

PHÂN TÍCH THIẾT KẾ HỆ THỐNG THÔNG TIN

Information System Analysis and Design

Số tín chỉ: 4(3,1)

Số tiết: 75 tiết (45 LT- 30 TH)

Giảng viên: **TS. Đinh Thị Thu Hương.**

(Mobile: 0903087599 – e-mail: huongdtt@sgu.edu.vn)

Bộ môn: Khoa học máy tính – Khoa CNTT, SGU.

Chương 7: Lý thuyết chuẩn hóa cơ sở dữ liệu

- ☐ Phụ thuộc hàm.
- ☐ Bao đóng.
- ☐ Khóa của lược đồ quan hệ.
- ☐ Phủ tối thiểu.
- ☐ Dạng chuẩn.
- ☐ Chuẩn hóa cơ sở dữ liệu.
- ☐ Phép tách lược đồ quan hệ.
- ☐ Bài tập

7.5 Dạng chuẩn

7.5.0 Các loại PTH

❖ Phụ thuộc hàm riêng phần/bộ phận

$X \rightarrow A$ được gọi là *phụ thuộc hàm riêng phần* nếu tồn tại $Y \subset X$ để cho $Y \rightarrow A$.

❖ Phụ thuộc hàm đầy đủ

$X \rightarrow A$ được gọi là *phụ thuộc hàm* đầy đủ nếu không tồn tại $Y \subset X$ để cho $Y \rightarrow A$.

❖ Ví dụ

Cho lược đồ quan hệ $R(A,B,C)$ và tập PTH

$F = \{A \rightarrow B; A \rightarrow C; AB \rightarrow C\}$ thì $A \rightarrow B; A \rightarrow C$ là các PTH đầy đủ.

$AB \rightarrow C$ không là PTH đầy đủ vì có $A \rightarrow C$.

7.5 Dạng chuẩn

7.5.0 Các loại PTH

❖ Phụ thuộc hàm bậc cao

$X \rightarrow A$ được gọi là *phụ thuộc hàm* bậc cao nếu tồn tại Y để cho $X \rightarrow Y$, $Y \rightarrow A$, $Y \not\rightarrow X$ và $A \notin XY$.

❖ Ví dụ

Cho lược đồ quan hệ $R(A,B,C,D,E)$ và tập PTH $F=\{A \rightarrow BCE; B \rightarrow DC\}$

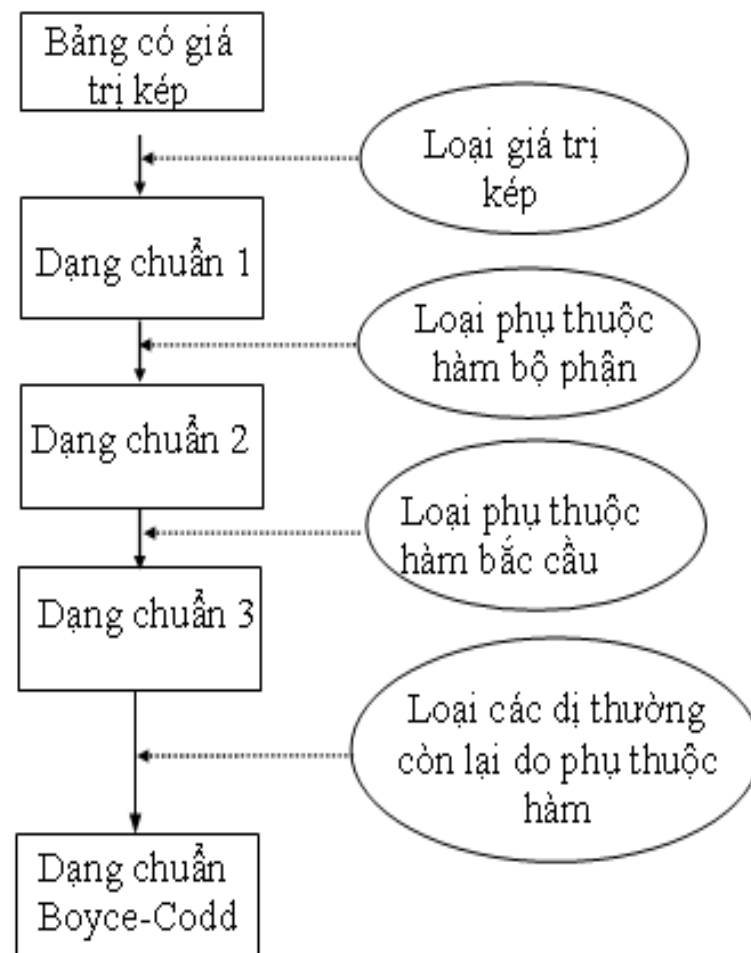
~~Thuộc tính~~ D,C là PTH bậc cao vào A .

