

Bài 8

Chuẩn hóa cơ sở dữ liệu quan hệ

Nội dung trình bày

- Chất lượng thiết kế csdl quan hệ
- Phụ thuộc hàm
- Các dạng chuẩn
- Chuẩn hóa quan hệ

Chất lượng thiết kế csdl quan hệ (1)

- Thước đo nào để đánh giá chất lượng của csdl quan hệ được thiết kế theo phương pháp phân tích (top - down)?
 - Ý nghĩa của lược đồ quan hệ dễ hiểu, ngữ nghĩa của thuộc tính rõ ràng.
 - Quan hệ không lưu trữ trùng lặp thông tin.
 - Quan hệ không lưu trữ nhiều giá trị null.
 - Không tạo ra các bộ dữ liệu ảo.

DEPT_PROJ có thuộc tính với ngữ nghĩa rõ ràng nhưng khó giải thích ý nghĩa của quan hệ, lưu trữ trùng lặp thông tin gây ra các dị thường khi cập nhật dữ liệu.

DEPT_PROJ

Dnumber	Dname	<u>Dlocation</u>	<u>Pnumber</u>	Pname	Plocation
5	Research	Bellaire	1	ProductX	Bellaire
5	Research	Houston	3	ProductZ	Houston
4	Administration	Stafford	10	Computerization	Stafford
4	Administration	Stafford	30	Newbenefits	Stafford
1	Headquarters	Houston	20	Reorganization	Houston

3

Chất lượng thiết kế csdl quan hệ (2)

DEPT

<u>Dnumber</u>	<u>Location</u>	Dname
5	Bellaire	Research
5	Houston	Research
4	Stafford	Administration
1	Houston	Headquarters

PROJ

<u>Pnumber</u>	Pname	Location
1	ProductX	Bellaire
3	ProductZ	Houston
10	Computerization	Stafford
20	Reorganization	Houston

*DEPT * PROJ tạo dữ liệu không chính xác do liên kết trên thuộc tính không phù hợp.*

Dnumber	Location	Dname	Pnumber	Pname	Location
5	Bellaire	Research	1	ProductX	Bellaire
5	Houston	Research	3	ProductZ	Houston
5	Houston	Research	20	Reorganization	Houston
4	Stafford	Administration	10	Computerization	Stafford
1	Houston	Headquarters	3	ProductZ	Houston

4

Phụ thuộc hàm (1)

- **Phụ thuộc hàm** là ràng buộc giữa hai tập con của tập thuộc tính của một lược đồ quan hệ, xác định tính duy nhất các giá trị của một số thuộc tính cụ thể.

- Ví dụ:

- $\{Dnumber\} \rightarrow \{Dname\}$ biểu diễn ràng buộc là với mỗi thể hiện của DEPT nếu 2 bộ có cùng giá trị Dnumber thì cũng phải có cùng giá trị Dname.
- $\{Dnumber, Location\} \rightarrow \{Dname\}$,
 $\{Dname\} \rightarrow \{Dnumber\}$.

DEPT

<u>Dnumber</u>	<u>Location</u>	Dname
5	Bellaire	Research
5	Houston	Research
4	Stafford	Administration
1	Houston	Headquarters

- Cơ sở để xây dựng tiêu chuẩn hình thức đánh giá chất lượng của lược đồ quan hệ.

5

Phụ thuộc hàm (2)

- Định nghĩa phụ thuộc hàm: cho X, Y là các tập con của tập thuộc tính của lược đồ quan hệ R ; khi đó, phụ thuộc hàm $X \rightarrow Y$ được xem là đúng trong R nếu và chỉ nếu, mỗi khi có hai bộ của R trùng nhau về giá trị trên X , thì chúng cũng trùng nhau về giá trị trên Y .

- X được gọi là vế xác định (vế trái), và Y là vế phụ thuộc (vế phải).
- Pth này có thể được đọc là “ X xác định hàm Y ”, hoặc “ Y phụ thuộc hàm vào X ”, hay đơn giản là “ X mũi tên Y ”.
- Pth $X \rightarrow Y$ là hiển nhiên nếu và chỉ nếu $Y \subseteq X$.

