

Chương 6

Chuẩn hoá sơ đồ quan hệ

Phạm Thị Ngọc Diễm
Bộ môn HTTT - ĐHCT

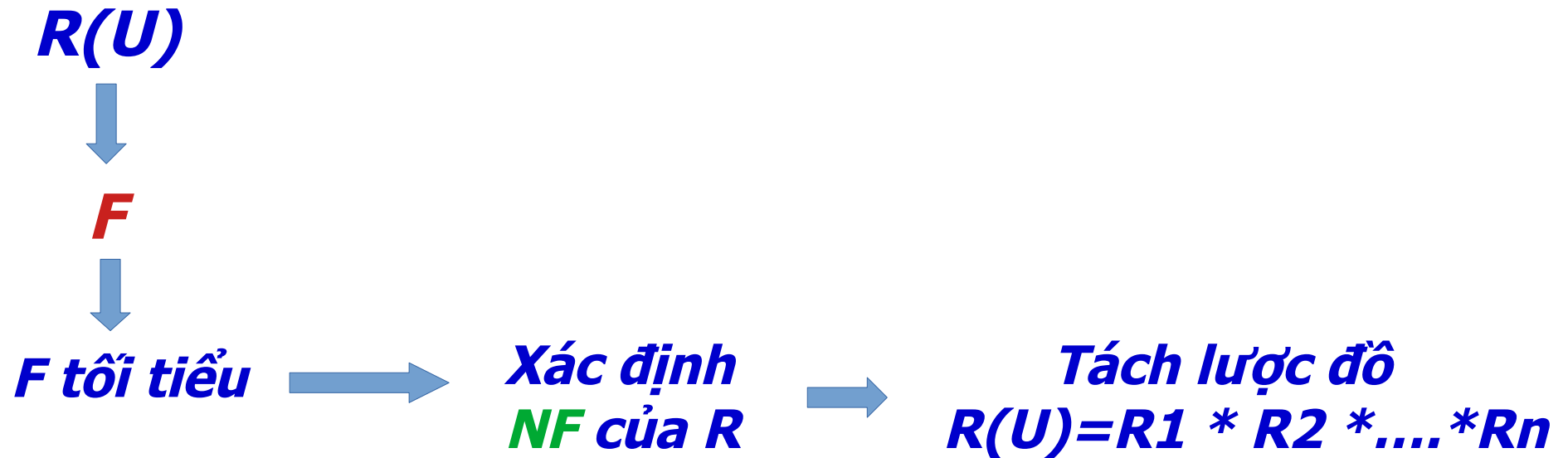
Nội dung

- Khóa của sơ đồ quan hệ
- Các dạng chuẩn của sơ đồ quan hệ

Nội dung

- Khóa của sơ đồ quan hệ
- Các dạng chuẩn của sơ đồ quan hệ

Các bước thiết kế CSDL theo dạng chuẩn



Khóa của sơ đồ quan hệ

- Cho lược đồ quan hệ $R(U)$, $U=\{A_1,A_2,...,A_n\}$ và tập PTH F
- **Định nghĩa khóa** : $K \subseteq U$ được gọi là khóa của lược đồ quan hệ $R(U)$ trên F nếu và chỉ nếu:
 - (1) Mọi thuộc tính đều PTH vào K tức là $K \rightarrow U$
 - (2) Không tồn tại tập con thực sự $X \subset K$ mà $X \rightarrow U$
- Siêu khóa:
 - Nếu K chỉ thỏa mãn điều kiện (1) mà không thỏa mãn điều kiện (2) thì K được gọi là siêu khóa, ký hiệu S
- Nhận xét:
 - R có thể có nhiều khóa
 - $A \subset K$ được gọi là thuộc tính khóa
- S và K xác định tất cả các thuộc tính của R

Xác định một khóa của sơ đồ quan hệ

- Cho lược đồ quan hệ $R(U)$, $U = \{A_1, A_2, \dots, A_n\}$ và tập PTH F
- Giải thuật tìm **một** khóa K
 - Bước 1: $K=U$, $i = 1$
 - Bước 2: Nếu $U \subseteq (K - A_j)_F^+$ thì $K = K - A_j$, $i = i+1$
 - Nếu $i > n$ thì dừng, ngược lại lặp lại bước 2
 - Bước 3 : Kết quả là K

Xác định một khóa của sơ đồ quan hệ - Ví dụ

- Cho $R(U)$
 - $U = \{A, B, C, D, E, F, G\}$
 - $F = \{B \rightarrow A, D \rightarrow C, D \rightarrow BE, DF \rightarrow G\}$
- Tìm khóa K của R ?