7.5 Dạng chuẩn

7.5.1 Quy trình chuẩn hóa

❖ Khi thiết kế và cài đặt các hệ CSDL, chuẩn hoá là quá trình khảo sát danh sách các thuộc tính và áp dụng tập các quy tắc phân tích vào danh sách đó, biến đổi chúng thành nhiều tập nhỏ hơn sao cho:

- Tối thiểu việc lặp lại
- Tránh dị thường thông tin
- Xác định và giải quyết được sự không rõ ràng, nhập nhằng trong suy diễn.



7.5.2 Các dạng chuẩn

Dạng chuẩn Một (1NF-first normal form)

Lược đồ quan hệ R ở dạng chuẩn 1 nếu và chỉ nếu toàn bộ các miền có mặt trong R đều chỉ chứa các giá trị nguyên tố (không phân chia được nữa).

Ví dụ:

Canbo

MACB	HOTEN	NGAYSINH	QUEQUAN	NGOAINGU	TRINHDO_NN
01	Lê Văn An	5/5/1977	Nghệ An	Tiếng Anh Tiếng Pháp	C A
02	Nguyễn Văn Bình	10/04/1974	Hà Tĩnh	Tiếng Anh Tiếng Nhật Tiếng Hàn	B B A

Không ở
dạng 1NF

Ở dạng
1NF

Canbo

MACB	HOTEN	NGAYSINH	QUEQUAN	NGOAINGU	TRINHDO_NN
01	Lê Văn An	5/5/1977	Nghệ An	Tiếng Anh	C
01	Lê Văn An	5/5/1977	Nghệ An	Tiếng Pháp	A
02	Nguyễn Văn Bình	10/04/1974	Hà Tĩnh	Tiếng Anh	B
02	Nguyễn Văn Bình	10/04/1974	Hà Tĩnh	Tiếng Nhật	B
02	Nguyễn Văn Bình	10/04/1974	Hà Tĩnh	Tiếng Hàn	A

Dư thừa
dữ liệu

❖ Đưa về dạng chuẩn 1:

- Biến cột đa trị thành đơn trị
- Điền đủ dữ liệu vào các cột khác

Dạng chuẩn 2 (2NF-Second Normal Form)

Lược đồ quan hệ R ở dạng chuẩn 2 NF nếu R ở dạng chuẩn 1 và mọi thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào mọi thuộc tính khóa của R.

Ví dụ:

Cho lược đồ quan hệ R(ABCDEFG) thỏa mãn phụ thuộc hàm:

$$F = \{ A \rightarrow BC, C \rightarrow DE, E \rightarrow G \}$$

Kiểm tra R có thỏa dạng chuẩn 2NF không ?

Giải: Ta có: $(A)^+ = U \rightarrow K = A$

Các thuộc tính không khóa { B, C, D, E, G }

Do khóa chỉ có 1 thuộc tính nên R ở dạng 2NF

Dạng chuẩn 3(Third Normal Form)

Lược đồ quan hệ R ở dạng 3NF nếu có PTH $X \rightarrow A$ thỏa trên R thì:

- X là siêu khóa của R, hay
- A là thuộc tính khóa của R

Ví dụ:

Cho lược đồ quan hệ R (A, B, C, D) thỏa mãn phụ thuộc hàm:

$$F = \{ AB \rightarrow CD, D \rightarrow A \}$$

Kiểm tra R có thỏa dạng chuẩn 3NF không ?

