

Chương 3

Đại số quan hệ

(Relational algebra)

Phạm Thị Ngọc Diễm
Bộ môn HTTT - ĐHCT

Tháng 12/2015

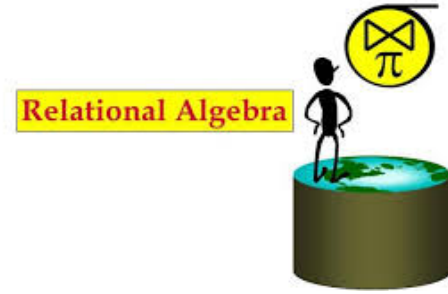
Nội dung

- Giới thiệu
- Một số khái niệm
- Các phép toán cơ bản
- Các phép toán khác

Nội dung

- Giới thiệu
- Một số khái niệm
- Các phép toán cơ bản
- Các phép toán khác

Giới thiệu



- Là ngôn ngữ hình thức cho mô hình quan hệ
- Được phát triển trước ngôn ngữ SQL
- Tập các thao tác trên mô hình quan hệ chính là ĐSQH
- Ý nghĩa và tầm quan trọng của ĐSQH:
 - Cung cấp một nền tảng cho các thao tác trên mô hình quan hệ.
 - Nó được sử dụng như là cơ sở cho việc cài đặt và tối ưu hóa các câu truy vấn
 - Một số khái niệm của nó được tích hợp vào ngôn ngữ truy vấn chuẩn SQL.

Nội dung

- Giới thiệu
- Một số khái niệm
- Các phép toán cơ bản
- Các phép toán khác

Ký hiệu

- Tập các thuộc tính $U = \{A_1, A_2, \dots, A_n\}$
- D_1, D_2, \dots, D_n tương ứng là miền giá trị của các thuộc tính A_1, A_2, \dots, A_n
- R, S là các quan hệ
- Quan hệ R gồm tập các thuộc tính U , ký hiệu $R(U)$

CSDL Ví dụ

- Sử dụng lược đồ CSDL của CSDL về trường đại học
 - SINHVIEN (MASV, hoten, namsinh, diachi, #ML)
 - LOP (ML, tenlop)
 - MONHOC (MM, tenmon, TC, LT, TH)
 - DIEM (#MASV, #MM, hk, nk, diem)
 - GIAOVIEN (MAGV, hotenGv, namsinhGv, diachiGv)
 - DAY (#MAGV, #MM, hk, nk)

Định nghĩa quan hệ

- Định nghĩa

- Quan hệ R trên tập thuộc tính $U = \{A_1, A_2, \dots, A_n\}$ là *tập con* của tích Descartes $D_1 \times D_2 \times \dots \times D_n$
- Mỗi phần tử $t = (d_1, d_2, \dots, d_n) / d_i \in D_i, i = 1, 2, \dots, n$ là một bộ của quan hệ

- Ví dụ:** xét tập các thuộc tính $U = \{ML, tenlop\}$

- $Dom(ML) = \{DI1295A1, DI1295A2, DI1395A3\} = D_1$
- $Dom(tenlop) = \{HTTT 1, HTTT 2, HTTT 3\} = D_2$

=> tích Descartes $D_1 \times D_2$ có 3x3 bộ

=> Có thể có nhiều quan hệ trong số các quan hệ

ML	tenlop
DI1295A1	HTTT 1
DI1295A2	HTTT 2
DI1395A3	HTTT 3

Đại số quan hệ

- Định nghĩa:
 - ĐSQH ký hiệu là $\alpha=(R,O)$
 - R: quan hệ
 - O: là tập các phép toán quan hệ
 - Các phép toán quan hệ:
 - Đầu vào: *một hoặc nhiều* quan hệ
 - Đầu ra hay kết quả: là *một quan hệ mới* hay quan hệ kết quả
- Biểu thức quan hệ:
 - Biểu thức gồm các quan hệ trong CSDL quan hệ và các phép toán quan hệ

Nội dung

- Giới thiệu
- Một số khái niệm
- Các phép toán cơ bản
- Các phép toán khác

Các phép toán cơ bản

- Các phép toán một ngôi
- Các phép toán tập hợp
- Các phép toán nối kết

Phép toán một ngôi

- Là phép toán chỉ thao tác trên một quan hệ. Đầu vào của chúng là một quan hệ và trả về kết quả là một quan hệ mới.
- Gồm các phép toán:
 - Chọn,
 - Chiếu,
 - Đổi tên

Phép chọn

- **Phép chọn** (Selection) được sử dụng để chọn ra một tập hợp con các bộ từ một quan hệ thỏa mãn một điều kiện chọn. Ký hiệu σ

