=B không chứa C, không loại được
Vậy phủ tối thiểu: F = {A → B, B → C}

19

TÌM TẬP PTH TỐI THIỂU

- VD1: Cho lược đồ quan hệ R(U,F), U = ABCDEG
- F = { B→EC; BC→AE; CD→AB; C→AD; AC→BD}

  F = { B→C; B→E; C→B; C→A; C→D}

- VD2: Cho lược đồ quan hệ R(U,F), U = ABCDEG
  F = { AB → C; C → A; BC → D; D → EG; CG → BD; ACD → B; CD → AG}.

20

TÌM TẬP PTH TỐI THIỂU

- VD1: Cho lược đồ quan hệ R(U,F), U = ABCDEG
- F = { B→EC; BC→AE; CD→AB; C→AD; AC→BD}

Tách vế phải
F={ B→E; B → C; BC→A; BC → E; CD→A; CD → B; C→A; C → D; AC→B; AC → D}
Loại bỏ dư thừa thuộc tính:
Xét: BC → A
Xét C+ chứa A, loại B: C → A
F={ B→E; B → C; C→A; BC → E; CD→A; CD → B; C→A; C → D; AC→B; AC → D}
Xét: BC → E
Xét C+ =CABDE, loại được B: C → E
F={ B→E; B → C; C→A; C → E; CD→A; CD → B; C→A; C → D; AC→B; AC → D}
Xét: CD → A
Xét D+ =D, không loại được C
Xét C+=CABD loại được D: C → A
F={ B→E; B → C; C→A; C → E; CD → B; C→A; C → D; AC→B; AC → D}
Xét: CD → B
Xét D+ =D không loại được C
Xét C+=CAEDB chứa B, D, loại được D: C → B
F={ B→E; B → C; C→A; C → E; C → B; C → D; AC→B; AC → D}
Xét: AC → B
Xét C+ =CEABD chứa B, loại được A: C → B
F={ B→E; B → C; C→A; C → E; C → B; C → D; AC → D}
Xét: AC → D
Xét C+ =CEABD chứa D, loại được A: C → D

21

TÌM TẬP PTH TỐI THIỂU

- VD1: Cho lược đồ quan hệ R(U,F), U = ABCDEG
- F = { B→EC; BC→AE; CD→AB; C→AD; AC→BD}
- VD2: Cho lược đồ quan hệ R(U,F), U = ABCDEG
  F = { AB → C; C → A; BC → D; D → EG; CG → BD; ACD → B; CD → AG}.

AB → C; C → A; BC → D; D → G; D → E; CG → D; CD → B;

22

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.

23

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.

24

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=U ⇒ K = BCDEFG.
    - Lặp 2: $(CDEFG)^+_F$ = CDEFGBA=U ⇒ K = CDEFG.
    - Lặp 3: $(DEFG)^+_F$ = DEFGCBA =U⇒ K = DEFG.
    - Lặp 4: $(EFG)^+_F$ = EFG khác U, không loại được D.
    - Lặp 5: $(DFG)^+_F$ = DFGCBEA=U ⇒ K = DFG.
    - Lặp 6: $(DG)^+_F$ = DGCBEA, khác U, không loại được F.
    - Lặp 7: $(DF)^+_F$ = DFCBEAG=U ⇒ K = DF.
  - B3:
    - Khóa là K = DF.

Bài số 1:

Cho lược đồ quan hệ: α=(U,F)
V ới U=ABCDEGH
F={AB → CDE, AC → BCG, BD→G, ACH→HE, CG → BDE }
và K = ACGH

Hỏi rằng K có là khoá của lược đồ hay không?