Xác định một khóa của sơ đồ quan hệ - Ví dụ

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

Tìm các khoá ứng viên của 1 LĐQR

Cho lược đồ quan hệ $S(U)$, $U = \{A_1, A_2, \dots, A_n\}$ và tập PTH F tối tiểu. Giải thuật tìm **các khoá** của quan hệ S gồm 3 bước

- **Bước 1, xác định các tập sau**
 - R : tập tất cả các thuộc tính **chỉ xuất hiện ở vế phải** của các PTH trong $F \Rightarrow R = \bigcup R_i - \bigcup L_i$
 - B : tập tất cả các thuộc tính **xuất hiện ở cả vế phải và vế trái** của các PTH trong $F \Rightarrow B = \bigcup R_i \cap \bigcup L_i$
 - $L = U \setminus \{R \cup B\}$ tập tất cả các thuộc tính **chỉ xuất hiện ở vế trái hoặc không xuất hiện trong** các PTH trong F
 $\Rightarrow L \cap R = \emptyset, L \cap B = \emptyset, R \cap B = \emptyset$

Tìm các khoá ứng viên của 1 LĐQR

- **Bước 2, Xét L**

- Nếu $L \neq \emptyset$: Các thuộc tính trong L **đều** là thuộc tính khoá
 - Nếu $(L)^+ = U \Rightarrow L$ là khoá K duy nhất của S, dừng
 - Ngược lại, tức là $(L)^+ \subset U$, qua bước 3

Tìm các khoá ứng viên của 1 LĐQR

- **Bước 3, tìm các khoá ứng viên**

- Tìm tất cả các tập con Xi của B và thêm lần lượt từng Xi vào L và tính bao đóng $L \cup Xi$.
 - Tìm tất cả $(L \cup Xi)$ là siêu khoá, tức là $(L \cup Xi)^+ = U$
 - Chọn các siêu khoá $(L \cup Xi)$ nhỏ nhất
- **Chú ý:**
 - Các thuộc tính nằm trong R không cần xem xét vì chúng không phải là thuộc tính khoá.
 - Nếu L rỗng thì tính bao đóng của các thuộc tính trong tập B

Tìm các khoá ứng viên của 1 LĐQR

- **Ví dụ 1:** Cho $S(\text{CTHRSG})$, tìm tất cả các khoá ứng viên của S với F như sau:

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

- Bước 1 : $U = \{\text{CTHRSG}\}$
 - $R = \cup Ri - \cup Li = \text{TCRG} - \text{CHRTS} = \{G\}$
 - $B = \cup Ri \cap \cup Li = \text{TCRG} \cap \text{CHRTS} = \{\text{CRT}\}$
 - $L = U - (R \cup B) = \text{CTHRSG} - \{\text{CRTG}\} = \{\text{HS}\} \neq \emptyset$
 $\Rightarrow \text{HS}$ là một phần của mọi khoá của S .
- Bước 2:
 - $(L)^+ = (\mathbf{HS})^+ = \{\text{CTHRSG}\} = U \Rightarrow \text{HS}$ là khoá duy nhất của S

Tìm các khoá ứng viên của 1 LĐQR

Ví dụ 2: Cho $S(ABCDEF)$, tìm tất cả các khoá ứng viên của S với F như sau:

$$F = \{AD \rightarrow B, AB \rightarrow ED, C \rightarrow B, B \rightarrow C, AC \rightarrow F\}$$

- Bước 1 :
 - $R = \cup Ri - \cup Li = BEDCF - ADBC = \{EF\}$
 - $B = \cup Ri \cap \cup Li = \{BCD\}$
 - $L = U - (R \cup B) = \{A\} \neq \emptyset \Rightarrow A$ là một phần của mọi khoá của S .
- Bước 2:
 - $(L)^+ = (\mathbf{A})^+ = \{A\} \subset \mathbf{U} \Rightarrow A$ không phải là khoá duy nhất của S ,
sang **bước 3**