- Biểu diễn trên lược đồ:

DEPT

<u>Dnumber</u>	<u>Location</u>	Dname

6

Hệ tiên đề Armstrong

- Hệ tiên đề Armstrong là các luật suy diễn cho phép xác định thêm các pth từ một tập pth cho trước của lược đồ quan hệ R .
 - Luật phản xạ: nếu $Y \subseteq X$ thì $X \rightarrow Y$
 - Luật tăng trưởng: nếu $X \rightarrow Y$ thì $XZ \rightarrow YZ$ (XZ viết tắt của $X \cup Z$)
 - Luật bắc cầu: nếu $X \rightarrow Y$ và $Y \rightarrow Z$ thì $X \rightarrow Z$
- Hệ quả của các tiên đề Armstrong
 - Luật phân rã: nếu $X \rightarrow YZ$ thì $X \rightarrow Y$ và $X \rightarrow Z$
 - Luật hợp: nếu $X \rightarrow Y$ và $X \rightarrow Z$ thì $X \rightarrow YZ$
 - Luật bắc cầu giả: nếu $X \rightarrow Y$ và $WY \rightarrow Z$ thì $WX \rightarrow Z$
- *Ví dụ: cho lược đồ quan hệ $R(A, B, C, D)$ và $F = \{A \rightarrow B, B \rightarrow H\}$ là một tập pth đúng trên R . Từ các pth $A \rightarrow B, B \rightarrow H$, áp dụng luật bắc cầu thì $A \rightarrow H$ cũng là pth đúng trên R .*

7

Các dạng chuẩn (1)

- Dạng chuẩn là một biểu diễn chuẩn cho dữ liệu dưới dạng quan hệ đảm bảo
 - Ý nghĩa của quan hệ dễ hiểu.
 - Hạn chế lưu trữ trùng lặp thông tin.
 - Không tạo ra các bộ dữ liệu ảo.
- Các dạng chuẩn:
 - Dạng chuẩn 1 (1NF).
 - Dạng chuẩn 2 (2NF).
 - Dạng chuẩn 3 (3NF).
 - Dạng chuẩn Boy - Codd (BCNF).
 - Dạng chuẩn 4 (4NF).
 - Dạng chuẩn 5 (5NF).

8

Các dạng chuẩn (2)

- **Siêu khóa:** cho SK là một tập con của tập thuộc tính của lược đồ quan hệ R . Khi đó, SK được gọi là một siêu khóa của R nếu và chỉ nếu không tồn tại một thể hiện hợp lệ nào của R chứa hai bộ khác nhau có cùng giá trị trên tập thuộc tính SK .
- **Khóa:** cho K là một tập con của tập thuộc tính của lược đồ quan hệ R . Khi đó, K được gọi là một khóa của R nếu và chỉ nếu K thỏa mãn đồng thời hai tính chất sau:
 - K là siêu khóa của R .
 - Không tồn tại tập con thực sự nào của K cũng là siêu khóa của R .
- Một thuộc tính của lược đồ quan hệ R được gọi là **thuộc tính khóa** nếu và chỉ nếu nó là một phần của một khóa nào đó của R ; ngược lại, được gọi là **thuộc tính không khóa**.

9

Dạng chuẩn 1 (1)

- **Định nghĩa 1NF:** một quan hệ R là dạng chuẩn 1 (1NF) nếu và chỉ nếu, mỗi bộ của R có đúng một giá trị cho mỗi thuộc tính.

DEPT_PROJ không là 1NF do Dlocation có nhiều hơn một giá trị hay Project có giá trị dạng bảng.