K+=ACGHBED=U
K LÀ SIÊU KHÓA
(K-{A})+=CGH+=CGHBDE <>U
(K-{C})+=AGH+=AGH<>U
(K-{G})+=ACH+=ACHBGED=U
(K-{G}) CŨNG LÀ SIÊU KHÓA
SUY RA K KHÔNG LÀ KHÓA

Cho lược đồ quan hệ (=(U, F) với
U=ABCDEGH
F={ ABC→DE, BCD→G, ABH→EG, CE→GH}. Hãy tìm một khóa của lược đồ

K=ABCDEGH
K=K-{A}; K+=BCDEGH≠U→K=ABCDEGH
K=K-{B}; K+=ACDEGH≠U→K=ABCDEGH
K=K-{C}; K+=ABDEGH≠U→K=ABCDEGH
K=K-{D}; K+=ABCDEGH=U→K=ABCEGH
K=K-{E}; K+=ABCDEGH=U→K=ABCGH
K=K-{G}; K+=ABCDEGH=U→K=ABCH
K=K-{H}; K+=ABCDEGH=U→K=ABC

27

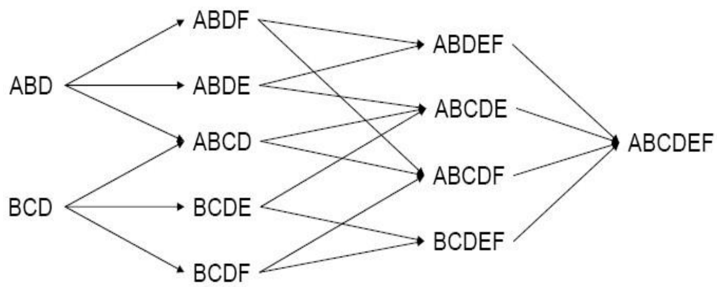
XÁC ĐỊNH TẤT CẢ KHÓA CỦA LƯỢC ĐỒ

- Input: tập PTH F xác định trên lược đồ R(U).
- Output: tất cả khóa của R.
- Thuật toán
  - *B1*:
    - Xây dựng 2^n-1 tập con của U = {A_1, …, A_n}
    - S = {};
  - *B2*:
    - Với mỗi tập con X ⊆ U
    - Nếu U ⊆ X_F^+ thì S = S ∪{X}
  - *B3*:
    - ∀X, Y ∈S, nếu X ⊂ Y thì S = S - {Y}
  - *B4*:
    - S là tập các khóa của R

28

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}.



29

F = {AE → C, CF → A, BD → F, AF → E}.

A	BF	ADE	ABCE	ABCDE
B	CD	ADF	ABCF	ABCDF
C	CE	AEF	ABDE	ABCEF
D	CF	**BCD**	ABDF	BCDEF
E	DE	BCE	ABEF	ACDEF
F	DF	BCF	ACDE	ABDEF
AB	EF	BDE	ACDF	ABCDEF
AC	ABC	BDF	ACEF	
AD	**ABD**	BEF	ADEF	
AE	ABE	CDE	BCDE	
AF	ABF	CDF	BCDF	
BC	ACD	CEF	BDEF	
BD	ACE	DEF	CDEF	
BE	ACF	ABCD	BCEF	

30

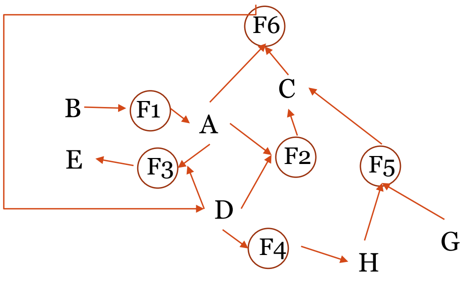
THUẬT TOÁN TÌM MỘT KHÓA BẰNG ĐỒ THỊ CÓ HƯỚNG

- Biểu diễn lược đồ quan hệ R (U, F) bằng đồ thị có hướng như sau:
- \- Mỗi nút của đồ thị là tên một thuộc tính của R
- \- Cung nối hai thuộc tính A và B thể hiện phụ thuộc hàm A->B
- Thuộc tính chỉ có mũi tên đi ra gọi là nút gốc
- Thuộc tính chỉ có mũi tên đi tới gọi là nút lá
- Khi đó: Khóa của lược đồ bao phủ tập các nút gốc mà không chứa bất kỳ nút lá nào
- **Thuật toán**
- B1: Xuất phát từ tập nút gốc X
- B2: Tính bao đóng của X: X^+
- B3: Nếu X^+=U thì X là khóa, ngược lại thì bổ sung một thuộc tính không thuộc nút lá vào X, rồi tìm bao đóng. Quay lại bước 2

