

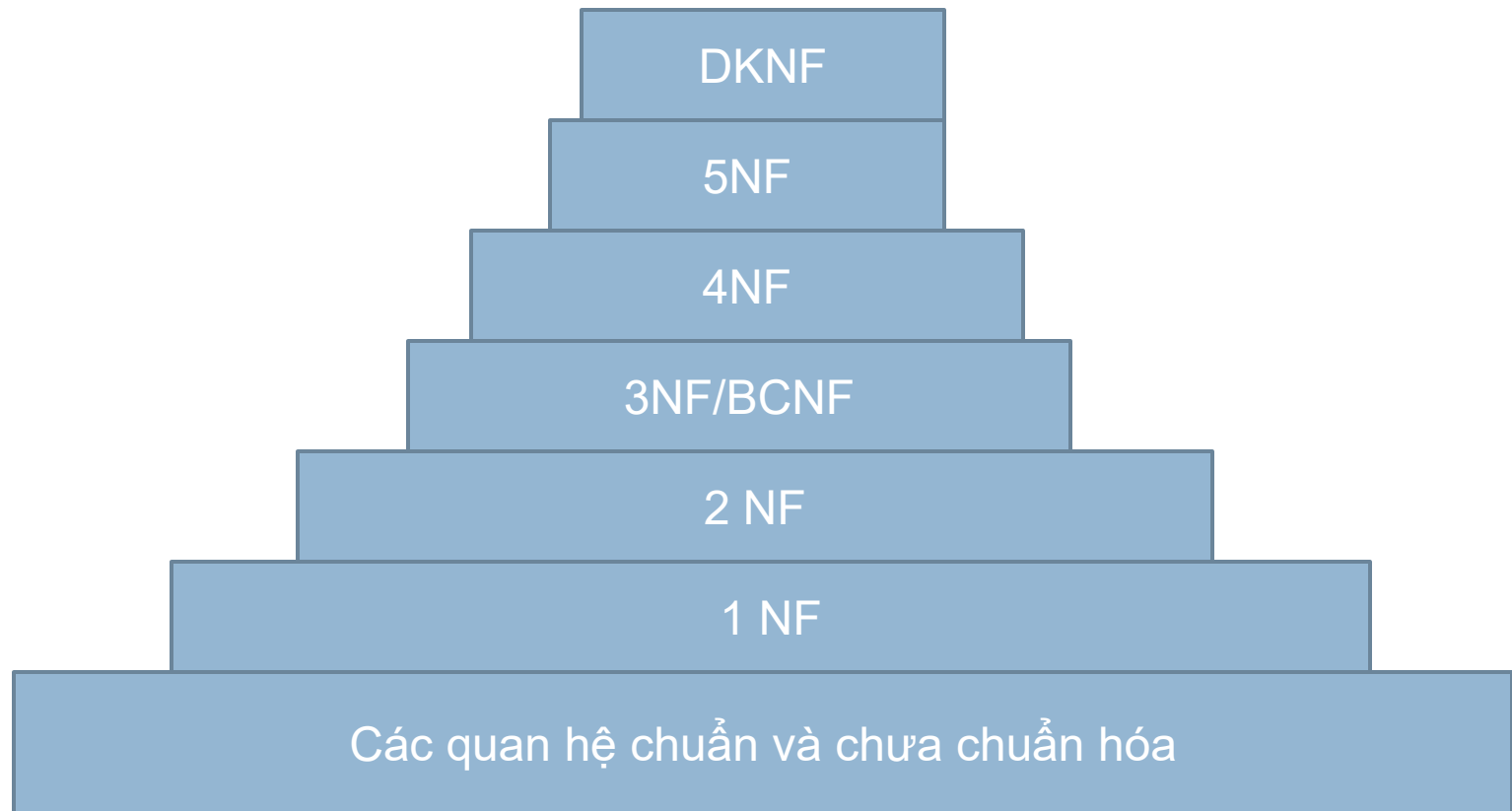
Chuẩn hóa Cơ sở dữ liệu

- ❑ Loại bỏ dư thừa
- ❑ Các dạng chuẩn

Tại sao phải chuẩn hóa CSDL

- Triệt tiêu sự ***Dị thường thông tin***
- Chuẩn hóa là quá trình:
 - ▣ Khảo sát các thuộc tính
 - ▣ Áp dụng các quy tắc phân tích
 - ▣ Biến đổi các t/t thành tập nhỏ hơn
 - Tối thiểu việc lặp lại
 - Tránh dị thường thông tin
 - Xác định và giải quyết sự không rõ ràng trong suy dẫn
- ***Chuẩn hóa là quá trình tách LDQH về 1 nhóm tương đương các LDQH chiếu sao cho khi kết nối tự nhiên không làm tổn thất thông tin và bảo toàn các PTH***