DEPT_PROJ

<u>Dnumber</u>	Dname	Dlocation	Project		
			Pnumber	Pname	Plocation
5	Research	{Bellaire, Houston}	1	ProductX	Bellaire
			3	ProductZ	Houston
4	Administration	Stafford	10	Computerization	Stafford
			30	Newbenefits	Stafford
1	Headquarters	Houston	20	Reorganization	Houston

10

Dạng chuẩn 1 (2)

DEPT_PROJ là 1NF nhưng có lưu trữ trùng lặp thông tin.

DEPT_PROJ

Dnumber	Dname	<u>Dlocation</u>	<u>Pnumber</u>	Pname	Plocation
5	Research	Bellaire	1	ProductX	Bellaire
5	Research	Houston	3	ProductZ	Houston
4	Administration	Stafford	10	Computerization	Stafford
4	Administration	Stafford	30	Newbenefits	Stafford
1	Headquarters	Houston	20	Reorganization	Houston

- Do đặc tính của quan hệ (không có thuộc tính đa trị và thuộc tính gộp), mọi quan hệ đều là dạng chuẩn 1.

11

Dạng chuẩn 2 (1)

- Định nghĩa phụ thuộc hàm tối giản:** pth $X \rightarrow Y$ được gọi là tối giản nếu và chỉ nếu nó đúng trong R và không tồn tại một tập con thực sự của X là X' sao cho pth $X' \rightarrow Y$ cũng đúng trong R .
- Định nghĩa 2NF:** một quan hệ R là dạng chuẩn 2 (2NF) nếu và chỉ nếu với mỗi khóa K của R và với mỗi thuộc tính không khóa A của R , pth $K \rightarrow \{A\}$ là tối giản.

DEPT_PROJ

	Dnumber	Dname	<u>Dlocation</u>	<u>Pnumber</u>	Pname	Plocation
f_1	↑	↑			↑	↑
f_2		↑				
f_3	↑					
f_4				↑	↑	↑
f_5				↑		
f_6	↑					

DEPT_PROJ có khóa là $\{Dlocation, Pnumber\}$; không là 2NF, do pth $\{Dlocation, Pnumber\} \rightarrow \{Dnumber\}$ không tối giản.

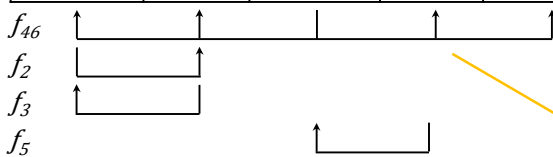
12

Dạng chuẩn 2 (2)

- **Định nghĩa 2NF tương đương:** một quan hệ R là dạng chuẩn 2 (2NF) nếu và chỉ nếu với mỗi pth không hiển nhiên $X \rightarrow Y$ đúng trong R , ít nhất một trong các điều sau đúng
 - X là một siêu khóa của R .
 - Y là một tập con thực sự của một khóa của R .
 - X không là một tập con thực sự của một khóa của R .

DEPT_PROJ1

Dnumber	Dname	<u>Pnumber</u>	Pname	Plocation
---------	-------	----------------	-------	-----------



DEPT_LOCATIONS

<u>Dnumber</u>	<u>Dlocation</u>
----------------	------------------

DEPT_PROJ1 có 2 khóa là Pnumber và Pname; là 2NF do các pth có vế trái là khóa và vế phải là thuộc tính không khóa đều tối giản.

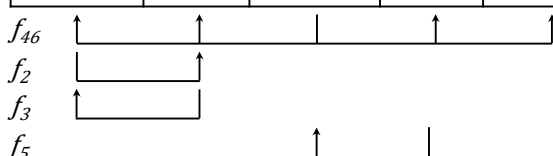
13

Dạng chuẩn 3 (1)

- **Định nghĩa 3NF:** một quan hệ R là dạng chuẩn 3 (3NF) nếu và chỉ nếu với mỗi pth không hiển nhiên $X \rightarrow Y$ đúng trong R , ít nhất một trong các điều sau đúng
 - X là một siêu khóa của R .
 - Y là một tập con thực sự của một khóa của R .

