

Chương 7

Phụ thuộc hàm và Chuẩn hóa cơ sở dữ liệu quan hệ

Nội dung trình bày

- Thiết kế cơ sở dữ liệu
 - Phụ thuộc hàm
 - Các dạng chuẩn dựa trên khóa chính
 - Tổng quát cho dạng chuẩn 2 và 3
 - Dạng chuẩn Boy - Codd
-

Thiết kế cơ sở dữ liệu (1)

- Phương pháp dưới - lên (thiết kế tổng hợp)
 - Xem các mối quan hệ cơ bản giữa các thuộc tính riêng lẻ như là điểm khởi đầu và sử dụng các mối quan hệ đó để xây dựng các lược đồ quan hệ.
- Phương pháp trên - xuống (thiết kế theo phân tích)
 - Bắt đầu với một số nhóm của các thuộc tính thành các quan hệ tồn tại với nhau một cách tự nhiên.
 - Ví dụ: hóa đơn, biểu mẫu, báo cáo
 - Các quan hệ được phân tích riêng lẻ và cùng với nhau, dẫn đến sự phân tách sâu hơn cho đến khi có được tất cả các thuộc tính mong muốn.

Thiết kế cơ sở dữ liệu (2)

- Làm thế nào để thiết kế cơ sở dữ liệu với các lược đồ quan hệ tốt?
- Ví dụ một thiết kế quan hệ khó giải thích ý nghĩa của nó và trùng lặp thông tin

			Redundancy	Redundancy	
EMP_PROJ					
Ssn	Pnumber	Hours	Ename	Pname	Plocation
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford

Thiết kế cơ sở dữ liệu (3)

- Ví dụ một thiết kế tạo ra bộ giả

EMP_LOCS

Ename	Plocation
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford

EMP_PROJ1

Ssn	Pnumber	Hours	Pname	Plocation
123456789	1	32.5	ProductX	Bellaire
123456789	2	7.5	ProductY	Sugarland
666884444	3	40.0	ProductZ	Houston
453453453	1	20.0	ProductX	Bellaire
453453453	2	20.0	ProductY	Sugarland
333445555	2	10.0	ProductY	Sugarland
333445555	3	10.0	ProductZ	Houston
333445555	10	10.0	Computerization	Stafford

EMP_LOCS * EMP_PROJ1

Ssn	Pnumber	Hours	Pname	Plocation	Ename
123456789	1	32.5	ProductX	Bellaire	Smith, John B.
* 123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith, John B.
* 123456789	2	7.5	ProductY	Sugarland	English, Joyce A.
* 123456789	2	7.5	ProductY	Sugarland	Wong, Franklin T.
666884444	3	40.0	ProductZ	Houston	Narayan, Ramesh K.
* 666884444	3	40.0	ProductZ	Houston	Wong, Franklin T.
* 453453453	1	20.0	ProductX	Bellaire	Smith, John B.

Thiết kế cơ sở dữ liệu (4)

- Thiết kế một quan hệ sao cho có thể dễ dàng giải thích ý nghĩa của nó, ngữ nghĩa của các thuộc tính phải dễ hiểu.
 - Không kết hợp các thuộc tính từ nhiều kiểu thực thể và kiểu mối kết hợp vào cùng một quan hệ.
 - Nếu một quan hệ tương ứng với một kiểu thực thể hoặc một kiểu mối kết hợp thì sẽ dễ dàng giải thích ý nghĩa của nó. Ngược lại, nếu một quan hệ tương ứng với một hỗn hợp của nhiều thực thể và mối kết hợp, sẽ dẫn đến sự mơ hồ về ngữ nghĩa và quan hệ không thể dễ dàng giải thích được.
- Thiết kế các quan hệ cơ sở để không có các dị thường cập nhật. Nếu có các dị thường, hãy ghi lại chúng một cách rõ ràng và đảm bảo rằng các chương trình cập nhật cơ sở dữ liệu sẽ hoạt động chính xác.