31

VÍ DỤ

- Cho lược đồ quan hệ R(A,B,C,D,E,G,H) và tập phụ thuộc hàm:
- F = {B→A; DA→CE; D→H; GH→C; AC→D}
- Tìm một khóa?
- Phân rã F: F = {B→A; DA→C; DA→E; D→H; GH→C; AC→D}



32

TÌM TẤT CẢ CÁC KHÓA BẰNG ĐỒ THỊ CÓ HƯỚNG

- - Tập thuộc tính nguồn (TN): bao gồm các thuộc tính chỉ xuất hiện ở vế trái, không xuất hiện ở vế phải của pth và các thuộc tính không xuất hiện ở vế trái lẫn vế phải của pth.
  - Tập thuộc tính đích (TĐ) : bao gồm các thuộc tính chỉ xuất hiện ở vế phải không xuất hiện ở vế trái của pth.
  - Tập thuộc tính trung gian (TG): Chứa thuộc tính ở vế trái lẫn vế phải của pth.

Cho lược đồ quan hệ R(A,B,C,D,E,G,H) và tập phụ thuộc hàm:

F = {B→A; DA→CE; D→H; GH→C; AC→D}

Trái={B, D, A, G, H, C} ; Phải={A, C, E, H, D};

Nguồn: {B, G}; Trung gian: {A, C, D, H}

33

TÌM TẤT CẢ CÁC KHÓA BẰNG ĐỒ THỊ CÓ HƯỚNG

- Thuật toán:
  Bước 1: Tạo tập nguồn TN và tập trung gian TG
  Bước 2: Nếu TG=ø thì K=TN, kết thúc. ngược lại qua bước 3.
- Bước 3: tính TN+, nếu TN+=U, K=TN, kết thúc, ngược lại qua bước 4.
  Bước 4: tìm tất cả tập con Xi của tập trung gian.
  Bước 5: tìm siêu khóa Si bằng cách với mọi Xi, nếu (TN U Xi)+=Q+ thì Si = TN U Xi
  Bước 6: tìm khóa bằng cách loại bỏ các siêu khóa không tối thiểu - với mọi Si, Sj thuộc S nếu Si chứa trong Sj thì loại bỏ tập Sj ra khỏi siêu khóa (VD: Si=AB, Sj=ABC thì loại bỏ Sj ra khỏi tập siêu khóa)
  S còn lại chính là tập khóa cần tìm.

Cho lược đồ quan hệ R(A,B,C,D,E,G,H) và tập phụ thuộc hàm:

F = {B→A; DA→CE; D→H; GH→C; AC→D}

Trái={B, D, A, G, H, C} ; Phải={A, C, E, H, D};

Nguồn: {B, G}; Trung gian: {A, C, D, H}

TGi={A, C, D, H, AC, AD,...}

Xét Ki=BG ∪ TGi, Ki+=U

34

TẬP NGUỒN: BG
TẬP TRUNG GIAN: A, C, D, H
F = {B→A; DA→CE; D→H; GH→C; AC→D}

BGA	BGCDH			
BGC	BGACDH			
BGD				
BGH				
BGAC				
BGAD				
BGAH				
BGCD				
BGCH				
BGDH				
BGACD				
BGACH				
BGADH				

35

Cho lược đồ quan hệ α = (U, F) với
U=ABCDEGH, F={ ABC→ADH, ABG→AEH, AE→DG}
Hãy tìm tất cả các khoá của lược đồ

- Trái: {A, B, C, G,E}
- Phải: {A,D, H, E, G}
- TN: BC
- TG: AGE

BCA	BCG	BCE	BCAG
BCAE	BCGE	BCAGE	

F = {A→G; DG→CE; D→H; BH→C; GC→D}

36

F = {A→G; DG→CE; D→H; BH→C; GC→D}

Trái: A,D,G,B,H,C
PhảI: G, C, E, H, D
TN: A, B
TG: C, D, G, H

ABC	ABD	ABG	ABH	ABCD
ABCG	Abch	ABDG	ABDH	ABGH
ABCDG	ABCGH	ABDGH	ABCDH	ABCDGH

37

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)

38

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)