DEPT_PROJ1

Dnumber	Dname	<u>Pnumber</u>	Pname	Plocation
---------	-------	----------------	-------	-----------



DEPT_PROJ1 không là 3NF do các pth f_2, f_3 có vế trái không là siêu khóa và vế phải cũng không là tập con thực sự của một khóa.

14

Dạng chuẩn 3 (2)

DEPT_PROJ1

Dnumber	Dname	Pnumber	Pname	Plocation
5	Research	1	ProductX	Bellaire
5	Research	3	ProductZ	Houston
4	Administration	10	Computerization	Stafford
4	Administration	30	Newbenefits	Stafford
1	Headquarters	20	Reorganization	Houston

DEPT_PROJ1 vẫn có sự trùng lặp thông tin.

DEPARTMENT

<u>Dnumber</u>	Dname
f_2	
f_3	

PROJECT

Dnumber	<u>Pnumber</u>	Pname	Plocation
f_{46}			
f_5			

DEPARTMENT, PROJECT là 3NF.

15

Dạng chuẩn Boyce-Codd (1)

- **Định nghĩa BCNF:** một quan hệ R là dạng chuẩn Boyce-Codd (BCNF) nếu và chỉ nếu với mỗi pth không hiển nhiên $X \rightarrow Y$ đúng trong R, điều sau đúng
 - X là một siêu khóa của R.
- **Ví dụ:** mỗi dự án cũng cần lưu trữ tổng chi phí, xếp hạng. Biết rằng không có 2 dự án cùng hạng được triển khai cùng một địa điểm, xếp hạng của dự án dựa trên tổng chi phí.

PROJ_COST_RANK

Dnumber	<u>Pnumber</u>	Pname	Plocation	Pcost	Prank
f_{46}					
f_5					
f_7					
f_8					

PROJ_COST_RANK có 3 khóa là Pnumber, Pname và {Plocation, Prank}; là 3NF nhưng không là BCNF do pth f_8 có vế trái không là siêu khóa.

16

Dạng chuẩn Boyce-Codd (2)

PROJ_COST_RANK

Dnumber	<u>Pnumber</u>	Pname	Plocation	Pcost	Prank
5	1	ProductX	Bellaire	400	B
5	3	ProductZ	Houston	400	B
4	10	Computerization	Stafford	200	C
4	30	Newbenefits	Stafford	150	C
1	20	Reorganization	Houston	600	A

PROJ_COST_RANK vẫn có sự trùng lặp thông tin.

PROJ_COST

Dnumber	<u>Pnumber</u>	Pname	Plocation	Pcost
---------	----------------	-------	-----------	-------

f_{46} ↑
 f_5 ↑

COST_RANK

<u>Pcost</u>	Prank
--------------	-------

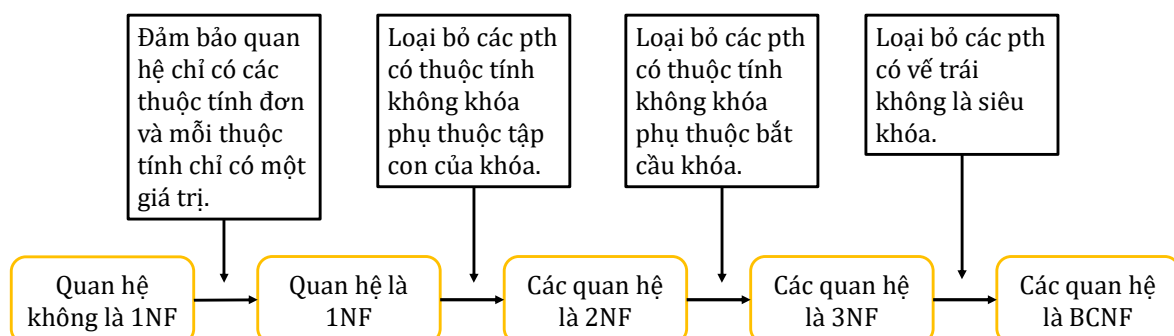
f_7 ↑

PROJ_COST, COST_RANK là BCNF.