Tìm các khoá ứng viên của 1 LĐQR

- **Bước 3:**

- Các tập con X_i của B là $\{B, C, D, BC, BD, CD\}$
- Tính bao đóng của $(L \cup X_i)$

X_i	$L \cup X_i$	$(L \cup X_i)^+$	SK	Khoá
B	AB	$(AB)^+ = U$	x	x
C	AC	$(AC)^+ = U$	x	x
D	AD	$(AD)^+ = U$	x	x
BC	ABC	$(ABC)^+ = U$	x	
BD	ABD	$(ABD)^+ = U$	x	
CD	ACD	$(ACD)^+ = U$	x	

Phạm Thị Ngọc Diễm

14

Vậy S có các khoá $K = \{AB, AC, AD\}$

Tìm các khoá ứng viên của 1 LĐQR

- **NHẬN XÉT :**

- A có trong mọi khoá
- Không cần tính bao đóng của L với các tập con còn lại của B, vì chúng sẽ đều là siêu khoá.

Xi	$L \cup Xi$	$(L \cup Xi)^+$	Khoá
B	AB	$(AB)^+ = U$	x
C	AC	$(AC)^+ = U$	x
D	AD	$(AD)^+ = U$	x

Nội dung

- Khóa của sơ đồ quan hệ
- Các dạng chuẩn của sơ đồ quan hệ

Lịch sử / tổng quan

- Chuẩn hóa Cơ sở dữ liệu được đề xuất lần đầu tiên (1970) bởi Edgar F. Codd.
 - Codd định nghĩa ba dạng chuẩn đầu tiên.
 - Codd và Raymond F. Boyce định nghĩa dạng chuẩn BCNF vào năm 1974
- Để chuẩn hóa, chúng ta phải biết những yêu cầu cho mỗi một dạng chuẩn.
- Một trong những yêu cầu quan trọng cần nhớ là : để có 3 NF chúng ta phải có 2 NF và để có 2 NF chúng ta phải có 1 NF.

Chuẩn hóa CSDL

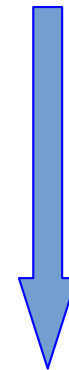
- Mục tiêu chính của chuẩn hóa cơ sở dữ liệu là để cơ cấu lại mô hình dữ liệu luận lý của một cơ sở dữ liệu để :
 - Loại bỏ sự dư thừa, nghĩa là không có các thông tin lặp lại
 - Tổ chức dữ liệu một cách hiệu quả
 - Giảm khả năng *dị thường dữ liệu* (data anomaly).
 - Tránh mất thông tin khi xóa.

Chuẩn hoá CSDL

=> Có thể ngăn chặn dị thường dữ liệu bằng cách thực hiện các cấp độ khác nhau của chuẩn hóa thường được gọi là các dạng chuẩn (Normal Form – NF)

- **4 dạng chuẩn cơ bản**

- Dạng chuẩn 1 (1 NF)
- Dạng chuẩn 2 (2 NF)
- Dạng chuẩn 3 (3 NF)
- Dạng chuẩn BCNF (Boyce Codd NF)



Các loại phụ thuộc hàm

- PTH tầm thường (trivial)
 - $X \rightarrow Y$ là tầm thường nếu $Y \subseteq X$
- PTH từng phần
 - $X \rightarrow Y$ là từng phần nếu $X \subset K$ và Y là thuộc tính không khoá
- PTH truyền
 - $X \rightarrow Y$ là truyền nếu X và Y là các thuộc tính không khoá

Dạng chuẩn 1

- Một quan hệ R được gọi là thỏa dạng chuẩn thứ nhất nếu và chỉ nếu :
 - Quan hệ bao gồm **một khóa chính**
 - Mọi thuộc tính của quan hệ R đều chứa các **giá trị nguyên tố** (thuộc tính đa trị không được phép).
 - Không có nhóm lặp lại: hai thuộc tính không lưu trữ thông tin tương tự trong cùng một quan hệ.

