

THIẾT KẾ CƠ SỞ DỮ LIỆU QUAN HỆ (Relational Database Designing)

Phần I – MÔ HÌNH CƠ SỞ DỮ LIỆU QUAN HỆ (Relational Database Modal)

Cơ Sở Dữ Liệu Là Gì ?

- Dữ liệu = các thông tin cần lưu trữ vào máy tính để có thể *truy xuất* (*access*) và *truy vấn* (*query*)
- CSDL = không gian chứa dữ liệu có tổ chức :
 - Có tương quan (*relational*)
 - Có cấu trúc : các mẫu tin (*record*), cột thuộc tính (*column*)

Sự cần thiết của CSDL

- Cung cấp khả năng truy vấn dữ liệu
 - Tránh sự dư thừa, trùng lặp dữ liệu
 - Liên kết chặt chẽ giữa dữ liệu và chương trình ứng dụng khai thác.
- Có nhiều mô hình CSDL khác nhau

Hệ Quản Trị CSDL (Database Management System)

- Là 1 phần mềm quản lý CSDL, cung cấp các dịch vụ xử lý (truy cập – truy vấn) CSDL cho
 - Nhà phát triển ứng dụng (lập trình viên)
 - Tập các hàm, thư viện mã lệnh truy xuất
 - Người dùng cuối (End-user)
 - Giao diện sử dụng để truy xuất

Các Mô Hình CSDL

- Mô hình phân cấp
 - Mô hình mạng
 - Mô hình quan hệ
 - Mô hình đối tượng
 - ...
- ➔ Các mô hình CSDL khác nhau có các Hệ Quản Trị CSDL tương ứng.

Mô Hình CSDL Quan Hệ

Do E.F.Codd đề xuất năm 1971, gồm:

- Dữ liệu được mô tả thành tập các bảng (table) (*dòng dữ liệu và cột thuộc tính*). Các bảng có thể mô tả nội dung của 1 *đối tượng/ thực thể* (Entity) hoặc 1 *mối kết hợp* (Relationship)
- Một hệ thống các **ký hiệu, phép toán và khái niệm** (thuật ngữ) để mô tả dữ liệu.

Trong phần trình bày tiếp theo, ta sẽ dựa vào ví dụ là CSDL Quản Lý Sinh Viên của 1 trường Đại Học

Các Khái Niệm : *Thuộc tính*_(Attribute)

- Thuộc vào 1 kiểu dữ liệu đơn (Data Type) nhất định và có một miền giá trị.
- Trong 1 *Lược đồ quan hệ* không được có 2 thuộc tính trùng tên.
- Ký hiệu : $A_1, A_2, \dots B, C, \dots$

Ví dụ : với đối tượng Học Sinh, ta có thể có các thuộc tính : Họ, Tên, Ngày sinh, Điểm trung bình, ...

Lược đồ quan hệ (Relation Schema) (1)

- Tập các thuộc tính của 1 đối tượng hoặc 1 mối kết hợp (1 bảng !)
- Ký hiệu :
 - Q
 - Q với tập thuộc tính : $Q(A_1, A_2, \dots, A_n)$
 - Tập thuộc tính của Q : Q^+

Lược đồ quan hệ (Relation Schema) (2)

- **Tân từ** của *Lược đồ quan hệ* :
 - Là phát biểu mô tả các ý nghĩa, điều kiện ràng buộc (giữa các thuộc tính) của lược đồ quan hệ.
 - Được phát biểu ở dạng ngôn ngữ tự nhiên hoặc ngôn ngữ hình thức, mã giả, ...

Ví dụ :

- Mỗi sinh viên chỉ có 1 mã sinh viên và không trùng với các sinh viên khác.
- Ngày kết thúc học phần phải lớn hơn ngày bắt đầu học phần.
- ...

Lược đồ quan hệ (Relation Schema) (3)

- Tập tất cả các *Lược đồ quan hệ* có trong 1 CSDL được gọi là *Lược đồ CSDL*.
- Ký hiệu : **R**

Ví dụ : trong CSDL Quản Lý Sinh Viên, **R** =

Q1 =

SINHVIEN(MaSV, Ho, Ten, NgaySinh,
DiemTB)

Q2 =

HOCPHAN(MaHP, TenHP, SoTinChi)

Quan hệ (Relation)

- Sự thể hiện của *Lược đồ quan hệ* tại một thời điểm nào đó được gọi là **Quan hệ**.
- Ký hiệu : q, r, s, t, ... tương ứng với các *Lược đồ quan hệ* Q, R, S, T, ...