17

Chuẩn hóa quan hệ

- Chuẩn hóa** là quá trình biến đổi một lược đồ quan hệ thành một tập hợp lược đồ quan hệ có cấu trúc tốt hơn (mục tiêu là 3NF, BCNF) tương đương về biểu diễn dữ liệu.



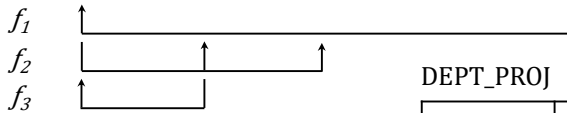
18

Chuẩn hóa về 1NF (1)

- Loại bỏ dạng đa trị và dạng gộp của thuộc tính.

DEPT_PROJ

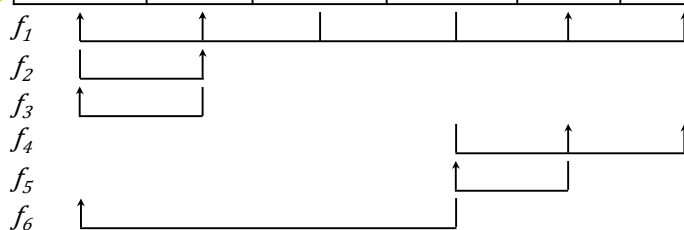
<u>Dnumber</u>	Dname	Dlocation	Project		
			Pnumber	Pname	Plocation



DEPT_PROJ là 1NF sau khi loại bỏ dạng đa trị và dạng gộp của các thuộc tính Dlocation, Project

DEPT_PROJ

<u>Dnumber</u>	Dname	<u>Dlocation</u>	<u>Pnumber</u>	Pname	Plocation
----------------	-------	------------------	----------------	-------	-----------



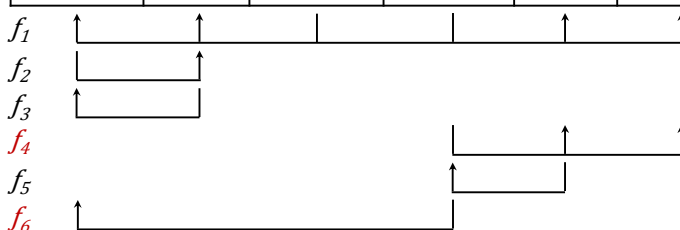
19

Chuẩn hóa về 2NF (1)

- Phân tách quan hệ là 1NF như sau:
 - Với mỗi pth có thuộc tính không khóa phụ thuộc tập con của khóa, tạo một quan hệ với tập thuộc tính gồm vế trái và vế phải của pth.
 - Tạo một quan hệ với tập thuộc tính gồm các thuộc tính còn lại.
 - Đảm bảo các quan hệ được liên kết bởi khóa ngoại phù hợp.

DEPT_PROJ

<u>Dnumber</u>	Dname	<u>Dlocation</u>	<u>Pnumber</u>	Pname	Plocation
----------------	-------	------------------	----------------	-------	-----------



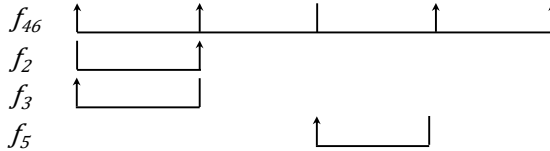
DEPT_PROJ là 1NF nhưng không là 2NF, do pth f_4 , f_6 có thuộc tính không khóa phụ thuộc tập con của khóa.

20

Chuẩn hóa về 2NF (2)

DEPT_PROJ1

Dnumber	Dname	<u>Pnumber</u>	Pname	Plocation
---------	-------	----------------	-------	-----------



DEPT_LOCATIONS

<u>Dnumber</u>	<u>Dlocation</u>
----------------	------------------

Quan hệ DEPT_PROJ được biến đổi thành 2 quan hệ DEPT_PROJ1 và DEPT_LOCATIONS là 2NF. Hơn nữa, DEPT_LOCATIONS là BCNF.