Dạng chuẩn 1 – ví dụ

NSX	loai	socho	congty
Airbus	A340	228	Air France
Boeing	B747	432	British Airways Qantas

=> Quan hệ không thỏa dạng chuẩn 1

NSX	loai	socho	congty
Airbus	A340	228	Air France
Boeing	B747	432	British Airways
Boeing	B747	432	Qantas

=> Quan hệ thỏa dạng chuẩn 1

Dạng chuẩn 1 – ví dụ

MSSV	Hoten	tpho	Mon1	Diem1	Mon2	Diem2
123	James	Paris	CS123	7	UE111	6
124	Smith	London	UE111	6	CS123	7

=> Quan hệ không thỏa dạng chuẩn 1

MSSV	Hoten	tpho	Mon	Diem
123	James	Paris	CS123	7
123	James	Paris	UE111	6
124	Smith	London	UE111	6
124	Smith	London	CS123	7

=> Quan hệ thỏa dạng chuẩn 1

Nhưng có trùng lặp dữ liệu

=> vấn đề dị thường khi thêm

=> khó khăn cập nhật

=> mất thông tin khi xóa

Dạng chuẩn 1 – ví dụ

MSSV	Hoten	tpho	Mon	Diem
123	James	Paris	CS123	7
123	James	Paris	UE111	6
124	Smith	London	UE111	6
124	Smith	London	CS123	7
123	James	Munich	UE112	6
125	Patrick	Tokyo	CS123	8

=> Quan sát khi thêm dòng 5, xóa dòng 6 và cập nhật thành phố cho James ???

- Thêm dòng 5 : **mâu thuẫn dữ liệu** (cùng sinh viên nhưng **tpho** khác nhau)
- Cập nhật : phải cập nhật thuộc tính **tpho** cho tất cả các dòng cho James
- Xóa dòng 6: **Mất thông tin** về sinh viên Patrick
=> Dạng chuẩn 2

Dạng chuẩn 2

- Một quan hệ được gọi là thỏa dạng chuẩn thứ 2 nếu và chỉ nếu:
 - Quan hệ thỏa dạng chuẩn 1 và
 - Mọi thuộc tính không khóa đều PTH vào khóa (Nghĩa là không tồn tại PTH mà vế trái là con của khóa hay không tồn tại **PTH từng phần**).

Dạng chuẩn 2 – Ví dụ

- Xét quan hệ:

- Các PTH:

MSSV \rightarrow hoten, tpho

MSSV, Mon \rightarrow diem

- Khóa của quan hệ là {MSSV, Mon}

=> Quan hệ thỏa dạng chuẩn 1, dữ liệu trùng lặp

- Xét PTH MSSV \rightarrow hoten, tpho

- **hoten, tpho** là thuộc tính không khóa nhưng không PTH vào khóa

=> vi phạm dạng chuẩn 2

MSSV	Hoten	tpho	Mon	Diem
123	James	Paris	CS123	7
123	James	Paris	UE111	6
124	Smith	London	UE111	6
124	Smith	London	CS123	7

Dạng chuẩn 2

- Để biết một lược đồ quan hệ ở dạng 1NF có thỏa dạng chuẩn 2 không:
 - Tìm khóa
 - Nếu khóa có từ 2 thuộc tính:
 - Tìm PTH từng phần là PTH mà vế trái là con của khóa (VP là thuộc tính không khóa).
 - Nếu tồn tại PTH như thế => không thỏa dạng chuẩn 2
 - Nếu không tồn tại PTH như thế => thỏa dạng chuẩn 2
 - Nếu khóa nhỏ hơn 2 thuộc tính => thỏa dạng chuẩn 2

Dạng chuẩn 2 – Ví dụ

=> Đưa về dạng chuẩn 2

$F = \{ \text{MSSV} \rightarrow \text{hoten, tpho}$
 $\text{MSSV, Mon} \rightarrow \text{diem} \}$
Khóa {MSSV, Mon}

MSSV	Hoten	tpho	Mon	Diem
123	James	Paris	CS123	7
123	James	Paris	UE111	6
124	Smith	London	UE111	6
124	Smith	London	CS123	7



Phân rã thành 2 quan hệ

1.

MSSV	Hoten	tpho
123	James	Paris
124	Smith	London

$F = \{ \text{MSSV} \rightarrow \text{hoten, tpho} \}$
Khóa: {MSSV}

2.