Ví dụ : lược đồ Q = SINHVIEN(MaSV, Ho, Ten, NgaySinh, DiemTB) tại 1 thời điểm t nào đó có quan hệ q =

{
 SV01, Nguyễn Văn, Tài, 12/10/1980, 6.9
 SV02, Trần Thị, Chi, 19/12/1980, 7.5
 SV03, Lý Văn, Sử, 27/8/1978, 8
}

BỘ (Record)

- Là 1 dòng / phần tử của *Quan hệ*
- Là 1 tập các giá trị tương ứng với các thuộc tính của quan hệ
- Ký hiệu : p, q, t, \dots

Ví dụ : *Quan hệ* $q =$

{

}

SV01, Nguyễn Văn, Tài, 12/10/1980, 6.9

SV02, Trần Thị, Chi, 19/12/1980, 7.5

SV03, Lý Văn, Sử, 27/8/1978, 8

có 3 bộ

Kết luận

- *Bộ ∈ Quan hệ*
- *Thuộc tính ∈ Lược đồ quan hệ ∈ Lược đồ CSDL*
- *Tân từ mô tả điều kiện, ràng buộc của các Thuộc tính trong Lược đồ quan hệ*
- *Quan hệ là 1 thể hiện (instance) tại 1 thời điểm của Lược đồ quan hệ*

Siêu Khóa (Super Key)

Cho *Lược đồ quan hệ* Q có thể hiện là quan hệ q , s là 1 tập thuộc tính thuộc Q^+ . t_1, t_2 là 2 bộ tùy ý thuộc q

Ta ký hiệu $t_1.s$ là tập các giá trị của t_1 trên tập thuộc tính s

$$s' = s + A_i$$

s được gọi là 1 siêu khóa nếu và chỉ nếu :

$$\forall q, t_1, t_2 : t_1.s \neq t_2.s$$

Khóa / Khóa chỉ định

- Siêu khóa có ít thuộc tính nhất được gọi là ***Khóa*** hay ***Khóa chỉ định***.
- 1 lược đồ quan hệ có thể có nhiều *khóa*
- Trong lược đồ quan hệ, 1 *khóa* được chọn làm ***khóa chính*** (primary key).
- *Thuộc tính* có tham gia vào *khóa* được gọi là ***thuộc tính khóa***, ngược lại, được gọi là ***thuộc tính không khóa***.

Khóa ngoại (foreign key) (1)

- 1 tập *thuộc tính* được gọi là ***Khóa ngoại*** của *Lược đồ quan hệ* Q nếu và chỉ nếu nó là tập *thuộc tính* của Q^+ và là khóa của 1 *lược đồ quan hệ* khác.

Ví dụ : Cho lược đồ cơ sở dữ liệu Quản Lý Sinh Viên, các lược đồ quan hệ :

$Q_1 = \text{SINHVIEN}(\underline{\text{MaSV}}, \text{Ho}, \text{Ten}, \text{DiemTB})$

$Q_2 = \text{HOCPHAN}(\underline{\text{MaHP}}, \text{TenHP}, \text{SoTinChi})$

$Q_3 = \text{DANGKY_HOCPHAN}(\underline{\text{MaSV}}, \underline{\text{MaHP}})$

Khóa ngoại (foreign key) (2)

- {MaSV} là *khóa* của Q_1
- {MaHP} là *khóa* của Q_2
- {MaSV, MaHP} là *khóa* của Q_3
- {MaSV}, {MaHP} là các *khóa ngoại* của Q_3

Trong thiết kế CSDL, nhà thiết kế thường dựa vào các **Tân từ** để xác định khóa cho **Lược đồ quan hệ**.

Các Phép Toán : Phép toán tập hợp

Phép hợp (Union) : \cup

Cho lược đồ quan hệ Q , với các quan hệ thể hiện q_1, q_2

Ta có :

$$q_1 \cup q_2 = q_3 = \{r : r \in q_1 \vee r \in q_2\}$$

Các Phép Toán : Phép giao (Intersection)

Phép giao : \cap

Cho lược đồ quan hệ Q , với các quan hệ thể hiện q_1, q_2

Ta có :

$$q_1 \cap q_2 = q_3 = \{r : r \in q_1 \wedge r \in q_2\}$$

Các Phép Toán : Phép trừ (Minus, difference)