Giải

Khóa của lược đồ quan hệ R là: $K = \{AB, BD\}$

Xét $D \rightarrow A$: A là thuộc tính khóa

Vậy R thỏa dạng chuẩn 3NF

Hệ quả: Một lược đồ quan hệ gọi là ở dạng chuẩn thứ 3 nếu nó ở dạng chuẩn thứ 2 và không có phụ thuộc hàm bắc cầu

Ví dụ 2:

Xét quan hệ CNHAN như sau:

CNHAN(MACN, LOAINGHE, HESOTHUONG)

- Khóa của quan hệ là MACN
- Ta thấy có các PTH trong quan hệ:

MACN → LOAINGHE

MACN → HESOTHUONG

LOAINGHE → HESOTHUONG

Dạng chuẩn cao nhất của quan hệ này?

Giải:

Ta có:

- Pth bắc cầu: **MACN** → LOAINGHE và
LOAINGHE → **HESOTHUONG**
- Thuộc tính **không khóa** HESOTHUONG **phụ thuộc bắc cầu vào thuộc tính khóa** MACN

Do đó: CNHAN không phải là 3NF.

Dạng chuẩn BCNF (Boyce - Codd)

Lược đồ quan hệ R được gọi là BCNF nếu $\forall X \rightarrow A$ trên R ($A \notin X$) thì: **X là một siêu khóa của R.**

Ví dụ 1: Cho lược đồ quan hệ R(A, B, C) thỏa mãn phụ thuộc hàm $F = \{AB \rightarrow C, BC \rightarrow A\}$

Kiểm tra R có thỏa dạng chuẩn BCNF không?

Giải: Khóa của R là : $K = \{AB, BC\}$

Xét $AB \rightarrow C$: Ta có AB là khóa, $C \notin AB$

Xét $BC \rightarrow A$: Ta có BC là khóa, $A \notin BC$

Vậy R thỏa dạng chuẩn BCNF

Ví dụ 2: Cho lược đồ của quan hệ PROJ(PNO, PNAME, BUDGET) thỏa mãn phụ thuộc hàm duy nhất

$PNO \rightarrow (PNAME, BUDGET)$,

Kiểm tra PROJ có thỏa dạng chuẩn BCNF không?

Định lý: Các lớp dạng chuẩn của 1 lược đồ quan hệ có quan hệ lồng nhau nghĩa là lớp sau nằm trọn trong lớp trước: $BCNF \subset 3NF \subset 2NF \subset 1NF$

Ví dụ:

Cho lược đồ quan hệ R (A, B, C, D) và $F = \{ AB \rightarrow C, D \rightarrow B, C \rightarrow ABD \}$

Kiểm tra R có thuộc dạng chuẩn BCNF không? Giải thích?

Giải:

- Khóa của lược đồ quan hệ R là : $K_1=AB$, $K_2=AD$, $K_3=C$
Nên R không có thuộc tính không khóa. Vậy R là dạng chuẩn 3NF nhưng không là BCNF.

Chú ý:

- Một quan hệ ở BCNF thì cũng đạt 3NF.
- Trong thực hành các quan hệ đạt chuẩn 3NF là đủ.
Tuy nhiên một quan hệ ở 3NF không đảm bảo đã loại bỏ được tất cả các lỗi khi thao tác dữ liệu.

Bài tập

1/ Cho lược đồ quan hệ $R(A, B, C, D, E)$ và tập phụ thuộc hàm F :

$$F = \{ AB \rightarrow C, AD \rightarrow E, B \rightarrow D \}$$

Kiểm tra R có thuộc dạng chuẩn 3NF không? Giải thích?

2/ Cho lược đồ quan hệ $Q (A, B, C, D, E)$ và tập phụ thuộc hàm F như sau: $F = \{AD \rightarrow C, AB \rightarrow E, D \rightarrow B\}$

a. Kiểm tra phụ thuộc hàm: $AD \rightarrow E$ có thuộc F^+ ?

b. Kiểm tra R có thuộc dạng chuẩn 2NF không?

BTVN

1/ Xét LĐQH R với $U=ABCDE$ và tập PTH

$$F = \{AB \rightarrow CE, E \rightarrow AB, C \rightarrow D\}$$

Dạng chuẩn cao nhất của quan hệ này là gì?