MSSV	Mon	Diem
123	CS123	7
123	UE111	6
124	UE111	6
124	CS123	7

$F = \{ \text{MSSV, Mon} \rightarrow \text{diem} \}$
Khóa : {MSSV, Mon}

Dạng chuẩn 3

- Một quan hệ được gọi là thỏa dạng chuẩn thứ 3 nếu và chỉ nếu:
 - Quan hệ thỏa dạng chuẩn 2 và
 - Mọi thuộc tính không khóa không phụ thuộc bắc cầu vào khóa chính (Hay không có PTH truyền).

Dạng chuẩn 3 – Ví dụ

- Xét quan hệ: SINHVIEN (MSSV, hoten, MLOP, tenlop, namvao)

MSSV	Hoten	MLOP	tenlop	namvao
123	James	CS04	Computer science 04	2004
124	Smith	SE03	Software 03	2003
125	Patrick	CS04	Computer science 04	2004

- Các PTH:

- MSSV → hoten, MLOP
- MLOP → tenlop, namvao

- Khóa {MSSV}

- Xét : MSSV → MLOP

MLOP → tenlop, namvao

=> MSSV → tenlop, namvao là PTH
bắc cầu

tenlop, namvao là các thuộc tính không
phụ thuộc bắc cầu vào khóa

=> không thỏa chuẩn 3

Dạng chuẩn 3 – Ví dụ

- PTH bắc cầu là nguyên nhân dẫn đến trùng lặp dữ liệu
=> dị thường dữ liệu
- Ví dụ: thêm dòng cuối vào quan hệ SINHVIEN

MSSV	Hoten	MLOP	tenlop	namvao
123	James	CS04	Computer science 04	2004
124	Smith	SE03	Software 03	2003
125	Patrick	CS04	Computer science 04	2004
126	Nathalie	CS04	System information 04	2005

Trùng lặp dữ
liệu

Dị thường dữ liệu

=> Dạng chuẩn 3 là dạng chuẩn tối thiểu mà một thiết kế CSDL phải thỏa mãn

Dạng chuẩn 3

- Để biết một lược đồ quan hệ ở dạng 2NF có thỏa dạng chuẩn 3 không:
 - Tìm khóa
 - Tìm PTH bắc cầu vào khóa hay truyền.
 - Nếu tồn tại PTH như thế => không thỏa dạng chuẩn 3
 - Nếu không tồn tại PTH như thế => thỏa dạng chuẩn 3

Dạng chuẩn 3 – Ví dụ

=> Đưa SINHVIEN về dạng chuẩn 3

$F = \{MSSV \rightarrow \text{hoten}, MLOP$
 $MLOP \rightarrow \text{tenlop}, \text{namvao}\}$
Khóa {MSSV}

MSSV	Hoten	MLOP	tenlop	namvao
123	James	CS04	Computer science 04	2004
124	Smith	SE03	Software 03	2003
125	Patrick	CS04	Computer science 04	2004
126	Nathalie	CS04	System information 04	2005



1.

MLOP	tenlop	namvao
CS04	Computer science 04	2004
SE03	Software 03	2003

$MLOP \rightarrow \text{tenlop}, \text{namvao}$
Khóa {MLOP}

2.

MSSV	Hoten	MLOP
123	James	CS04
124	Smith	SE03
125	Patrick	CS04
126	Nathalie	CS04

$MSSV \rightarrow \text{hoten}, MLOP$
Khóa {MSSV}

Dạng chuẩn BCNF (Boyce Codd)

- Một quan hệ ở dạng chuẩn BCNF nếu:
 - Quan hệ thỏa dạng chuẩn 3 và
 - Các PTH tầm thường hoặc nếu PTH không tầm thường thì **về trái phải là siêu khóa**.
- Định nghĩa 3NF không xử lý trường hợp một quan hệ:
 - Có nhiều khóa ứng viên, trong đó:
 - Những khóa ứng viên này là nhiều thuộc tính, và
 - Các khóa ứng viên chồng chéo lên nhau (nghĩa là, có ít nhất một thuộc tính chung)

Dạng chuẩn BCNF (Boyce Codd)

- Xét quan hệ:

PHIM (tua, nam, dodai, loaiphim, nsx, dienvien)

tua	nam	dodai	loaiphim	NSX	dienvien
Star Wars	1977	124	color	Fox	Fisher
Star Wars	1977	124	color	Fox	Hamill
Star Wars	1977	124	color	Fox	Ford
Mighty Ducks	1991	104	color	Disney	Esteves
Wayne's World	1992	95	color	Paramount	Carvey
Wayne's World	1992	95	color	Paramount	Meyers

Dạng chuẩn BCNF (Boyce Codd)

- Xét quan hệ:

PHIM (tua, nam, dodai, loaiphim, nsx, dienvien)

- PTH :

tua, nam \rightarrow dodai, loaiphim, nsx

- **{tua, nam, dienvien}** là một khóa ứng viên

=> PTH này vi phạm dạng chuẩn BCNF vì {tua, nam}
không xác định được dienvien

(nói cách khác, về trái PTH không phải là siêu khóa)

Dạng chuẩn BCNF (Boyce Codd)

=> Đưa về BCNF : Phân rã PHIM thành 2 quan hệ:

PHIM (tua, nam, dodai, loaiphim, nsx, dienvien)

PTH : tua, nam \rightarrow dodai, loaiphim, nsx



1. Quan hệ bao gồm các thuộc tính của PTH {**tua, nam**, dodai, loaiphim, NSX}

2. Quan hệ gồm tất cả các thuộc tính của PHIM trừ đi các thuộc tính đã xuất hiện ở vế phải của PTH {**tua, nam, dienvien**}

Dạng chuẩn BCNF (Boyce Codd)

- Một ví dụ khác:

NGUOI(id, hoten, diachi, sothich)

- PTH : $id \rightarrow hoten, diachi$
- Khóa {id, sothich}

=> Quan hệ NGUOI không thỏa BCNF vì **id** không xác định được *sothich* (nói cách khác, về trái PTH không phải là siêu khóa)