Định nghĩa

- Cho quan hệ $R(U)$
- Cho E : biểu thức logic gồm
 - Các phép toán số học và
 - Logic : $<, \leq, >, \geq, \neq, \wedge, \vee, \neg$
- *Phép chọn σ trên quan hệ R theo điều kiện E , ký hiệu $\sigma_E(R)$ có kết quả là 1 tập hợp con của R , gồm các bộ $t \in R$ thỏa điều kiện E*

$$\sigma_E(R) = \{ t / t \in R \wedge t(E) \}$$

$t(E)$: bộ t thỏa điều kiện E

Phép chọn

- Ví dụ 1 : Cho quan hệ R(A,B,C) với các thể hiện như sau:

R	A	B	C
	1	a	2
	5	x	4
	2	b	2

$\sigma_{C=2}(R) ?$

$\sigma_{A=C}(R) ?$

- Ví dụ 2: Tìm các sinh viên sinh năm 1990

$\sigma_{\text{namsinh}=1990}(\text{SINHVIEN})$

Phép chiếu

- **Phép Chiếu** (Projection) được sử dụng để chọn một vài cột hay thuộc tính từ một bảng, ký hiệu π

Định nghĩa

- Cho quan hệ $R(U)$
- X là tập con của U : $X \subset U, X \neq \emptyset$
- *Phép chiếu π trên quan hệ R theo tập thuộc tính X , ký hiệu $\pi_X(R)$ có kết quả là tập hợp các bộ $t \in R$ nhưng chỉ lấy giá trị trên X*

$$\pi_X(R) = \{ t / t \in R \wedge t[X] \}$$

$t[X]$: t nhận giá trị trên X

- $\pi_X(R)$ chỉ gồm các bộ phân biệt (Các bộ trùng nhau trong kết quả chỉ giữ lại một)

Phép chiếu

- Ví dụ 1:** Cho quan hệ R(A,B,C) với các thể hiện như sau:

R	A	B	C
	1	a	2
	5	x	4
	2	b	2

$\pi_C(R) ?$

$\pi_{A,C}(R) ?$

- Ví dụ 2 :** - Liệt kê mã sinh viên và tên sinh viên của tất cả các sinh viên.
- Liệt kê mã sinh viên và tên sinh viên của tất cả các sinh viên sinh năm 1980.

$\pi_{MSSV,hoten}(SINHVIEN)$

$\pi_{MSSV,hoten}(\sigma_{namsinh=1990}(SINHVIEN))$

Phép đặt lại tên

- **Phép đặt lại tên ρ** cho phép đặt lại tên cho thuộc tính hoặc quan hệ.
 - $\rho_{S(A_1, A_2, \dots, A_n)}(R)$: đặt lại tên quan hệ và các thuộc tính
 - $\rho_S(R)$: đặt lại tên quan hệ, giữ nguyên các thuộc tính
 - Ví dụ:
 - $\rho_S(\pi_{MSSV, hoten}(SINHVIEN))$
 - $\rho_{S(ma, ten)}(\pi_{MSSV, hoten}(SINHVIEN))$

Các phép toán trên tập hợp

- Hai quan hệ tương thích
 - Hai quan hệ tương thích nếu
 - Có cùng bậc n
 - Thuộc tính thứ $j = 1, 2, \dots, n$ của hai quan hệ có cùng miền giá trị
- Các phép toán trên tập hợp gồm:
 - Hợp
 - Giao
 - Trừ

Phép hợp

- Ký hiệu \cup

Định nghĩa

Hợp của hai quan hệ tương thích R và S , ký hiệu $R \cup S$, gồm các bộ thuộc ít nhất một trong 2 quan hệ đã cho.

$$R \cup S = \{ t / t \in R \vee t \in S \}$$

- Ví dụ 1: Cho hai quan hệ R và S .

R	A	B	C	S	A	B	C
	1	a	2		1	a	2
	5	x	4		6	y	4
	2	b	2		2	b	2

- Ví dụ 2** : Liệt kê danh sách tên sinh viên và tên giáo viên

Phép giao

- Ký hiệu \cap

Định nghĩa

Giao của hai quan hệ tương thích R và S , ký hiệu $R \cap S$, gồm các bộ thuộc về cả 2 quan hệ đã cho.

$$R \cap S = \{ t / t \in R \wedge t \in S \}$$

- Ví dụ 1: Tìm mã số sinh viên vừa học môn TH409 vừa học môn TH408.