Thiết kế cơ sở dữ liệu (5)

- Tránh đặt các thuộc tính trong một quan hệ cơ sở mà giá trị của nó thường là NULL. Nếu NULL là không thể tránh khỏi, đảm bảo rằng nó không áp dụng cho phần lớn các bộ trong quan hệ.
 - Thiết kế các quan hệ để có thể thực hiện phép kết với điều kiện so sánh bằng trên các thuộc tính là các cặp (khóa chính, khóa ngoại) liên quan thích hợp theo cách đảm bảo rằng không có bộ giả nào được tạo ra. Tránh các quan hệ có chứa các thuộc tính để thực hiện phép kết không phải là tổ hợp (khóa ngoại, khóa chính) vì việc kết trên các thuộc tính đó có thể tạo ra bộ giả.
-

Phụ thuộc hàm

- Các phụ thuộc hàm là các ràng buộc bắt nguồn từ ý nghĩa và mối quan hệ qua lại của 2 tập thuộc tính trong cơ sở dữ liệu.
 - Các phụ thuộc hàm được sử dụng
 - Làm thước đo về "tính tốt" của chất lượng thiết kế các quan hệ.
 - Là yếu tố chính để xác định các dạng chuẩn cho các quan hệ.
-

Định nghĩa phụ thuộc hàm

- Một phụ thuộc hàm giữa hai tập thuộc tính X và Y của quan hệ R xác định một ràng buộc đối với các bộ trong tất cả các thể hiện quan hệ $r(R)$, đó là:
 - Với hai bộ bất kỳ t_1 và t_2 trong r mà có $t_1[X] = t_2[X]$, cũng phải có $t_1[Y] = t_2[Y]$.
 - Kí hiệu là: $X \rightarrow Y$. X là vế trái, Y là vế phải của phụ thuộc hàm.
 - Ví dụ về phụ thuộc hàm
 - Mã nhân viên xác định tên nhân viên: $MaNV \rightarrow TenNV$
 - Mã dự án xác định tên dự án và địa điểm: $MaDA \rightarrow \{TenDA, DiaDiem\}$
 - Nếu K là khóa của R thì K xác định hàm tất cả các thuộc tính của R.
-

Xác định phụ thuộc hàm (1)

- Để xác định PTH, chúng ta cần hiểu ý nghĩa của các thuộc tính và mối quan hệ giữa chúng.
 - Một PTH là một tính chất (của các thuộc tính) của quan hệ R.
 - Với một thể hiện của quan hệ
 - Chúng ta chỉ có thể kết luận là một PTH *có thể tồn tại* giữa các thuộc tính nào đó.
 - Chúng ta chắc chắn kết luận là một số PTH *nhất định không tồn tại* bởi vì có những bộ cho thấy sự vi phạm những phụ thuộc đó.
-

Xác định phụ thuộc hàm (2)

- Với thể hiện của quan hệ NHANVIEN_DUAN

NHANVIEN_DUAN

MaNV	TenNV	MaDA	TenDA	DiaDiem	SoGio
123456789	Hung	1	Du an X	Tan Binh	8
123456789	Hung	2	Du an Y	Phu Nhuan	4
333445555	Nghia	1	Du an X	Tan Binh	4
444556666	Bao	3	Du an Z	Go Vap	8

- Các PTH sau có thể tồn tại: $\text{MaNV} \rightarrow \text{TenNV}$, $\text{MaDA} \rightarrow \text{TenDA}$, $\{\text{MaNV}, \text{MaDA}\} \rightarrow \text{SoGio}$, $\text{TenDA} \rightarrow \text{DiaDiem}$
- Các PTH sau chắc chắn không tồn tại: $\text{MaNV} \rightarrow \{\text{MaDA}, \text{TenDA}, \text{DiaDiem}\}$, $\text{MaNV} \rightarrow \text{SoGio}$.

Luật dẫn Armstrong

- Cho F là một tập các PTH tồn tại trên quan hệ R. Một PTH $X \rightarrow Y$ là **được suy ra từ F** nếu nó tồn tại với mọi thể hiện $r(R)$; nghĩa là, khi r thỏa tất cả các PTH trong F, $X \rightarrow Y$ cũng tồn tại trong r .
- Luật dẫn Armstrong
 - 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 $X \rightarrow Z$
- Một vài luật dẫn bổ sung:
 - 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 bất cầu giả: Nếu $X \rightarrow Y$ và $WY \rightarrow Z$ thì $WX \rightarrow Z$

Chuẩn hóa các quan hệ (1)