=> *CANHAN(id, hoten, diachi) $F=\{ id \rightarrow hoten, diachi\}$*

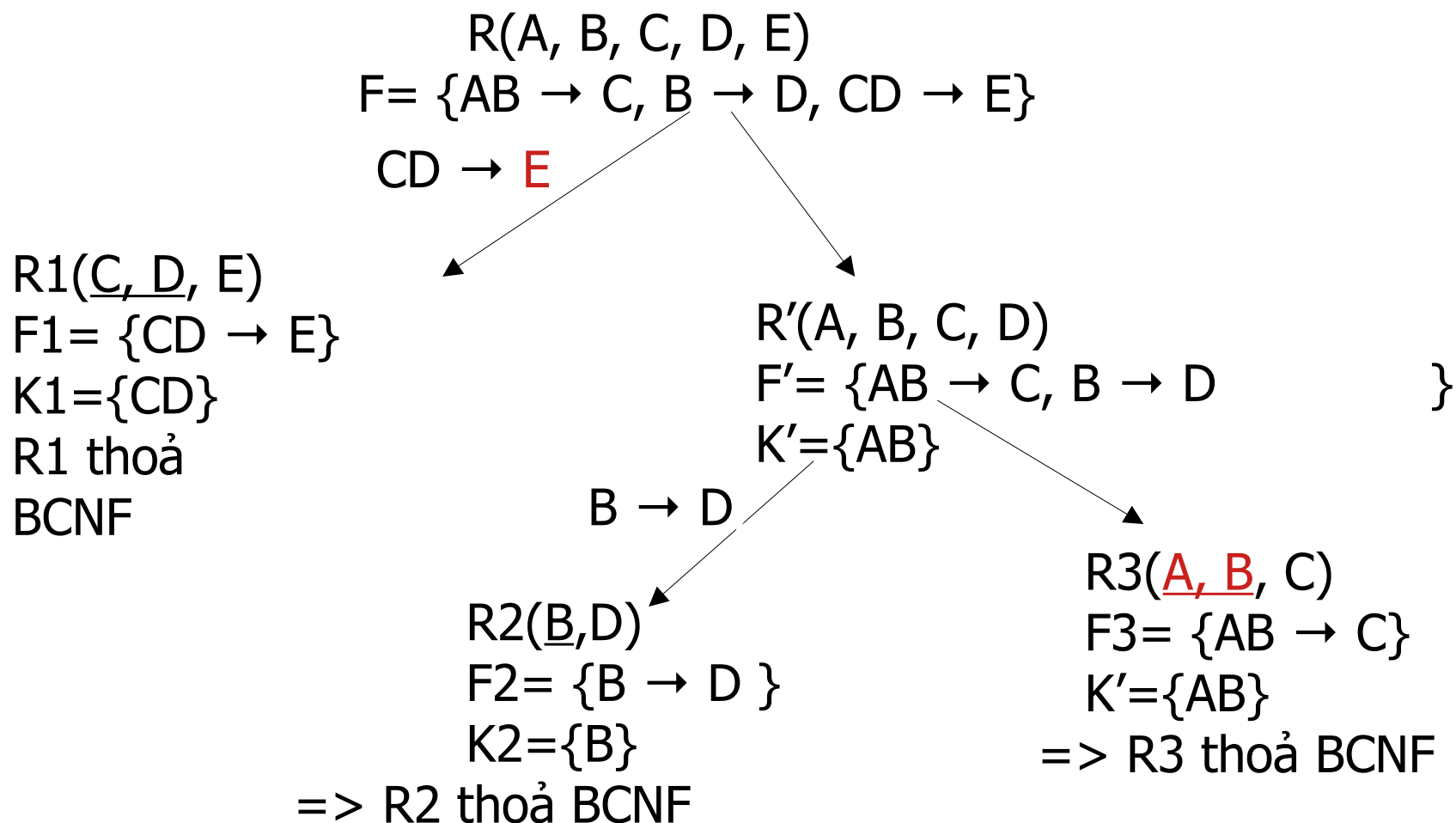
=> *THICH(id, sothich)*

Tóm tắt

- Để xác định dạng chuẩn của một lược đồ quan hệ $R(U)$ và tập F tối thiểu:
 - B0: Tìm khóa
 - B1: Kiểm tra tính nguyên tố của các thuộc tính,... => kiểm tra 1NF
 - B2: Tìm PTH từng phần => kiểm tra 2NF
 - B3: Tìm PTH bắc cầu vào Khóa (truyền) => kiểm tra 3NF
 - B4: Kiểm tra các PTH có vế trái là siêu khóa => Kiểm tra BCNF

Bài tập

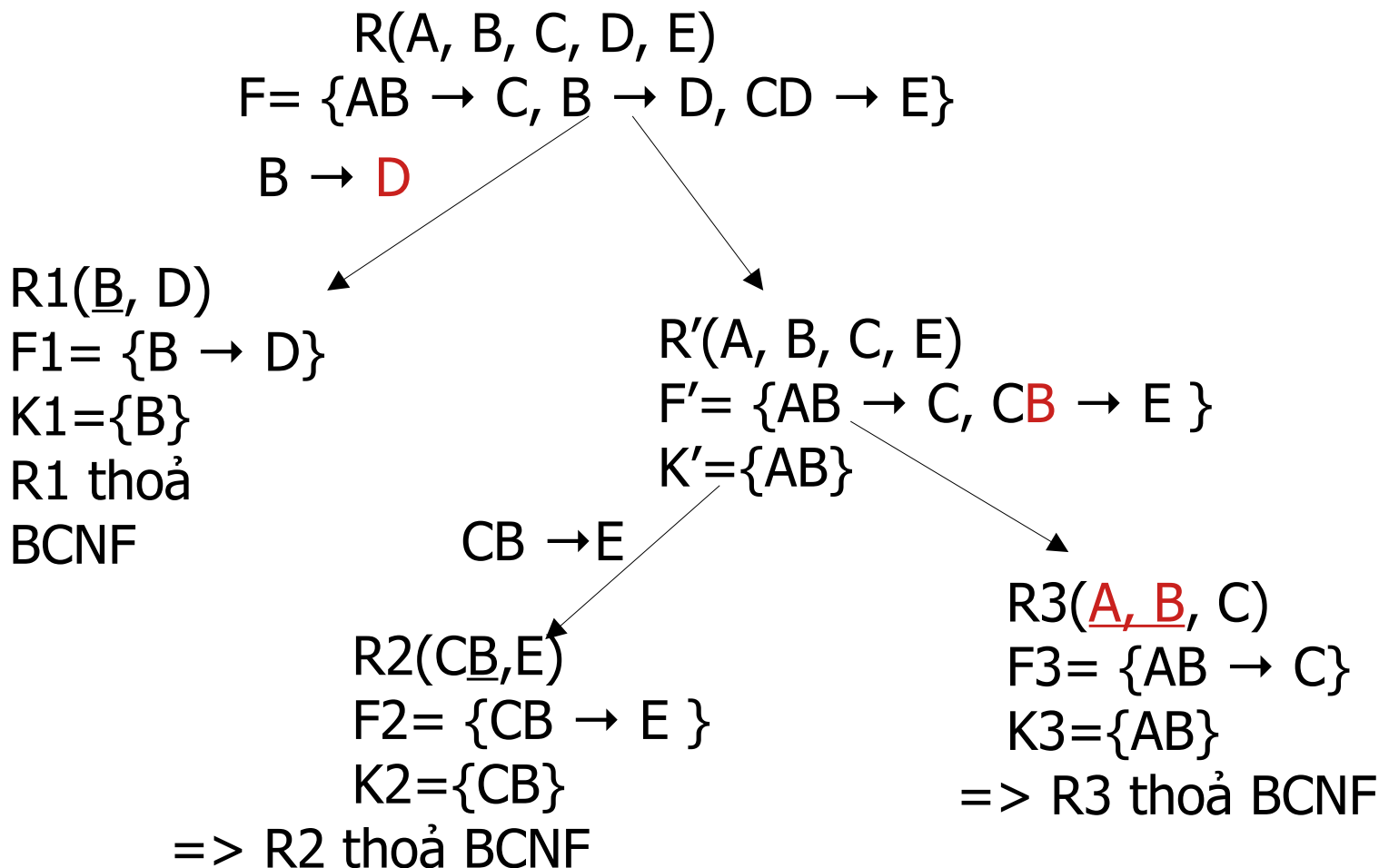
- Đưa $R(A, B, C, D, E)$ và $F = \{AB \rightarrow C, B \rightarrow D, CD \rightarrow E\}$ về dạng **BCNF**
- Khoá $K = \{AB\} \Rightarrow$ thoả 1NF do có PTH từng phần $B \rightarrow D$