Các dạng chuẩn



Dạng chuẩn 1

- Một lược đồ quan hệ R ở dạng chuẩn 1 nếu toàn bộ các thuộc tính của mọi bộ đều mang giá trị đơn.

MASV	HOTEN	KHOA	MONHOC	DIEMTHI
1234	HUYNH CHI THANH	CONG NGHE THONG TIN	KY THUAT LAP TRINH CO SO DU LIEU HE DIEU HANH	6 7 8
4321	MAI THI LUU	VIEN THONG	MANG VIEN THONG	9

MASV	HOTEN	KHOA	MONHOC	DIEMTHI
1234	HUYNH CHI THANH	CONG NGHE THONG TIN	KY THUAT LAP TRINH	6
1234	HUYNH CHI THANH	CONG NGHE THONG TIN	CO SO DU LIEU	7
1234	HUYNH CHI THANH	CONG NGHE THONG TIN	HE DIEU HANH	8
4321	MAI THI LUU	VIEN THONG	MANG VIEN THONG	9

Dạng chuẩn 2

- **Một lược đồ quan hệ R ở dạng chuẩn 2 nếu R đạt chuẩn 1 và mọi thuộc tính không khóa của R đều phụ thuộc đầy đủ vào khóa.**
- **Thuật toán kiểm tra dạng chuẩn 2**
 - ▣ Vào: lược đồ quan hệ R , tập phụ thuộc hàm F
 - ▣ Ra: khẳng định R đạt chuẩn 2 hay không đạt chuẩn 2.
 - ▣ **Bước 1: Tìm tất cả khóa của R**
 - ▣ **Bước 2: Với mỗi khóa K , tìm bao đóng của tất cả tập con thật sự S của K .**
 - ▣ **Bước 3: Nếu có bao đóng S^+ chứa thuộc tính không khóa thì R không đạt chuẩn 2, ngược lại thì R đạt chuẩn 2**

Ví dụ chuẩn 2

- ❑ **Ví dụ:** $R(A,B,C,D)$; $F=\{AB \rightarrow C, B \rightarrow D, BC \rightarrow A\}$; R có đạt chuẩn 2 không?
- ❑ $TN=\{B\}; TG=\{AC\}$

X_i	$X_i=(TN \cup TG_i)$	$(X_i)^+$	Siêu khóa	Khóa
ϕ	B	BD		
A	AB	ABCD	AB	AB
C	BC	ABCD	BC	BC
AC	ABC	ABCD	ABC	

- ❑ Khóa là $K1=AB$ và $K2=BC$.
- ❑ Vì $B \subset K1; B \rightarrow D; D$ là t/t không khóa \Rightarrow t/t không khóa KHÔNG phụ thuộc đầy đủ vào khóa \Rightarrow R không đạt chuẩn 2.

Ví dụ chuẩn 2 (tt)

- **Ví dụ:** $R(G, M, V, N, H, P)$;
 $F = \{G \rightarrow M, G \rightarrow N, G \rightarrow HP, M \rightarrow V, MHP \rightarrow M\}$; R có đạt chuẩn 2 không?
- $TN = \{G\}$; $TG = \{M, N, H, P\}$
- R chỉ có một khóa và khóa chỉ có một thuộc tính nên mọi thuộc tính đều phụ thuộc đầy đủ vào khóa \Rightarrow R đạt chuẩn 2.