- Chuẩn hóa:
 - Là quá trình phân tách các quan hệ chất lượng xấu bằng cách chia nhỏ các thuộc tính thành các quan hệ nhỏ hơn.
 - Quá trình chuẩn hóa bằng cách phân tách cũng phải xác nhận sự tồn tại của các tính chất:
 - Tính chất không mất thông tin.
 - Tính bảo toàn phụ thuộc hàm.
-

Chuẩn hóa các quan hệ (2)

- Dạng chuẩn:
 - Là điều kiện sử dụng các khóa và các PTH của một quan hệ để xác nhận xem một quan hệ có đạt một dạng chuẩn cụ thể hay không.
 - Dạng chuẩn dựa trên các khóa và các PTH của một quan hệ:
 - Dạng chuẩn 2 (2NF), dạng chuẩn 3 (3NF)
 - Dạng chuẩn Boy - Codd (BCNF)
 - Dạng chuẩn dựa trên các khóa và các phụ thuộc đa trị.
 - Dạng chuẩn 4 (4NF)
 - Dạng chuẩn dựa trên các khóa và các phụ thuộc kết nối
 - Dạng chuẩn 5 (5NF)
-

Khóa và các thuộc tính khóa

- Tập thuộc tính S của quan hệ $R(A_1, A_2, \dots, A_n)$ gọi là **siêu khóa** nếu với mọi thể hiện $r(R)$ không có hai bộ khác nhau $t_1, t_2 \in r$ mà $t_1[S] = t_2[S]$.
 - Tập thuộc tính K của R gọi là **khóa** nếu nó là một siêu khóa đồng thời việc loại bỏ thuộc tính bất kỳ khỏi K sẽ khiến K không còn là một siêu khóa.
 - Khóa dự tuyển
 - Nếu một quan hệ có nhiều khóa thì mỗi khóa được gọi là khóa dự tuyển.
 - Một trong các khóa dự tuyển được chọn tùy ý làm **khóa chính**, các khóa còn lại được gọi là **khóa phụ**.
 - **Thuộc tính khóa** là thuộc tính thành viên của một khóa dự tuyển.
 - **Thuộc tính không khóa** là thuộc tính không phải thành viên của bất kỳ khóa dự tuyển nào.
-

Dạng chuẩn 1 - 1NF (1)

- **Định nghĩa 1NF:** Một quan hệ ở dạng chuẩn 1 nếu nó không có
 - Thuộc tính gộp
 - Thuộc tính đa trị
 - Quan hệ lồng - do sự kết hợp của các thuộc tính đa trị
 - Được xem là một phần của định nghĩa về quan hệ trong mô hình quan hệ.
 - Hầu hết các hệ quản trị chỉ cho phép những quan hệ ở dạng chuẩn 1.
-

Dạng chuẩn 1 - 1NF (2)

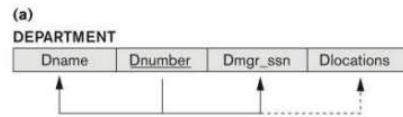


Figure 14.9
Normalization into 1NF. (a) A relation schema that is not in 1NF. (b) Sample state of relation DEPARTMENT. (c) 1NF version of the same relation with redundancy.

(b)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

(c)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Dạng chuẩn 1 - 1NF (3)

(a)

EMP_PROJ

Ssn	Ename	Proj	
		Pnumber	Hours

(b)

EMP_PROJ

Ssn	Ename	Pnumber	Hours
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya, Alicia J.	30	30.0
		10	10.0
987987987	Jabbar, Ahmad V.	10	35.0
		30	5.0
987654321	Wallace, Jennifer S.	30	20.0
		20	15.0
888665555	Borg, James E.	20	NULL

(c)

EMP_PROJ1

Ssn	Ename

EMP_PROJ2

Ssn	Pnumber	Hours

Figure 14.10
Normalizing nested relations into 1NF. (a) Schema of the EMP_PROJ relation with a *nested relation* attribute PROJ. (b) Sample extension of the EMP_PROJ relation showing nested relations within each tuple. (c) Decomposition of EMP_PROJ into relations EMP_PROJ1 and EMP_PROJ2 by propagating the primary key.

