

Các dạng chuẩn của CSDL quan hệ

Lê Thành Văn

06-12-2023

Khoa Hệ thống thông tin quản lý

Giới thiệu

Giới thiệu

Định nghĩa

Chuẩn hóa cơ sở dữ liệu (csdl) là quá trình tách bảng dữ liệu nhằm:

- giảm thiểu việc dư thừa dữ liệu, và
- hạn chế lỗi có thể phát sinh trong quá trình thay đổi dữ liệu.

Việc chuẩn hóa csdl dựa trên các dạng chuẩn (normal form) được Edgar F. Codd đề ra trong mô hình quan hệ của mình.

Giới thiệu

Các loại lỗi

Lỗi khi thay đổi dữ liệu (thêm, sửa hoặc xóa) là dạng lỗi khi dữ liệu được thay đổi không tương thích với cấu trúc hoặc dữ liệu hiện có.

Giả sử thông tin giảng viên cần bao gồm mã môn học họ giảng dạy. Tuy nhiên, một giảng viên mới có thể chưa có môn, dẫn đến không thể thêm thông tin của họ.

Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201

424	Dr. Newsome	29-Mar-2007	?
-----	-------------	-------------	---

Hình 1: Lỗi khi thêm dữ liệu

Một thông tin có thể xuất hiện nhiều lần trong một bảng, nên khi thay đổi có thể bị sót, dẫn đến thông tin không nhất quán trong csdl.

Employees' Skills

Employee ID	Employee Address	Skill
426	87 Sycamore Grove	Typing
426	87 Sycamore Grove	Shorthand
519	94 Chestnut Street	Public Speaking
519	96 Walnut Avenue	Carpentry

Hình 2: Lỗi khi sửa dữ liệu

Lỗi khi xóa dữ liệu là dạng lỗi mà khi xóa một thông tin này có thể xóa những thông tin (không cần xóa) khác.

Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201



DELETE

Hình 3: Lỗi khi xóa dữ liệu

Giới thiệu

Các dạng chuẩn

Các dạng chuẩn thường gặp bao gồm (từ thấp đến cao):

- Dạng chuẩn 1 (First Normal Form, 1NF).
- Dạng chuẩn 2 (Second Normal Form, 2NF).
- Dạng chuẩn 3 (Third Normal Form, 3NF).
- Dạng chuẩn Boyce-Codd (Boyce-Codd Normal Form, BCNF, 3.5NF).

Việc đạt được một dạng chuẩn cao có nghĩa là phải thỏa mãn điều kiện của các dạng chuẩn thấp hơn nó.

Tức là, không thể đạt được dạng chuẩn 3 nếu như không đạt được dạng chuẩn 1 hoặc dạng chuẩn 2.

Một csdl quan hệ được gọi là **đã chuẩn hóa** nếu như nó đạt được dạng chuẩn 3.

Các dạng chuẩn

Chúng ta sẽ dùng dữ liệu sau đây để tìm hiểu về các dạng chuẩn cũng như quá trình chuẩn hóa csdl.

orders.xls											
	A	B	C	D	E	F	G	H	I	J	K
1	Invoice No.	Date	Cust. No.	Cust. Name	Cust. Address	Cust. City	Cust. State	Item ID	Item Descri	Item Qty.	Item Price
2	125	9/13/2002	56	Foo, Inc.	23 Main St., Thorpleburg	Thorpleburg	TX	563	56" Blue Fre	4	\$ 3.50
3								851	Spline End	32	\$ 0.25
4								652	3" Red Free	5	\$ 12.00
5	126	9/14/2002	2	Freens R Us	1600 Pennsylvania Avenue	Washington	DC	563	56" Blue Fre	500	\$ 3.50
6								652	3" Red Free	750	\$ 12.00

Hình 4: CSDL chưa chuẩn hóa

Các dạng chuẩn

Dạng chuẩn 1

Dạng chuẩn 1 đạt được khi giá trị của mỗi ô là nguyên tố, tức là không thể chia nhỏ giá trị đó thành nhiều phần có ý nghĩa tương đương.