Bài 3:

Tìm phủ tối thiểu của tập phụ thuộc hàm:

$$T = \{ABH \rightarrow CK, A \rightarrow D, C \rightarrow E, BGH \rightarrow F, F \rightarrow AD, E \rightarrow F, BH \rightarrow E\}$$

7.6 Phép tách lược đồ quan hệ

7.6.1 Phép tách bảo toàn thông tin

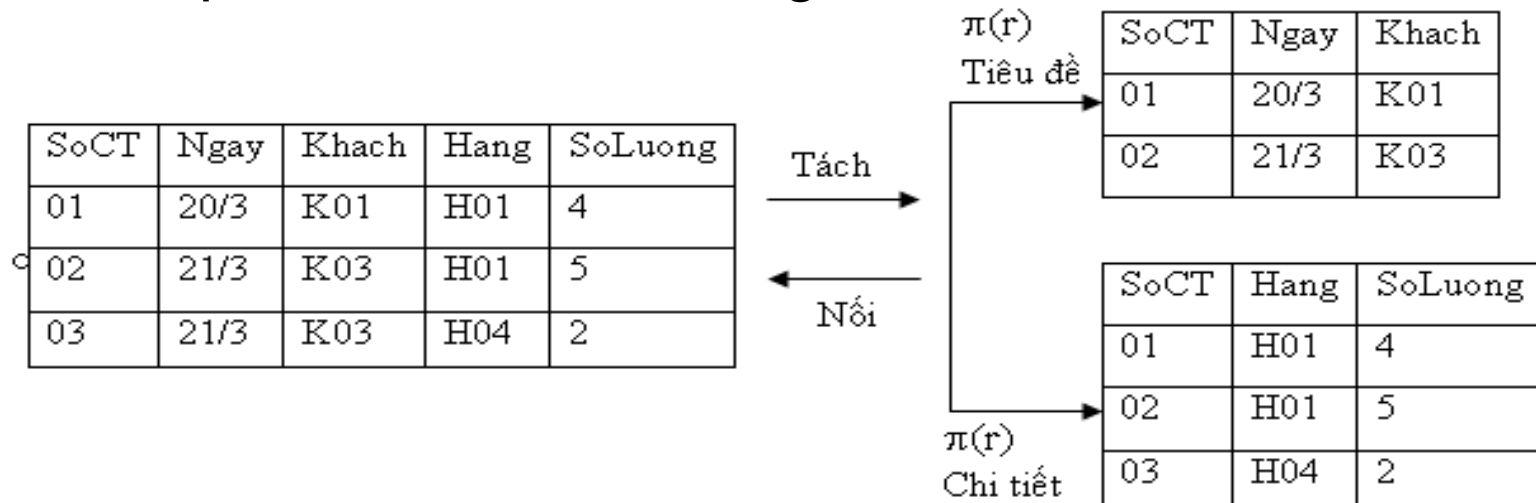
❖ **Định nghĩa:** Cho lược đồ quan hệ $R = A_1A_2...A_n$.

Tách lược đồ quan hệ R là thay thế R bằng các lược đồ con R_1, R_2, \dots, R_m sao cho $R_1 \cup R_2 \cup \dots \cup R_m = R$ và $R_i \neq R_j$ khi $i \neq j$

$\rho(R) = (R_1, R_2, \dots, R_m)$ bảo toàn thông tin

$$\Leftrightarrow \forall r(R) = \pi_{R_1}(r) * \pi_{R_2}(r) * \dots * \pi_{R_m}(r)$$

Ví dụ 1: Phép tách bảo toàn thông tin



Cho quan hệ r

S	P	D
s1	p1	d1
s2	p2	d2
s3	p1	d3

Instance r

Kết nối tự nhiên 2 quan hệ phân rã này:

Phân rã thành 2 quan hệ

S	P
s1	p1
s2	p2
s3	p1

$\pi_{SP}(r)$

P	D
p1	d1
p2	d2
p1	d3

$\pi_{PD}(r)$

S	P	D
s1	p1	d1
s2	p2	d2
s3	p1	d3
s1	p1	d3
s3	p1	d1

$\pi_{SP}(r) \bowtie \pi_{PD}(r)$

→ Có những bộ không thuộc quan hệ gốc ban đầu

Thuật toán kiểm tra phép tách ko mất thông tin

Vào: Lược đồ quan hệ $R=\{A_1, A_2, \dots, A_n\}$, tập các PTH F và phép tách: $\rho(R_1, R_2, \dots, R_k)$

Ra: Kết luận phép tách ρ không mất tin.