21

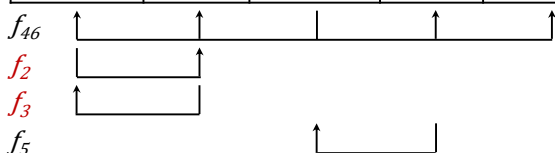
Chuẩn hóa về 3NF (1)

▪ Phân tách quan hệ là 2NF như sau:

- Với mỗi pth có thuộc tính không khóa phụ thuộc bất cầu khóa, tạo một quan hệ với tập thuộc tính gồm vế trái và vế phải của pth.
- Tạo một quan hệ với tập thuộc tính gồm các thuộc tính còn lại.
- Đảm bảo các quan hệ được liên kết bởi khóa ngoại phù hợp.

DEPT_PROJ1

Dnumber	Dname	<u>Pnumber</u>	Pname	Plocation
---------	-------	----------------	-------	-----------



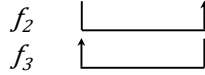
DEPT_PROJ1 là 2NF nhưng không là 3NF, do pth f_2, f_3 có thuộc tính không khóa phụ thuộc bất cầu khóa.

22

Chuẩn hóa về 3NF (2)

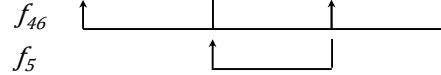
DEPARTMENT

<u>Dnumber</u>	Dname
----------------	-------



PROJECT

<u>Dnumber</u>	<u>Pnumber</u>	Pname	Plocation
----------------	----------------	-------	-----------



Quan hệ DEPT_PROJ1 được biến đổi thành 2 quan hệ DEPARTMENT, PROJECT là 3NF (và cũng là BCNF).
 Kết luận, quan hệ DEPT_PROJ được biến đổi thành 3 quan hệ DEPT_LOCATIONS, DEPARTMENT, PROJECT là 3NF (và cũng là BCNF).

23

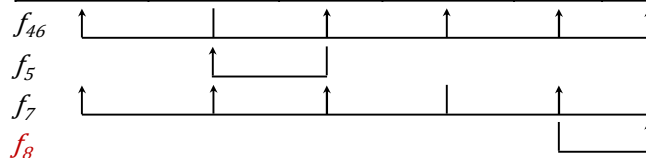
Chuẩn hóa về BCNF (1)

▪ Phân tách quan hệ là 3NF như sau:

- Với mỗi pth có vế trái không là siêu khóa, tạo một quan hệ với tập thuộc tính gồm vế trái và vế phải của pth.
- Tạo một quan hệ với tập thuộc tính gồm các thuộc tính còn lại.
- Đảm bảo các quan hệ được liên kết bởi khóa ngoại phù hợp.

PROJ_COST_RANK

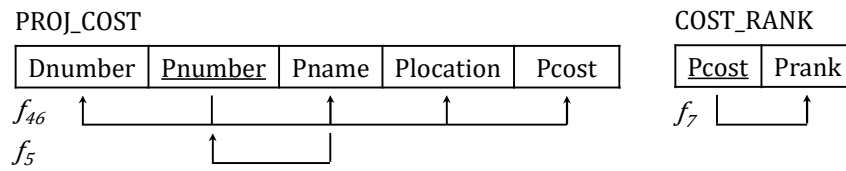
<u>Dnumber</u>	<u>Pnumber</u>	Pname	Plocation	Pcost	Prank
----------------	----------------	-------	-----------	-------	-------



PROJ_COST_RANK là 3NF nhưng không là BCNF do pth f_8 có vế trái không là siêu khóa.

24

Chuẩn hóa về BCNF (2)



Quan hệ PROJ_COST_RANK được biến đổi thành 2 quan hệ PROJ_COST, COST_RANK là BCNF.