Để đưa cơ sở dữ liệu về dạng chuẩn 1, chúng ta cần tách các giá trị không nguyên tố thành từng dòng riêng biệt.

orders.xls										
	A	B	C	D	E	F	G	H	I	J
1	Invoice No.	Date	Cust. No.	Cust. Name	Cust. Address	Cust. City	Cust. State	Item ID	Item Description	Item Qty.
2	125	9/13/2002	56	Foo, Inc.	23 Main St., Thorpleburg	Thorpleburg	TX	563	56" Blue Fre	4
3	125	9/13/2002	56	Foo, Inc.	23 Main St., Thorpleburg	Thorpleburg	TX	851	Spline End i	32
4	125	9/13/2002	56	Foo, Inc.	23 Main St., Thorpleburg	Thorpleburg	TX	652	3" Red Free	5
5	126	9/14/2002	2	Freens R Us	1600 Pennsylvania Avenue	Washington	DC	563	56" Blue Fre	500
6	126	9/14/2002	2	Freens R Us	1600 Pennsylvania Avenue	Washington	DC	652	3" Red Free	750

Hình 5: CSDL đạt dạng chuẩn 1

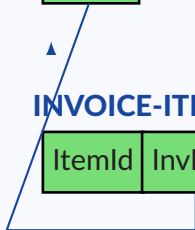
Hoặc là cũng có thể tách các giá trị không nguyên tố thành bảng riêng, kèm với một khóa ngoại tham chiếu đến bảng cũ.

INVOICE

InvNo	Date	CusNo	CusName	CusAddr	CusCity
-------	------	-------	---------	---------	---------

INVOICE-ITEM

ItemId	InvNo	ItemDesc	ItemQuan	ItemPrice
--------	-------	----------	----------	-----------



Các dạng chuẩn

Phụ thuộc hàm

Để hiểu về dạng chuẩn 2, chúng ta cần đến khái niệm phụ thuộc hàm.

Định nghĩa. Cho quan hệ $R(X, Y, \dots)$, ta nói Y phụ thuộc vào X nếu như khi biết được giá trị của X thì ta xác định được giá trị duy nhất của Y .

Ký hiệu: $X \rightarrow Y$, đọc là X xác định Y hoặc Y phụ thuộc vào X .

Với dữ liệu ở bảng bên, ta có thể thấy các phụ thuộc hàm sau:

- $A \rightarrow B, A \rightarrow C, \dots$

A	B	C	D	E
a1	b1	c1	d1	e1
a2	b1	c2	d2	e1
a3	b2	c1	d1	e1
a4	b2	c2	d2	e1
a5	b3	c3	d1	e1

Với dữ liệu ở bảng bên, ta có thể thấy các phụ thuộc hàm sau:

- $A \rightarrow B, A \rightarrow C, \dots$
- $BC \rightarrow A, BC \rightarrow DE$

A	B	C	D	E
a1	b1	c1	d1	e1
a2	b1	c2	d2	e1
a3	b2	c1	d1	e1
a4	b2	c2	d2	e1
a5	b3	c3	d1	e1

Với dữ liệu ở bảng bên, ta có thể thấy các phụ thuộc hàm sau:

- $A \rightarrow B, A \rightarrow C, \dots$
- $BC \rightarrow A, BC \rightarrow DE$
- ...

A	B	C	D	E
a1	b1	c1	d1	e1
a2	b1	c2	d2	e1
a3	b2	c1	d1	e1
a4	b2	c2	d2	e1
a5	b3	c3	d1	e1

Cho quan hệ $R(X, Y, Z)$, phụ thuộc hàm có một số tính chất:

- **Tính phản xạ.** $XY \rightarrow X$.
- **Tính tăng trưởng.** Nếu $X \rightarrow Y$ thì $XZ \rightarrow YZ$.
- **Tính bắc cầu.** Nếu $X \rightarrow Y$ và $Y \rightarrow Z$ thì $X \rightarrow Z$.

Câu hỏi. Chứng minh nếu $X \rightarrow Y$ và $X \rightarrow Z$ thì $X \rightarrow YZ$.

Các dạng chuẩn

Dạng chuẩn 2

Dạng chuẩn 2 đạt được khi:

- Đạt được dạng chuẩn 1.
- Không tồn tại phụ thuộc hàm một phần trong csdl.