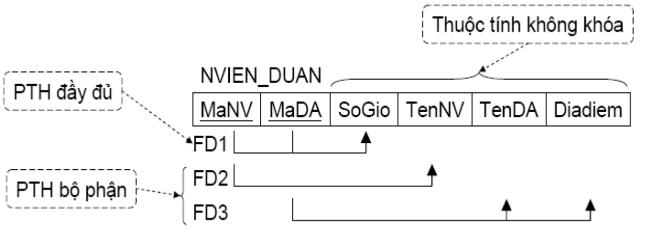
- Định nghĩa
  - Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 2 khi và chỉ khi:
    - R ở dạng chuẩn 1
    - Mọi thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào khóa chính.
  - R(U), K $\subseteq$ U là khóa chính của R
    - A $\in$ U là thuộc tính không khóa nếu A $\notin$ K.
    - X $\rightarrow$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

41

DẠNG CHUẨN 2 (2)

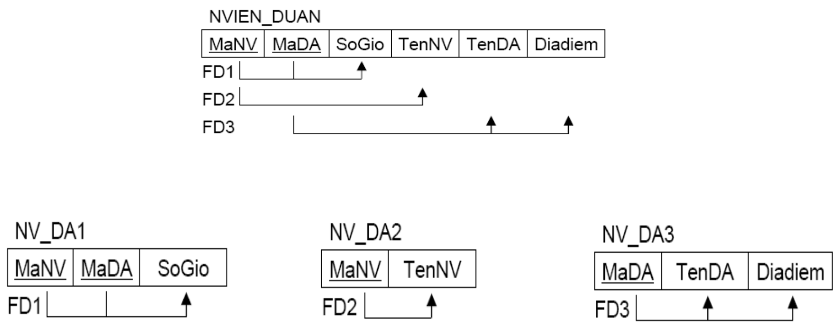
- Ví dụ 1:



42

DẠNG CHUẨN 2 (3)

- Ví dụ 2:



43

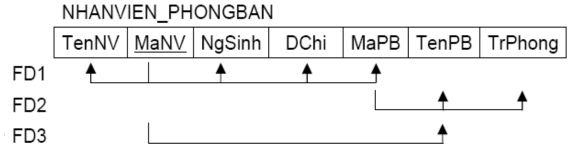
DẠNG CHUẨN 3 (1)

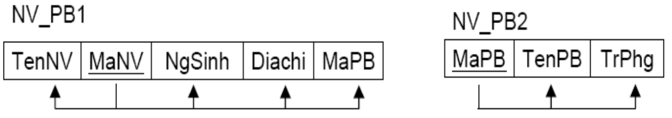
- Định nghĩa
  - Lược đồ quan hệ R được gọi là thuộc dạng chuẩn 3 khi và chỉ khi:
    - R ở dạng chuẩn 2
    - Mọi thuộc tính không khóa đều không phụ thuộc hàm bắc cầu vào khóa chính.
  - R(U)
    - $X \rightarrow Y$ là PTH bắc cầu nếu $\exists Z \subseteq U$, Z không là khóa và cũng không là tập con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

44

DẠNG CHUẨN 3 (2)

- Ví dụ:
  - FD3 là PTH bắc cầu

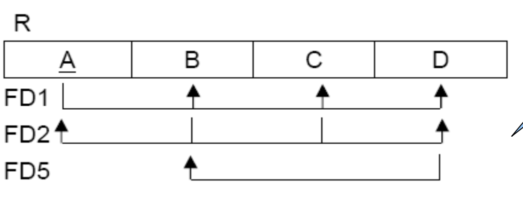




DẠNG CHUẨN BOYCE CODD (1)

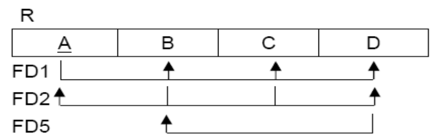
- Định nghĩa
  - Lược đồ quan hệ R được gọi là thuộc dạng chuẩn BCNF khi và chỉ khi:
    - PTH không hiển nhiên X → Y đúng trên R thì X là siêu khóa của R.
- Ví dụ
  - Cho lược đồ quan hệ R(ABCD)

R ở dạng chuẩn nào?