Phép trừ : -

Cho lược đồ quan hệ Q, với các quan hệ thể hiện q_1, q_2

Ta có :

$$q_1 - q_2 = q_3 = \{r : r \in q_1 \wedge r \notin q_2\}$$

Tích Descartes

Cho lược đồ quan hệ $Q_1(A_1, A_2, \dots, A_n)$ có
quan hệ thể hiện là q_1 ; $Q_2(B_1, B_2, \dots, B_m)$ có
quan hệ thể hiện là q_2

$q_1 \times q_2 = q_3$, ta có :

q_3 là 1 quan hệ của Q_3 với $Q_3^+ = Q_1^+ + Q_2^+$

$q_3 = \{r : r.(Q_1^+) \in q_1 \wedge r.(Q_2^+) \in q_2\}$

Tích Descartes – Ví dụ

Cho lược đồ quan hệ $Q_1(\text{MaSV}, \text{MaMH}, \text{DiemThi})$,
 $Q_2(\text{MaMH}, \text{TenMH})$ có các quan hệ q_1, q_2 :

$q_1 =$

MaSV	MaMH	DiemThi
99001	CSDL	5.0
99002	CTDL	2.0
99003	MANG	8.0

$q_2 =$

MaMH	TenMH
CSDL	Cơ sở dữ liệu
FOX	Foxpro

$\rightarrow q_3 =$

MaSV	MaMH	DiemThi	MaMH	TenMH
99001	CSDL	5.0	CSDL	Cơ sở dữ liệu
99001	CSDL	5.0	FOX	Foxpro
99002	CTDL	2.0	CSDL	Cơ sở dữ liệu
99002	CTDL	2.0	FOX	Foxpro
99003	MANG	8.0	CSDL	Cơ sở dữ liệu
99003	MANG	8.0	FOX	Foxpro

Các Phép Toán : Phép toán quan hệ

- **Phép chiếu** (Projection)

- _ Cho lược đồ quan hệ $Q(A_1, A_2, \dots, A_n)$ với quan hệ thể hiện là q , X là 1 tập con của Q^+

- _ **Phép chiếu của Q lên tập X** (ký hiệu là $Q[X]$) tạo thành lược đồ quan hệ $Q' = Q[X]$ có $Q'^+ = X$.

- _ t là 1 bộ nào đó của q , **Phép chiếu của t lên tập X** (ký hiệu là $t[X]$ hay $t.X$) là 1 tập con của t chứa các giá trị ứng với các thuộc tính trong X

- _ **Phép chiếu của q lên tập X** (ký hiệu là $q[X]$ hay $q.X$) tạo thành quan hệ q' :

$$q' = q[X] = q.X = \{t' : \exists t \in q, t.X = t'\}$$

Phép chiếu – Ví dụ

$q =$

MaSV	MaMH	DiemThi
99001	CSDL	5.0
99002	CTDL	2.0
99003	MANG	8.0

$q.\{\text{MaMH}\} = q[\{\text{MaMH}\}] =$

MaMH
CSDL
CTDL
MANG

Phép chọn (Selection)

Cho lược đồ quan hệ $Q(A_1, A_2, \dots, A_n)$ với quan hệ thể hiện là q , E là 1 biểu thức điều kiện trên các bộ của q , t là 1 bộ nào đó của q .

_ Nếu t thỏa điều kiện E , ta ký hiệu là $t(E)$ hay $t:E$

_ **Phép chọn q trên (với) điều kiện E** (ký hiệu là $q(E)$ hay $q:E$) tạo thành q' là tập con các bộ của q :

$$q' = q(E) = q:E = \{t : t \in q \wedge t(E)\}$$

Phép chọn – Ví dụ

$q =$

MaSV	MaMH	DiemThi
99001	CSDL	5.0
99002	CTDL	2.0
99003	MANG	8.0

$E : q.DiemThi \geq 5$

$q(E) = q:E =$

MaSV	MaMH	DiemThi
99001	CSDL	5.0
99003	MANG	8.0

Phép kết (Join)

Cho lược đồ quan hệ $Q_1(A_1, A_2, \dots, A_n)$ có quan hệ thể hiện là q_1 ; $Q_2(B_1, B_2, \dots, B_m)$ có quan hệ thể hiện là q_2 ; A_i và B_j là 2 thuộc tính có cùng miền giá trị, giả sử tồn tại 1 phép so sánh θ trên miền giá trị của A_i và $B_j \Rightarrow$

Gọi $q'_3 = q_1 \times q_2$ và Q_3 là lược đồ quan hệ của q'_3

$\rightarrow A_i \theta B_j$ là 1 biểu thức điều kiện trên q'_3

Phép kết (t.t)