Các bước của thuật toán:

B1: 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} .

B2. Xét $f = X \rightarrow Y \in F$ và xét các dòng trong bảng có giá trị bằng nhau trên thuộc tính X thì **làm bằng** các giá trị của chúng trên Y . Ngược lại **bằng** chúng bằng ký hiệu b_{ij} . Tiếp tục áp dụng các PTH cho bảng (kể cả việc lặp lại các phụ thuộc hàm đã áp dụng) cho tới khi không còn áp dụng được nữa.

B3. Nếu bảng có 1 hàng gồm các kí hiệu a_1, a_2, \dots, a_n thì phép tách là không mất mát thông tin.

Ví dụ 3: Cho LĐQH $r(R)$ với $R=\{A,B,C,D\}$, tập PTH F :
 $F=\{A \rightarrow B, AC \rightarrow D\}$, phép phân rã $\rho(AB, ACD)$. Hãy kiểm tra
 phép phân rã ρ có mất thông tin ko?

Giải:

Lập bảng

Dùng PTH $A \rightarrow B$ để làm bằng
 phần tử b_{22} thành a_2

	A	B	C	D
AB	a1	a2	b13	b14
ACD	a1	b22	a3	a4

	A	B	C	D
AB	a1	a2	b13	b14
ACD	a1	b22/a2	a3	a4

Vì bảng có dòng 2 chứa toàn a_i nên phân rã trên là không mất thông tin.

7.6.2 Phép tách bảo toàn tập phụ thuộc hàm

❖ Hình chiếu của một phụ thuộc hàm trên một tập thuộc tính:

Hình chiếu của F trên một tập các thuộc tính Z , ký hiệu là $\Pi_Z(F)$ là tập các phụ thuộc $X \rightarrow Y$ thuộc F^+ sao cho $XY \subset Z$ (chú ý $X \rightarrow Y$ có thể không thuộc F và chỉ thuộc F^+)

❖ Thuật toán tìm hình chiếu của F một phụ thuộc hàm lên một tập thuộc tính:

- + Tính X^+ ($X \subset Z$ làm VT của PTH)
- + $Y = (X^+ \cap Z) \setminus X$
- + Khi đó $X \rightarrow Y$ là một PTH F/Z

❖ **Định nghĩa:** Một phân rã $\rho = (R_1, R_2, \dots, R_k)$ của lược đồ quan hệ R trên tập PTH F bảo toàn phụ thuộc nếu có thể suy ra được F từ các hình chiếu của F trên R_i .

$$\left(\bigcup_i^n F_i \right)^+ = F^+$$

7.6.2 Phép tách bảo toàn tập phụ thuộc hàm

Ví dụ 1: $R = (A, B, C)$, $F = \{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$. Key: **A**
Tách R thành 2 bảng $R1(A, B)$, $R2(B, C)$. Kiểm tra tách LĐQH mất mát thông tin và bảo toàn phụ thuộc hàm không?

A	B	C
1	2	3
2	2	3
3	2	3
4	2	4

A	B
1	2
2	2
3	2
4	2

B	C
2	3
2	4

Giải:

- ✓ Phép tách bảo toàn phụ thuộc vì $F1 = \{A \rightarrow B\}$, $F2 = \{B \rightarrow C\}$ và $(F1 \cup F2)^+ = F^+$

Ví dụ 2: $R = (A, B, C)$, $F = \{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$. Key: **A**
Tách R thành 2 bảng $R1(A, B)$, $R2(A, C)$. Kiểm tra tách LĐQH mất mát thông tin và bảo toàn phụ thuộc hàm không?