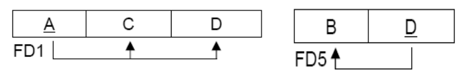
DẠNG CHUẨN BOYCE CODD (2)

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

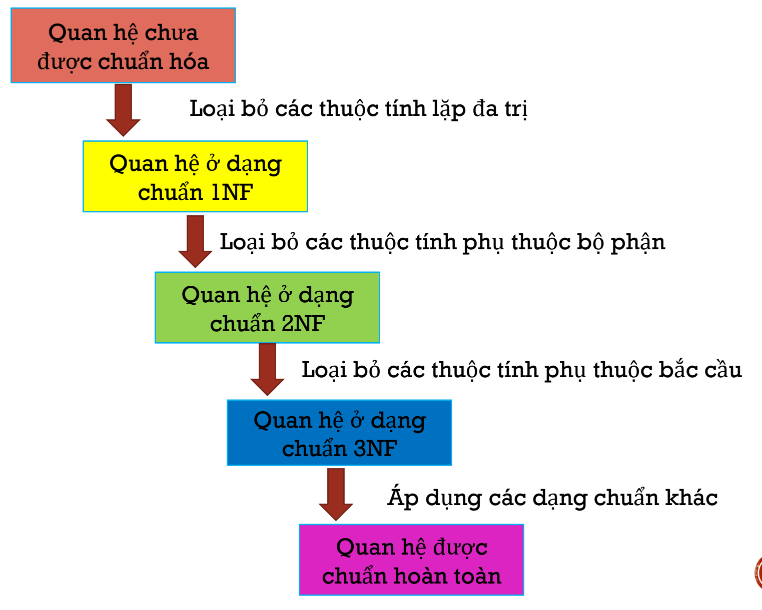
D	B
1	a
2	b



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.

CÁC DẠNG CHUẨN



49

MỤC ĐÍCH CỦA CHUẨN HÓA DỮ LIỆU

- Xác định được 1 tập các lược đồ quan hệ cho phép tìm kiếm thông tin một cách dễ dàng, đồng thời tránh được dư thừa dữ liệu

- Giải pháp:

Tách các lược đồ quan hệ "có vấn đề" thành những lược đồ quan hệ "chuẩn hơn"

50

PHÉP TÁCH CÁC LƯỢC ĐỒ QUAN HỆ

- **Mục đích**
  Thay thế một sơ đồ quan hệ $R(A_1, A_2, \ldots, A_n)$ bằng một tập các sơ đồ con $\{R_1, R_2, \ldots, R_k\}$ trong đó $R_i \subseteq R$ và $R = R_1 \cup R_2 \cup \ldots \cup R_k$

- **Yêu cầu của phép tách**
  - Bảo toàn thuộc tính, ràng buộc
  - Bảo toàn dữ liệu

51

PHÉP TÁCH KHÔNG MẤT MÁT THÔNG TIN (LOSSLESS JOIN)

- **Định nghĩa**: Cho lược đồ quan hệ R(U) phép tách R thành các sơ đồ con $\{R_1, R_2, \ldots, R_k\}$ được gọi là phép tách không mất mát thông tin đối với một tập phụ thuộc hàm F nếu với mọi quan hệ r xác định trên R thỏa mãn F thì:

$$r = \Pi_{R1}(r) \bowtie \Pi_{R2}(r) \bowtie \ldots \bowtie \Pi_{Rk}(r)$$

- **Ví dụ**:
  Supplier(sid, sname, pname,colour,quantity)

  S1(sid, sname, city)
  SP1(sid,pname,colour,quantity)

52

KIỂM TRA TÍNH KHÔNG MẤT MÁT THÔNG TIN

Vào: $R(A_1, A_2, \ldots, A_n)$, F, phép tách $\{R_1, R_2, \ldots, R_k\}$
Ra: phép tách là mất mát thông tin hay không
Thuật toán
B.1. Thiết lập một bảng k hàng, n cột
Nếu A_j là thuộc tính của R_i thì điền a_j vào ô (i,j).
Nếu không thì điền b_{ij}.
B.i. Xét $f = X \rightarrow Y \in F$.
Nếu $\exists$ 2 hàng t1, t2 thuộc bảng : t1[X] = t2[X]
thì t1[Y] = t2[Y], ưu tiên đồng nhất về giá trị a
Lặp cho tới khi không thể thay đổi được giá trị nào trong bảng
B.n. Nếu bảng có 1 hàng gồm các kí hiệu $a_1, a_2, \ldots, a_n$
thì phép tách là không mất mát thông tin.
ngược lại, phép tách không bảo toàn thông tin.