Xi	Xi=(TN \cup TGi)	(Xi) ⁺	Siêu khóa	Khóa
ϕ	G	R ⁺	G	G
M	GM	R ⁺	GM	
N	GN	R ⁺	GN	
MN	GMN	R ⁺	GMN	
H	GH	R ⁺	GH	
MH	GMH	R ⁺	GMH	
NH	GNH	R ⁺	GNH	
MNH	GMNH	R ⁺	GMNH	
P	GP	R ⁺	GP	
MP	GMP	R ⁺	GMP	
NP	GNP	R ⁺	GNP	
MNP	GMNP	R ⁺	GMNP	
HP	GHP	R ⁺	GHP	
MHP	GMHP	R ⁺	GMHP	
NHP	GNHP	R ⁺	GNHP	
MNHP	GMNHP	R ⁺	GMNHP	

Dạng chuẩn 3

- **Thuộc tính bắc cầu:** R là LDQH; X, Y là 2 tập con của R^+ , A là 1 t/t. Nói rằng A phụ thuộc bắc cầu vào X nếu cả 3 đ/k sau thỏa:
 - $X \rightarrow Y, Y \rightarrow A$
 - $Y \not\rightarrow X$
 - $A \not\subset XY$
- **ĐN1:** LDQH R ở dạng chuẩn 3 nếu mọi phụ thuộc hàm $X \rightarrow A \in F^+$ $A \not\subset X$ đều có:
 - Hoặc X là siêu khóa
 - Hoặc A là t/t khóa
- **ĐN2:** LDQH R ở dạng chuẩn 3 nếu mọi t/t không khóa của R đều không phụ thuộc bắc cầu vào một khóa bất kỳ của R

Dạng chuẩn 3

□ Thuật toán kiểm tra dạng chuẩn 3

- ▣ Vào: lược đồ quan hệ R , tập phụ thuộc hàm F
- ▣ Ra: khẳng định R đạt chuẩn 3 hay không đạt chuẩn 3.
- ▣ **Bước 1: Tìm tất cả khóa của R**
- ▣ **Bước 2: Từ F tạo tập phụ thuộc hàm tương đương F_{1tt} có vẻ phải 1 thuộc tính.**
- ▣ **Bước 3: Nếu mọi PTH $X \rightarrow A \in F_{1tt}$ với $A \notin X$ đều có X là siêu khóa hoặc A là t/t khóa thì R đạt chuẩn 3, ngược lại R không đạt chuẩn 3.**

Dạng chuẩn 3

- **Ví dụ:** $R(A,B,C,D)$
 $F=\{AB \rightarrow C, D \rightarrow B, C \rightarrow ABD\}$; R có đạt chuẩn 3 không?
- $TN=\phi$; $TG=\{ABCD\}$
- $K1=\{AB\}$; $K2=\{AD\}$; $K3=\{C\}$ là các khóa \Rightarrow mọi PTH $X \rightarrow A$ đều có A là t/t khóa. Vậy R đạt chuẩn 3.

X_i	$(TN \cup X_i)$	$(TN \cup X_i)^+$	Siêu khóa	Khóa
ϕ	ϕ	ϕ		
A	A	A		
B	B	B		
AB	AB	ABCD	AB	AB
C	C	ABCD	C	C
AC	AC	ABCD	AC	
BC	BC	ABCD	BC	
ABC	ABC	ABCD	ABC	
D	D	BD		
AD	AD	ABCD	AD	AD
BD	BD	BD		
ABD	ABD	ABCD	ABD	
CD	CD	ABCD	CD	
ACD	ACD	ABCD	ACD	
BCD	BCD	ABCD	BCD	
ABCD	ABCD	ABCD	ABCD	

Dạng chuẩn Boyce Codd

- **Một quan hệ Q ở dạng chuẩn BC nếu mọi phụ thuộc hàm $X \rightarrow A \in F^+$ với $A \notin X$ đều có X là siêu khóa.**
- **Thuật toán kiểm tra dạng chuẩn BC**
 - ▣ Vào: lược đồ quan hệ R , tập phụ thuộc hàm F
 - ▣ Ra: khẳng định R đạt chuẩn BC hay không đạt chuẩn BC.
 - ▣ **Bước 1: Tìm tất cả khóa của R**
 - ▣ **Bước 2: Từ F tạo tập phụ thuộc hàm tương đương F_{1tt} có vẻ phải 1 thuộc tính.**
 - ▣ **Bước 3: Nếu mọi PTH $X \rightarrow A \in F_{1tt}$ với $A \notin X$ đều có X là siêu khóa thì R đạt chuẩn BC, ngược lại R không đạt chuẩn BC.**

Chuẩn BC

- **Ví dụ:** $R(A, B, C, D, E, I)$
 $F = \{ACD \rightarrow EBI, CE \rightarrow AD\}$; R có đạt chuẩn BC không?