A	B	C
1	2	3
2	2	3
3	2	3
4	2	4

A	B
1	2
2	2
3	2
4	2

A	C
1	3
2	3
3	3
4	4

Giải:

- ✓ Phép tách không bảo toàn phụ thuộc vì $F1 = \{A \rightarrow B\}$, $F2 = \{A \rightarrow C\}$ và $(F1 \cup F2)^+ \neq F^+$ PTH $\{B \rightarrow C\}$. Đã bị mất

Hai tính chất của phép tách:

- Không mất thông tin và Bảo toàn phụ thuộc hàm.
- Vì chuẩn BCNF không có dư thừa và chuẩn 3NF tuy có dư thừa nhưng cũng hạn chế, nên việc phân rã lược đồ sẽ chỉ tập trung vào hai dạng này.

Thuật toán đưa về dạng chuẩn 3NF bảo toàn thông tin

Thuật toán:

Đầu vào: $\langle R, F \rangle$

Đầu ra: $\rho(R)$ thỏa 3NF bảo toàn thông tin

Phương pháp:

Bước 1: Tìm khoá của R và giả sử F là đầy đủ và không dư thừa

Bước 2: Nếu $X \rightarrow A$ và X không chứa khoá của R: $R = (XA, R \setminus A)$

Lặp lại bước 2 với $R \setminus A$ cho đến khi không tách được.

Ví dụ:

$R = \text{MTGPSL}$, $F = \{M \rightarrow T, \underline{GP} \rightarrow M, GT \rightarrow P, MS \rightarrow L, GS \rightarrow P\}$

Tách LĐQH thành thỏa 3NF bảo toàn thông tin.

Thuật toán đưa về dạng chuẩn Boyce-Codd bảo toàn thông tin

- ❖ Đầu vào: $\langle R, F \rangle$
- ❖ Đầu ra: $\rho(R)$ thoả BCNF bảo toàn thông tin
- ❖ Phương pháp:
 - Phương pháp chủ yếu của thuật toán là tách lược đồ $s = \langle R, F \rangle$ thành 2 lược đồ
 - Chọn bất kỳ $X \rightarrow A \in F^+$ sao cho X không là khoá và $A \notin X$. Khi đó lược đồ có tập các thuộc tính XA sẽ có dạng chuẩn BCNF và phụ thuộc hàm $X \rightarrow A$ sẽ thoả trên nó. Lược đồ thứ 2 có tập các thuộc tính $R \setminus A$. Hiển nhiên, khi kết nối lược đồ có tập thuộc tính $R \setminus A$ với lược đồ có tập thuộc tính XA không tổn thất thông tin. Tiếp tục tách $R \setminus A$ cho đến khi trở thành lược đồ có dạng chuẩn BCNF

Ví dụ:

Cho $\Omega = \text{CTHRSG}$, trong đó:

C : Khoá học, T: Thầy giáo, H: Giờ học R: Phòng học, S : Sinh viên,
G: Lớp

Biết rằng:

- Mỗi khoá học chỉ có một thầy dạy
- Một phòng học tại giờ xác định chỉ có một khoá học
- Thầy dạy tại giờ học cụ thể xác định phòng học cụ thể
- Khoá học với một sinh viên cụ thể xác định lớp học cụ thể
- Mỗi một sinh viên học trong một giờ xác định tại phòng học cụ thể

BTVN

1/ Cho lược đồ quan hệ $\alpha = (U, F)$ với

$U = \{A_1, A_2, A_3, A_4, A_5\}$

$F = \{A_1 \rightarrow A_2 A_3, A_2 A_4 \rightarrow A_5, A_2 \rightarrow A_3\}$

$\rho = \{A_1 A_2 A_4, A_2 A_3, A_1 A_4 A_5\}$

Kiểm tra phép tách ρ có mất thông tin không?

2/ Cho quan hệ $R(U, F)$ với

$U = \{A B C D E G H\}$

$F = \{A \rightarrow D, B \rightarrow CE, C \rightarrow G, E \rightarrow H\}$

Hãy kiểm tra tính mất mát thông tin khi tách R thành: $R_1(ABCE)$, $R_2(ABDH)$, $R_3(CGH)$.