KL: $R(A, B, C, D, E) = R1(\underline{C}, \underline{D}, E) * R2(\underline{B}, D) * R3(\underline{A}, \underline{B}, C)$

Bài tập

- Đưa $R(A, B, C, D, E)$ và $F = \{AB \rightarrow C, B \rightarrow D, CD \rightarrow E\}$ về dạng BCNF
- Khoá $K = \{AB\} \Rightarrow$ thoả 1NF do có PTH từng phần $B \rightarrow D$



KL: $R(A, B, C, D, E) = R1(\underline{B}, D) * R2(\underline{CB}, E) * R3(\underline{A}, \underline{B}, C)$

Bài tập

- **Bài 1** Hãy cho biết lược đồ sau ở dạng chuẩn cao nhất nào ?
 - **Câu 1** Cho $R(A, B, C, D, E)$ và $F = \{AB \rightarrow C, B \rightarrow D, CD \rightarrow E\}$
 - **Câu 2** : Cho $R(A, B, C, D)$ và
 $F = \{A \rightarrow BC, B \rightarrow CD, A \rightarrow D, D \rightarrow C\}$
 - **Câu 3** : Cho $R(A, B, C, D, E)$ và
 $F = \{AB \rightarrow C, A \rightarrow D, D \rightarrow E, AC \rightarrow B\}$
 - **Câu 4** : Cho $R(A, B, C, D)$ và $F = \{AB \rightarrow C, B \rightarrow D, BC \rightarrow A\}$.
 - **Câu 5**: Cho $R(\text{mamon}, \text{masv}, \text{tuoi}, \text{diem})$
 $F = \{\text{mamon}, \text{masv} \rightarrow \text{diem} ; \text{masv} \rightarrow \text{tuoi} \}$

Bài tập

- **Bài 1** Hãy cho biết lược đồ sau ở dạng chuẩn cao nhất nào ?
 - **Câu 6** Cho $R(A, B, C, D, E, F)$ và
 $F = \{A \rightarrow B, BC \rightarrow D, CD \rightarrow A, AE \rightarrow F, CE \rightarrow D\}$
 - Tìm Khoá
 - Quan hệ đã cho thoả dạng chuẩn BCNF ?