- $TN = C$; $TG = \{ADE\}$

- $F \cong F_{1tt} = \{ACD \rightarrow E, ACD \rightarrow B, ACD \rightarrow I, CE \rightarrow A, CE \rightarrow D\}$

- Mọi PTH của F_{1tt} đều có vế trái là siêu khóa \Rightarrow R đạt chuẩn BC

X_i	$(TN \cup TG_i)$	$(X_i)^+$	Siêu khóa	Khóa
ϕ	C	C		
A	AC	AC		
D	CD	CD		
AD	ACD	ABCDEI	ACD	ACD
E	CE	ABCDEI	CE	CE
AE	ACE	ABCDEI	ACE	
DE	CDE	ABCDEI	CDE	
ADE	ACDE	ABCDEI	ACDE	

- **Ví dụ:** $Q(SV, MH, THAY)$ $F = \{SV, MH \rightarrow THAY, THAY \rightarrow MH\}$; \Rightarrow Q đạt chuẩn 3 nhưng không đạt chuẩn BC

Phép tách về lược đồ dạng chuẩn 3NF có bảo toàn phụ thuộc

- Input: $s = \langle R, F \rangle$
- Output: Một phép tách bảo toàn phụ thuộc sao cho mỗi LDQH chiếu có dạng chuẩn 3NF tương ứng với tập phụ thuộc chiếu của F trên nó

Phương pháp tách về lược đồ dạng chuẩn 3NF có bảo toàn phụ thuộc

- Nếu có những t/t của R không có mặt trong các VP và VT của các PTH thì tập các t/t này sẽ tạo ra 1 LDQH chiếu.
 - ▣ Cần loại bỏ tất cả các t/t này ra khỏi R
- Nếu tồn tại $X \rightarrow A \in F$ sao cho VP và VT của nó chứa tất cả các t/t của R
 - ▣ Kết quả phép tách chính là LDQH s
- Ngược lại, $R_i = \{XA\}$ là một thành phần của phép tách

Ví dụ

- Cho $R = \{C, T, H, R, S, G\}$
- $F = \{C \rightarrow T, CS \rightarrow G, HR \rightarrow C, HT \rightarrow R, HS \rightarrow R\}$
- Khi đó $\varphi = [R1, R2, R3, R4, R5]$
- $R1 = \{CT\}; R2 = \{CHR\}; R3 = \{THR\}; R4 = \{CSG\};$
 $R5 = \{HRS\}$

Phép tách bảo toàn phụ thuộc và không tổn thất thông tin về các LDQH dạng 3NF

- Input: $s = \langle R, F \rangle$
- Output: Một phép tách bảo toàn phụ thuộc và không tổn thất thông tin sao cho mỗi LDQH chiếu có dạng chuẩn 3NF tương ứng với tập phụ thuộc chiếu của F trên nó

Phương pháp tách về lược đồ dạng chuẩn 3NF có bảo toàn phụ thuộc và không tổn thất thông tin

- B1: Xác định các t/t của R không có mặt trong VP và VT của các PTH (gọi là A)
 - ▣ Tạo ra LDQH chiếu trên các t/t A
 - ▣ Loại bỏ A; $R = R \setminus \{A\}$
- B2: Nếu tồn tại $X \rightarrow A \in F$ sao cho VP và VT của nó chứa tất cả các t/t của R, \Rightarrow kết quả phép tách $\phi[X, A]$
- B3: Với mọi $X \rightarrow A \in F$ trong đó t/t A là t/t đơn, $R_i = \{XA\}$
- B4: Nếu có 1 số PTH cùng VT: $X \rightarrow A_1 \in F, X \rightarrow A_2 \in F, \dots, X \rightarrow A_k \in F$ thì có thể hợp thành dạng: $R_j = \{XA_1A_2 \dots A_n\}$
- B5: Nếu các t/t của khóa K không xuất hiện trong tập R_j , khi đó 1 thành phần của phép tách sẽ được định nghĩa bởi khóa K
- B6: phép tách $\phi[R_1, R_2, \dots, R_p]$ bảo toàn phụ thuộc và không tổn thất thông tin