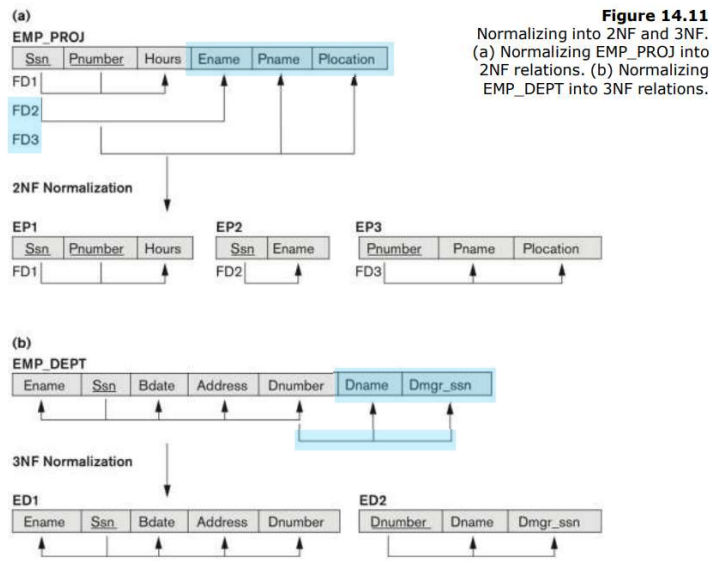
Dạng chuẩn 2 và 3 - 2NF và 3NF (1)

- Một PTH $X \rightarrow Y$ tồn tại trên quan hệ R gọi là **PTH đầy đủ** nếu loại bỏ bất kỳ thuộc tính A nào trong X thì PTH $(X - \{A\}) \rightarrow Y$ không tồn tại trên R.
 - Một PTH $X \rightarrow Y$ tồn tại trên quan hệ R gọi là **PTH bộ phận** nếu có một thuộc tính A trong X mà khi bị loại bỏ thì PTH $(X - \{A\}) \rightarrow Y$ vẫn tồn tại trên R.
 - Một PTH $X \rightarrow Y$ tồn tại trên quan hệ R gọi là **PTH bất cầu** nếu có một tập thuộc tính Z của R mà Z không là khóa hoặc là tập con của một khóa và cả hai PTH $X \rightarrow Z, Z \rightarrow Y$ cũng tồn tại trên R.
-

Dạng chuẩn 2 và 3 - 2NF và 3NF (2)

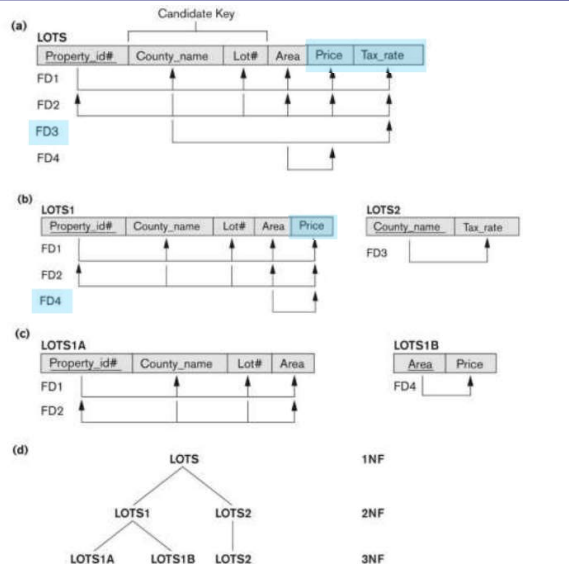
- **Định nghĩa 2NF:** Một quan hệ ở dạng chuẩn 2 nếu mọi thuộc tính không khóa của nó phụ thuộc đầy đủ vào khóa chính.
 - **Định nghĩa 3NF:** Một quan hệ ở dạng chuẩn 3 nếu nó ở dạng chuẩn 2 và không có thuộc tính không khóa nào của nó phụ thuộc bất cầu vào khóa chính.
-

Dạng chuẩn 2 và 3 - 2NF và 3NF (3)



Dạng chuẩn 2 và 3 - 2NF và 3NF (4)

Figure 14.12
Normalization into 2NF and 3NF. (a) The LOTS relation with its functional dependencies FD1 through FD4. (b) Decomposing into the 2NF relations LOTS1 and LOTS2. (c) Decomposing LOTS1 into the 3NF relations LOTS1A and LOTS1B. (d) Progressive normalization of LOTS into a 3NF design.