53

VÍ DỤ

- R(MONHOC, SOTIET, LOP, GV, HOCVI, DC)
- Kiểm tra: R_1(MONHOC, SOTIET, LOP, GV), R_2(GV, HOCVI, DC)

F = {MONHOC → SOTIET; MONHOC, LOP → GV; GV → HOCVI, DC}

	MONHOC	SOTIET	LOP	GV	HOCVI	DC
R_1	a_1	a_2	a_3	a_4	b_{15}	b_{16}
R_2	b_{21}	b_{22}	b_{23}	a_4	a_5	a_6

GV → HOCVI, DC

	MONHOC	SOTIET	LOP	GV	HOCVI	DC
R_1	a_1	a_2	a_3	a_4	a_5	a_6
R_2	b_{21}	b_{22}	b_{23}	a_4	a_5	a_6

54

BÀI SỐ 1: KIỂM TRA PHÉP TÁCH CÓ MẤT THÔNG TIN HAY KHÔNG?
CHO LƯỢC ĐỒ QUAN HỆ α= (U, F) VỚI
$U=\{A_1, A_2, A_3, A_4, A_5\}$
$F=\{ A_1 \rightarrow A_2 A_3 , A_2 A_4 \rightarrow A_5 , A_2 \rightarrow A_3\}$
$\delta=\{ A_1 A_2 A_4; A_2 A_3; A_1 A_4 A_5\}$

	A1	A2	A3	A4	A5
R1	a1	a2	*	a4	*
R2	*	a2	a3	*	*
R3	a1	*	*	a4	a5

	A	B	C	D	E	G
R1		a2	a3			
R2	a1	a2	a3			
R3	a1	a2	a3	a4	a5	
R4	a1	a2	a3	a4	a5	a6

Bài 3: α =(U, F) với U=(ABCDEGH), F={CD→H, E→B, D→G, BH→E, CH→DG, C→A}
δ ={ABCDE, BCH, CDEGH}

Bài 4: α =(U, F) với U=(ABCDEGH), F={CD→H, E→B, D→G, BH→E, CH→DG, C→A}
δ ={ABDE, BCH, DEGH}

	A	B	C	D	E	G	H
R1	a1	a2		a4	a5		
R2		a2	a3				a7
R3				a4	a5	a6	a7

PHÉP TÁCH BẢO TOÀN TẬP PHỤ THUỘC HÀM

Hình chiếu của tập phụ thuộc hàm

Cho sơ đồ quan hệ R, tập phụ thuộc hàm F, phép tách $\{R_1, R_2, \ldots, R_k\}$ của R trên F.

Hình chiếu F_i của F trên R_i là tập tất cả $X \rightarrow Y \in F+$:

$$XY \subseteq R_i .$$

Phép tách sơ đồ quan hệ R thành $\{R_1, R_2, \ldots, R_k\}$ là một phép tách bảo toàn tập phụ thuộc hàm F nếu

$$(F_1 \cup F_2 \ldots \cup F_k)+ = F+$$

hay hợp của tất cả các phụ thuộc hàm trong các hình chiếu của F lên các sơ đồ con sẽ suy diễn ra các phụ thuộc hàm trong F.

57

TÁCH BẢO TOÀN TẬP PHỤ THUỘC HÀM VỀ 3NF

Vào: R(U), F (giả thiết F là phủ tối thiểu)

Ra: Phép tách bảo toàn tập phụ thuộc hàm về 3NF

Thuật toán

B1. Với các $A_i \in U$, $A_i \notin F$ thì loại A_i khỏi R và lập 1 quan hệ mới cho các A_i

B2. Nếu $\exists f \in F$, f chứa tất cả các thuộc tính của R thì kết quả là R

B3. Ngược lại, với mỗi $X \rightarrow A \in F$, xác định một quan hệ $R_i(XA)$.

Nếu $\exists X \rightarrow A_i$, $X \rightarrow A_j$ thì tạo một quan hệ chung $R'(XA_iA_j)$

58

VÍ DỤ

Cho R = {A,B,C,D,E,F,G}

F = {A→B, ACD→E, EF→G}

- Xác định phép tách bảo toàn tập phụ thuộc hàm về 3NF

  B1. không lập được quan hệ nào mới.