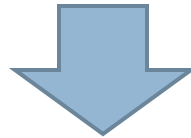
Ví dụ

- Cho $R=\{X,Y,Z,W,Q\}$
- $F=\{X\rightarrow Y, XZ\rightarrow W, YW\rightarrow Q\}$
- B1: XZ là khóa: $(XZ)^+ = R^+ = \{XYZWQ\}$
- B2: thực hiện phép tách:
 - ▣ Không tồn tại các t/t không xuất hiện trong các vế của PTH
 - ▣ Không tồn tại PTH chứa các t/t còn lại của R
 - ▣ $X\rightarrow Y$: $R1=\{XY\}$
 - ▣ $XZ\rightarrow W$: $R2=\{X,Z,W\}$
 - ▣ $YW\rightarrow Q$: $R3=\{Y,W,Q\}$
 - ▣ Khóa $XZ \subseteq R2 = \{X,Z,W\}$
- B3: Vậy phép tách $\phi[XY, XZW, YWQ]$ bảo toàn phụ thuộc
- B4: Kiểm tra tổn thất thông tin

Ví dụ (tt)

□ B4 (tt)

	X	Y	Z	W	Q
R1	a	a	b	b	b
R2	a	b	a	a	b
R3	b	a	b	a	a



	X	Y	Z	W	Q
R1	a	a	b	b	b
R2	a	a	a	a	a
R3	b	a	b	a	a

Phép tách về dạng chuẩn BC không tổn thất

- Input: $s = \langle R, F \rangle$
- Output: Một phép tách không tổn thất thông tin sao cho mỗi LDQH chiếu có dạng chuẩn BCNF tương ứng với tập phụ thuộc chiếu của F trên nó

Phương pháp tách về dạng chuẩn BCNF không tổn thất

□ Tách LDQH s thành 2 LDQH

▣ LDQH 1:

- Chọn bất kỳ $X \rightarrow A \in F^+$ sao cho X không là khóa và $A \notin X$.
- Khi đó LDQH có tập các t/t XA sẽ có dạng chuẩn BCNF và PTH $X \rightarrow A$ sẽ thỏa trên nó.

▣ LDQH 2:

- Tập các t/t $R \setminus \{A\}$
- Tiếp tục tách $R \setminus \{A\}$ cho đến khi thỏa điều kiện BCNF

Phụ thuộc đa trị MVD

- $X \twoheadrightarrow Y$ (X xác định đa trị Y , Y phụ thuộc đa trị vào X) khi và chỉ khi mỗi một giá trị của X xác định một tập giá trị của Y mà không liên quan gì đến các t/t còn lại $R \setminus X \setminus Y$
- $s_1[X] = s_2[X] = r_1[X] = r_2[X]$
- $s_1[Y] = r_1[Y]$ và $s_1[R \setminus X \setminus Y] = r_2[R \setminus X \setminus Y]$
- $s_2[Y] = r_2[Y]$ và $s_2[R \setminus X \setminus Y] = r_1[R \setminus X \setminus Y]$

Ví dụ về MVD

- $R = \{C, T, H, R, S, G\}$
- $C \twoheadrightarrow HR$ là MVD

C	T	H	R	S	G
LTM	Sơn	M9	222	Loan	8
LTM	Sơn	W9	333	Loan	8
LTM	Sơn	F9	222	Loan	8
LTM	Sơn	M9	222	Tùng	7
LTM	Sơn	W9	333	Tung	7
LTM	Sơn	F9	222	Tùng	7

Hệ tiên đề MVD

- Tính bù: Nếu $X \twoheadrightarrow Y \Rightarrow X \twoheadrightarrow (R \setminus X \setminus Y)$
- Tính tăng trưởng: Nếu $X \twoheadrightarrow Y$ và $V \subseteq W \subseteq Q \Rightarrow XW \twoheadrightarrow YV$
- Tính bắt cầu: Nếu $X \twoheadrightarrow Y$ và $Y \twoheadrightarrow Z \Rightarrow X \twoheadrightarrow Z \setminus Y$
- Các tính chất mở rộng:
 - Nếu $X \rightarrow Y \Rightarrow X \twoheadrightarrow Y$
 - Nếu $X \twoheadrightarrow Y$ và $Z \subseteq Y, W \cap Y = \emptyset, W \rightarrow Z \Rightarrow X \rightarrow Z$