Chuẩn hóa 1NF, 2NF và 3NF

- Chuẩn hóa 1NF
 - Tạo các quan hệ mới cho từng thuộc tính đa trị hoặc quan hệ lồng.
 - Chuẩn hóa 2NF
 - Phân tách và thiết lập một quan hệ mới cho mỗi PTH gây ra sự vi phạm 2NF. Đảm bảo giữ một quan hệ với khóa chính ban đầu và các thuộc tính phụ thuộc đầy đủ vào nó.
 - Chuẩn hóa 3NF
 - Phân tách và thiết lập một quan hệ mới cho mỗi PTH gây ra sự vi phạm 3NF.
-

Dạng chuẩn 2 và 3 tổng quát

- **Định nghĩa 2NF tổng quát:** Một quan hệ ở dạng chuẩn 2 nếu mọi thuộc tính không khóa của nó phụ thuộc đầy đủ và mọi khóa của nó.
 - **Định nghĩa 3NF tổng quát:** Một quan hệ ở dạng chuẩn 3 nếu PTH $X \rightarrow A$ tồn tại thì
 - a) X là siêu khóa của nó, hoặc
 - b) A là thuộc tính khóa của nó.
-

Dạng chuẩn Boyce-Codd - BCNF (1)

- **Định nghĩa BCNF:** Một quan hệ ở dạng chuẩn Boyce-Codd nếu PTH $X \rightarrow A$ tồn tại thì X là siêu khóa của nó.
- Mỗi dạng chuẩn đều mạnh hơn dạng trước
 - Mọi quan hệ đạt 2NF đều đạt 1NF.
 - Mọi quan hệ đạt 3NF đều đạt 2NF.
 - Mọi quan hệ đạt BCNF đều đạt 3NF.
- Trong thực tế, hầu hết các quan hệ đạt 3NF cũng đạt BCNF. Vẫn có các quan hệ đạt 3NF nhưng không đạt BCNF.
- Mục tiêu thiết kế là các quan hệ đạt BCNF (hoặc 3NF).

Dạng chuẩn Boyce-Codd - BCNF (2)

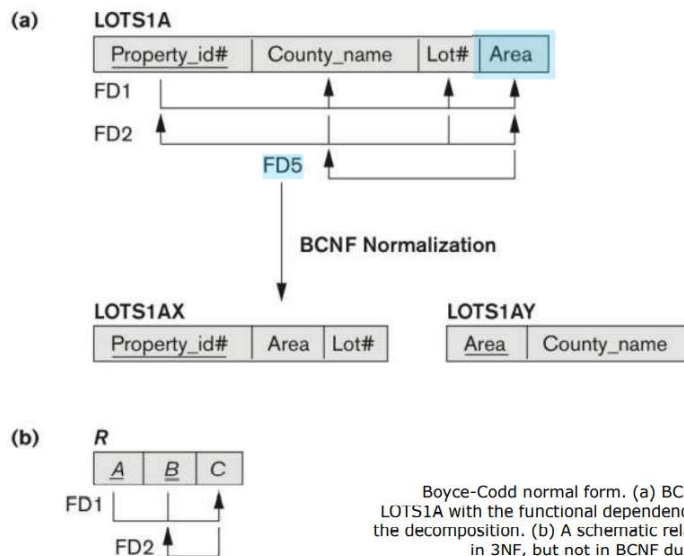


Figure 14.13
Boyce-Codd normal form. (a) BCNF normalization of LOTS1A with the functional dependency FD2 being lost in the decomposition. (b) A schematic relation with FDs; it is in 3NF, but not in BCNF due to the f.d. $C \rightarrow B$.

Chuẩn hóa BCNF

- Cho quan hệ $R(U)$ không đạt BCNF, nếu có PTH $X \rightarrow A$ gây ra sự vi phạm BCNF thì R có thể được phân tách thành 2 quan hệ
 - $R_1(U - \{A\})$ và $R_2(X \cup A)$
 - Nếu quan hệ R_1 hoặc R_2 không đạt BCNF thì lặp lại quá trình phân tách.
-