  B2. !∃ f ∈ F: f chứa tất cả các thuộc tính của R

  B3. A→B ⇨ R1(AB)

  ACD→E ⇨ R2(ACDE)

  EF→G ⇨ R3(EFG)

59

TÁCH KHÔNG MẤT MÁT THÔNG TIN VÀ BẢO TOÀN TẬP PHỤ THUỘC HÀM VỀ 3NF

- **Yêu cầu:**
  - Bảo toàn tập phụ thuộc hàm (như thuật toán trên)
  - Đảm bảo là có một lược đồ con chứa khóa của lược đồ được tách
- **Các bước tiến hành**
  - B1. Tìm một khóa tối thiểu của lược đồ quan hệ R đã cho
  - B2. Tách lược đồ quan hệ R theo phép tách bảo toàn tập phụ thuộc hàm
  - B3. Nếu 1 trong các sơ đồ con có chứa khóa tối thiểu thì kết quả của B2 là kết quả cuối cùng.
    Ngược lại, thêm vào kết quả đó một sơ đồ quan hệ được tạo bởi khóa tối thiểu tìm được ở 1.

60

VÍ DỤ

- Cho R(A,B,C,D,E,F,G).
  F = {A–>B, ACD->E, EF->G}

B1. Khóa tối thiểu cần tìm là ACDF
B2. Phép tách bảo toàn tập phụ thuộc hàm R cho 3 sơ đồ con
R_1(AB), R_2(ACDE), R_3(EFG)
B3. Do khóa ACDF không nằm trong bất kỳ một sơ đồ con nào trong 3 sơ đồ con trên, ta lập một sơ đồ con mới R_4(ACDF)

Kết quả cuối cùng ta có phép tách R thành 4 sơ đồ con {R_1, R_2, R_3, R_4} là một phép tách không mất mát thông tin và bảo toàn tập phụ thuộc hàm

61

BÀI TẬP

- Cho lược đồ quan hệ Q(A,B,C,D,E,G,H) và tập phụ thuộc hàm
  F = { E → C;H → E; A→ D;AE → H; DG → B;DG → C }
  a. Hãy xác định tất cả các khóa của Q
  b. Hãy xác định dạng chuẩn cao nhất của Q
  c. Phân rã Q về dạng chuẩn 3, yêu cầu phân rã bảo toàn thông tin và phụ thuộc hàm.

62

TÁCH KHÔNG MẤT MÁT THÔNG TIN VỀ BCNF

Vào: Sơ đồ quan hệ R, tập phụ thuộc hàm F.

Ra: phép tách không mất mát thông tin bao gồm một tập các sơ đồ con ở BCNF với các phụ thuộc hàm là hình chiếu của F lên sơ đồ đó.

Cách tiến hành

B1. KQ = {R},

B2. Với mỗi S ∈ KQ, S không ở BCNF, xét X→A ∈ S, với điều kiện X không chứa khóa của S và A ∉ X. Thay thế S bởi S1, S2 với S1=A ∪{X}, S2 = {S} \ A.

B3. Lặp (B2) cho đến khi ∀S ∈KQ đều ở BCNF

KQ gồm các sơ đồ con của phép tách yêu cầu

63

Cho lược đồ α = (U, F) với

U=CRHTSG (C : Course, T : Teacher, H Hour, R : Room, S : Student, G : Group)