Các quy tắc suy dẫn bổ sung cho MVD

- Quy tắc hợp: Nếu $X \twoheadrightarrow Y$ và $X \twoheadrightarrow Z$ thì $X \twoheadrightarrow YZ$
- Quy tắc tựa bắc cầu: Nếu $X \twoheadrightarrow Y$ và $WY \twoheadrightarrow Z$ thì $WX \twoheadrightarrow (Z \setminus WY)$
- Quy tắc bắc cầu trộn: Nếu $X \twoheadrightarrow Y$, $XY \twoheadrightarrow Z$ thì $X \twoheadrightarrow (Z \setminus Y)$
- Quy tắc phân rã: Nếu $X \twoheadrightarrow Y$ và $X \twoheadrightarrow Z$ thì $X \twoheadrightarrow (Z \cap Y)$, $X \twoheadrightarrow (Y \setminus Z)$ và $X \twoheadrightarrow (Z \setminus Y)$

Bao đóng MVD

- $X \twoheadrightarrow Y$ có thuộc F^+ không? \Rightarrow xác định cơ sở phụ thuộc của X và kiểm tra xem $Y \setminus X$ có phải là hợp của một số t/t trong cơ sở phụ thuộc của X hay không.
- Ví dụ: $C \twoheadrightarrow CTSG$ là một phần tử của D (tập phụ thuộc đa trị), vì cơ sở phụ thuộc của C là $\{T, HR, SG\}$ và $\{CTSG \setminus C\} = \{TSG\}$ là hợp của T và SG .

Thuật toán tính cơ sở phụ thuộc

- Input: $R, M \subseteq D$ (tập các MVD), $X \subseteq R$;
- Output: tính cơ sở phụ thuộc của X ứng với M tập các phụ thuộc đa trị
- Phương pháp:
 - ▣ Bắt đầu bằng tập $S = R \setminus X$
 - ▣ Lặp lại các phụ thuộc $V \twoheadrightarrow W \in M$ và một tập $Y \subseteq S$ sao cho:
 - $Y \cap W \neq \emptyset, Y \cap V = \emptyset$
 - ▣ Dừng khi S không thay đổi
 - ▣ Thay Y bằng $Y \cap W, (Y \setminus W)$ vào S
 - ▣ Tập S là cơ sở phụ thuộc của X

Dạng chuẩn 4 – 4NF

- $S = \langle R, D \rangle$ là dạng 4NF khi và chỉ khi $X \twoheadrightarrow Y \in D$ với $XY \not\subseteq R$ thì khi đó X là một siêu khóa của s
- Ví dụ: Cho $s = \langle R, D \rangle$ với $R = \{ABCEG\}$, $D = \{A \twoheadrightarrow BCEG\}$
 - ▣ Khi đó s ở dạng chuẩn 4NF. Vì các MVD $A \twoheadrightarrow B \in D^+$, $A \twoheadrightarrow C \in D^+$, $A \twoheadrightarrow E \in D^+$, $A \twoheadrightarrow G \in D^+$ đều thỏa các điều kiện của định nghĩa dạng chuẩn 4NF

Phép tách LDQH về dạng chuẩn 4NF

- Input: $s = \langle R, D \rangle$
- Output: Một phép tách không tổn thất thông tin sao cho mỗi LDQH chiếu có dạng chuẩn 4NF tương ứng với tập phụ thuộc chiếu của D trên nó
- Phương pháp (xem chi tiết trong giáo trình)

Dạng chuẩn 5NF

- Phụ thuộc kết nối: $\{R_1, R_2, \dots, R_n\}$ là kết nối các quan hệ chiếu r trên các tập R_1, R_2, \dots, R_n và các bộ thu được đều thuộc r
- LDQH dạng chuẩn 5NF: khi và chỉ khi tất cả các phụ thuộc kết nối được thực hiện trên khóa của LDQH

Dạng chuẩn DKMF

- Domain Key NF: khi và chỉ khi mỗi một ràng buộc trong QH là kết quả logic của các ràng buộc miền