Bài 1

Giải:

Xây dựng bảng gồm 3 dòng 5 cột
- Điền các tín hiệu vào bảng

	A_1	A_2	A_3	A_4	A_5
U_1	a_1	a_2	b_{13}	a_4	b_{15}
U_2	b_{12}	a_2	a_3	b_{24}	b_{25}
U_3	a_1	b_{32}	b_{33}	a_4	a_5

- Biến đổi bảng trên dựa vào tập phụ thuộc hàm F
- + Sử dụng phụ thuộc hàm $A_1 \rightarrow A_2 A_3$ ta biến đổi bảng

	A_1	A_2	A_3	A_4	A_5
U_1	a_1	a_2	b_{13}	a_4	b_{15}
U_2	b_{12}	a_2	a_3	b_{24}	b_{25}
U_3	a_1	a_2	b_{13}	a_4	a_5

+ Sử dụng phụ thuộc hàm $A_2 A_4 \rightarrow A_5$

	A_1	A_2	A_3	A_4	A_5
U_1	a_1	a_2	b_{13}	a_4	a_5
U_2	b_{12}	a_2	a_3	b_{24}	b_{25}
U_3	a_1	a_2	b_{13}	a_4	a_5

+ Sử dụng phụ thuộc hàm $A_2 \rightarrow A_3$

	A_1	A_2	A_3	A_4	A_5
U_1	a_1	a_2	a_3	a_4	a_5
U_2	b_{12}	a_2	a_3	b_{24}	b_{25}
U_3	a_1	a_2	a_3	a_4	a_5

Trong bảng này có hàng cuối cùng gồm toàn các tín hiệu chính, do vậy phép tách ρ là phép tách kết nối không mất thông tin.

Bài 2

Cho quan hệ $R(U, F)$: $U = \{A B C D E G H\}$ và $F = \{A \rightarrow D, B \rightarrow CE, C \rightarrow G, E \rightarrow H\}$. Hãy kiểm tra tính mất mát thông tin khi tách R thành: $R_1(ABCE)$, $R_2(ABDH)$, $R_3(CGH)$.

Giải:

Bài 3

Cho quan hệ $R(U, F)$: $U = \{A B C D E\}$ và
 $F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$.

Tách R thành: $R_1(ABC)$, $R_2(AD)$

- Phép tách có mất thông tin không?
- Phép tách có bảo toàn PTH không?

Bài 4

Cho LĐQH $r(R)$ với $R=\{A, B, C, D\}$ và tập PTH $F=\{AB \rightarrow CD, B \rightarrow C\}$.
R có là 2 NF?

Bài 6

Cho LĐQH $r(R)$ với $R=ABCD$

$F = \{A \rightarrow C, D \rightarrow B, C \rightarrow ABD\}$

Xác định dạng chuẩn cao nhất?

BTVN

1/ Cho lược đồ quan hệ HoaDon với HoaDon(SOHD, KHACH, NGAYLAP, MATHANG, DONGIA, SOLUONG) và tập các phụ thuộc hàm F như sau:

$F = \{ \text{SOHD} \rightarrow \text{KHACH}, \text{NGAYLAP}, \text{SOHD}, \text{MATHANG} \rightarrow \text{DONGIA}, \text{SOLUONG} \}$

Hãy cho biết lược đồ quan hệ HoaDon có đạt dạng chuẩn nào ? Tại sao?

2/ $R = (A, B, C, D); F = \{ABC \rightarrow D, D \rightarrow A\}$

- Xác định tất cả khóa của R
- Xác định dạng chuẩn cao nhất

BTVN

3/ Cho lược đồ quan hệ HoaDon với HoaDon(SOHD, KHACH, NGAYLAP, MATHANG, DONGIA, SOLUONG) và tập các phụ thuộc hàm F như sau:

$F = \{ \text{SOHD} \rightarrow \text{KHACH}, \quad \text{NGAYLAP}, \quad \text{SOHD}, \text{MATHANG} \rightarrow \text{DONGIA}, \text{SOLUONG} \}$

Hãy cho biết lược đồ quan hệ HoaDon có đạt dạng chuẩn nào ? Tại sao?