F ={C→T , HR → C, CH → R, CS→ G, HS→ R}

VD1: U = ABCDEG

F = { A → B, BE → C, EC → A, AD → G, ED → C}

VD2: U = A,B,C,D,E,F,G,H,I,J

F1 = {AB → C, A → DE, B → F, F → GH, D → IJ}

64

VD2: U = A,B,C,D,E,F,G,H,I,J
F1 = {AB → C, A → DE, B → F, F → GH, D → IJ}

BƯỚC 1: TÌM PHỦ TỐI THIỂU
F1 = {AB → C, A → D, A → E, B → F, F → G, F → H, D → I, D → J}
LOẠI BỎ DƯ THỪA THUỘC TÍNH:
XÉT AB → C: B+=BFGH, KHÔNG CHỨA A, C, KHÔNG LOẠI ĐƯỢC
A+=ADEIJ, KHÔNG CHỨA B, C, KHÔNG LOẠI ĐƯỢC
LOẠI BỎ DƯ THỪA PHỤ THUỘC HÀM
XÉT AB → C: AB+=ABDEFGHIJ KHÔNG CHỨA C, KHÔNG DƯ THỪA
XÉT A → D: A+=AEFGH, KHÔNG CHỨA D, KHÔNG DƯ THỪA
XÉT A → E: A+=ADGHIJ KHÔNG CHỨA E, KHÔNG DƯ THỪA
...
VẬY PHỦ TỐI THIỂU LÀ: F = {AB → C, A → D, A → E, B → F, F → G, F → H, D → I, D → J}
BƯỚC 2: TÌM TẤT CẢ CÁC KHÓA
TRÁI: ABFD; PHẢI: CDEFGIJ; TRUNG GIAN: DF; NGUỒN: AB
ABD+=ABDCEFGHIJ=U,
ABF=ABFCDEGHIJ=U
VẬY TẤT CẢ KHÓA LÀ ABD, ABF
BƯỚC 3: CÓ A → E, A LÀ BỘ PHẬN CỦA KHÓA, E LÀ THUỘC TÍNH KHÔNG KHÓA, VI PHẠM CHUẨN 2, (U, F) LÀ CHUẨN 1
BƯỚC 4: TÁCH

AB → C: R1(ABC)	A → D, A → E: R2(ADE)	B → F: R3(BF)
F → G, F → H: R4(FGH)	D → I, D → J: R5(DIJ)	R6(ABD)

Cho lược đồ α = (U, F) với
U=CRHTSG (C : Course, T : Teacher, H Hour, R : Room, S : Student, G : Group)
F ={C→T , HR → C, CH → R, CS→ G, HS→ R}

- C →T, HR → C, CH → R, CS → G, HS → R
- U=CRHTSG; Khóa: HS
- R1(CT); C →T,
- R_cl(CRHSG)
- R2(CSG); CS → G,
- R_cl(CRHS)
- R3(HRC); HR → C; CH → R
- R4(RHS): HS → R

VD2: U = A,B,C,D,E,F,G,H,I,J
F1 = {AB → C, A → DE, B → F, F → GH, D → IJ}

VD1: U = ABCDEG
F = { A → B, BE → C, EC → A, AD → G, ED → C}

U = ABCDEG
F = { A → B, BE → C, EC → A, AD → G, ED → C}

- Khóa: ED
- R1(ADG); AD → G, R_cl(ABCDE)
- R2(AB); A → B, R_cl(ACDE)
- R3(ECA); EC → A; R_cl(CDE)
- R4(CDE); ED → C

- TÁCH VỀ 3NF TRƯỚC
- CÁI NÀO KHÔNG BCNF, TÁCH VỀ BCNF

67

VD2: U = A,B,C,D,E,F,G,H,I,J
F1 = {AB → C, A → DE, B → F, F → GH, D → IJ}

- F1 = {AB → C, A → D, A → E, B → F, F → G, F → H, D → I, D → J}
- R(ABCDEFGHIJ)
- R1(AE), A → E, R_cl(ABCDFGHIJ)
- R2(DI), D → I, R_cl(ABCDFGHJ)
- R3(DJ), D → J, R_cl(ABCDFGH)
- R4(AD), A → D, R_cl(ABCFGH)
- R5(FG), F → G, R_cl(ABCFH)
- R6(FH), F → H, R_cl(ABCF)
- R7(BF), B → F, R_cl(ABC)
- R8(ABC), AB → C

68

VD2: U = A,B,C,D,E,F,G,H,I,J
F1 = {AB → C, A → D, A → E, B → F, F → G, F → H, D → I, D → J}

- Đưa về phủ tối thiểu
- Tìm tất cả các khóa: AB
- Tách về 3NF: R1(ABC), R2(AED), R3(BF), R4(FGH), R5(DIJ)
- Tách về BCNF:
  - R(ABCDEFGHIJ)
  - Xét A → D: R1(AD); R_cl(ABCEFGHIJ)
  - Xét A → E: R1(AE); R_cl(ABCFGHIJ)

69