- **Phép kết giữa q_1 và q_2 trên điều kiện $A_i \theta B_i$ (ký hiệu là $\overset{A_i \theta B_i}{\infty}$)** được xác định như sau :

$$q_3 = q_1 \overset{A_i \theta B_i}{\infty} q_2 = \left\{ t : \begin{array}{l} \exists t_1 \in q_1, t_2 \in q_2 \wedge \\ t.Q_1^+ = t_1 \wedge \\ t.Q_2^+ = t_2 \wedge \\ t_1.A_i \theta t_2.B_j \end{array} \right\}$$

Hoặc : $q_3 = q_1 \overset{A_i \theta B_i}{\infty} q_2 = q'_3(A_i \theta B_j)$

Phép kết – Ví dụ

$$q_1 =$$

MaSV	MaMH	DiemThi
99001	CSDL	5.0
99002	CSDL	2.0
99003	MANG	6.0

$$q_2 =$$

MaMH	DiemDau
CSDL	4.0
MANG	7.0

$A_i = \text{DiemThi}$; $B_j = \text{DiemDau}$; $\theta : '>='$

$$q_3 = q_1 \stackrel{A_i \theta B_j}{\infty} q_2 =$$

MaSV	MaMH	DiemThi	MaMH	DiemDau
99001	CSDL	5.0	CSDL	4.0
99003	MANG	6.0	CSDL	4.0

Phép kết – Kết luận

- Phép kết thực chất là gộp của *Tích Descartes* và phép chọn
- Trình tự thực hiện phép kết :
1. Tạo tích Descartes
 2. Thực hiện phép chọn với điều kiện $E = A_i \theta B_j$

Phép kết – Ứng dụng

- Phép kết có rất nhiều ứng dụng trong thiết kế CSDL, đặc biệt khi thực hiện các truy vấn (query)
- Ở slide 29, ta nhận thấy phép kết θ chưa thực hiện đúng ý đồ của nhà thiết kế là tìm ra các sinh viên đạt điểm đầu các môn học tương ứng (bộ thứ 2 của q_3 không đúng). Trong trường hợp này, ta dùng phép kết

$$q_3 = q_1 \underset{A_i \theta B_j, C_k \theta' D_l}{\text{kép}} q_2$$

trong đó : $C_k = q_1 \cdot \text{MaMH}$, $D_l = q_2 \cdot \text{MaMH}$ và θ' : '='

Phép kết tự nhiên (Natural Join)

- Trong trường hợp $\theta = '='$, phép kết được gọi là **Phép kết bằng**
- Trong trường hợp $A_i = B_j$, và $\theta = '='$, phép kết được gọi là **Phép kết tự nhiên**
- ➔ Ví dụ ở slide 31, phép kết giữa C_k , D_l và θ được gọi là **kết tự nhiên**.
- Phép kết tự nhiên thường được sử dụng nhất trong thực tế

Phép chia (Division)

Cho lược đồ quan hệ $Q_1(A_1, A_2, \dots, A_n)$ có quan hệ thể hiện là q_1 ; $Q_2(B_1, B_2, \dots, B_m)$ có quan hệ thể hiện là q_2 ; và $n > m$. Phép chia q_1 cho q_2 tạo thành q_3 với lược đồ quan hệ Q_3 được định nghĩa như sau :

$$Q_3^+ = \{A_1, A_2, \dots, A_{n-m}\}$$

$$q_3 = q_1 \div q_2 = \{t : \forall t_2 \in q_2, \exists t_1 \in q_1, t = t_1 \cdot Q_3^+ \wedge$$

$$t_2 = t_1 \cdot \{A_{n-m+1}, \dots, A_n\}$$

$$\}$$

Phép chia – Ví dụ

q_1				
A_1	A_2	A_3	A_4	A_5
a	b	d	c	g
a	b	d	e	f
b	c	e	e	f
e	g	c	c	g
e	g	c	e	f
a	b	e	g	c
a	b	d	t	l

q_2	
B_1	B_2
c	g
e	f

$$q_3 = q_1 \div q_2 =$$

q_3		
A_1	A_2	A_3
a	b	d
e	g	c

Phép chia – Ví dụ (t.t)

Đáp số là kết quả của phép chia quan hệ
DANGKY_HOCPHAN cho quan hệ
HOCPHAN.MAHP

DANGKY_HOCPHAN ÷ HOCPHAN.MAHP
MaSV
SV01
SV03

Các tính chất của Đại số quan hệ

Giáo trình trang 9