R	A	B	C	S	A	B	C
	1	a	2		1	a	2
	5	x	4		6	y	4
	2	b	2		2	b	2

- Ví dụ 2 :** Tìm mã số sinh viên vừa học môn TH409 vừa học

Phép trừ

- Ký hiệu \setminus

Định nghĩa

Hiệu của hai quan hệ tương thích R và S , ký hiệu $R \setminus S$, gồm các bộ thuộc R và không thuộc S .

$$R \setminus S = \{ t \mid t \in R \wedge t \notin S \}$$

- Ví dụ 1

R	A	B	C
	1	a	2
	5	x	4
	2	b	2

hệ R và S . Từ

S	A	B	C
	1	a	2
	6	y	4
	2	b	2

Các phép toán kết nối

- Là các phép toán cho phép kết hợp các bộ từ nhiều quan hệ.
- Các phép toán:
 - Phép kết nối Theta (θ)
 - Kết nối tự nhiên
 - Kết nối mở rộng

Phép kết nối Theta

- Ký hiệu $\triangleright \triangleleft$

Định nghĩa

- Cho quan hệ $R(U)$ và $S(V)$
- θ là một trong các phép toán so sánh : $<, \leq, >, \geq, =, \neq$
- Phép kết nối theo điều kiện θ giữa quan hệ R theo thuộc tính $A \in U$ và quan hệ S theo thuộc tính $B \in V$, ký hiệu $R \triangleright_{\theta} \triangleleft S$, là một quan hệ gồm các bộ có dạng (u, v) , trong đó $u \in R, v \in S$, giá trị của chúng trên A và B thỏa θ

$$R \triangleright_{\theta} \triangleleft S = \{ (u, v) / u \in R \wedge v \in S \wedge (u[A] \theta v[B]) \}$$

Chú ý: Phép kết nối chỉ thực hiện được khi θ thực hiện được giữa A và B

Phép kết nối Theta

- Ví dụ 1 : Cho hai quan hệ R và S

R	A	B	C
	1	a	2
	5	x	4
	2	b	2

S	E	F
	x	22
	y	14
	x	12

=> $R \bowtie_{B=E} S$
=?

$R \bowtie S$	A	B	C	E	F
	5	x	4	x	22
	5	x	4	x	12

- Ví dụ 2: Tìm MSSV và họ tên các sinh viên lớn tuổi hơn một giáo viên nào đó.

Phép kết nối tự nhiên

- Ký hiệu *

Định nghĩa

- *Phép kết nối tự nhiên trên 2 quan hệ $R(U)$ và $S(V)$, ký hiệu $R*S$, là phép kết nối θ dựa trên phép so sánh = của 2 thuộc tính cùng tên và cùng miền giá trị của 2 quan hệ R và S , một trong 2 thuộc tính cùng tên này sẽ bị loại bỏ qua phép chiếu.*

Phép kết nối tự nhiên

- Ví dụ 1:** Cho 2 quan hệ R và S

R	A	B	C
	1	a	2
	5	x	4
	2	b	2

S	A	E	F
	1	x	m
	6	w	n
	2	z	p

=> $R * S$

$R * S$	A	B	C	E	F
	1	a	2	x	m
	2	b	2	z	p

- Ví dụ 2:** Cho biết MSSV và họ tên các sinh viên học môn TH104

Phép kết nối mở rộng (outer join)

- Cho phép thực hiện nối kết tự nhiên trên các giá trị null của thuộc tính dùng nối kết :

Định nghĩa

- Cho $R(U)$, $S(V)$
- Đặt $N=R*S$
- Đặt P bao gồm các bộ (u,v) sao cho $u \in R$, $v \in S$, $u \notin N$, các giá trị của các thuộc tính trong v trên V đều là null
- Đặt Q bao gồm các bộ (u,v) sao cho $u \in R$, $v \in S$, $v \notin N$, các giá trị của các thuộc tính trong u trên U đều là null
- Kết nối mở rộng trái: $R \bowtie S = P \cup N$
- Kết nối mở rộng phải: $R \rightharpoonup S = N \cup Q$
- Kết nối mở rộng hai bên $R \bowtie S = P \cup N \cup Q$

Phép kết nối mở rộng

- Ví dụ:

- 1) Hãy cho biết mỗi sinh viên đã học môn nào, kể cả các sinh viên chưa học môn nào
- 2) hãy cho biết mỗi môn do các học sinh nào học, kể cả các môn không có học sinh nào học
- 3) Hãy cho biết mỗi sinh viên đã học môn nào, kể cả các sinh viên chưa học môn nào và các môn không có học sinh nào học

Phép kết nối mở rộng

- Ví dụ

$\pi_{\text{MSSV, hoten}}(\text{SINHVIEN})$

MSSV	Hoten
1940636	Phạm Thị Ngọc Diễm
1940647	Trần Thanh Điền
1940763	Nguyễn Văn Đạm
1940836	Lê Thị Thùy Linh
1940852	Đặng Thùy Lan

$\pi_{\text{MSSV, MM}}(\text{DIEM})$

MSSV	MM
1940636	CT104
1940647	CT165
1940763	CT304
1940763	CT106
1940630	CT114

Phép kết nối mở rộng

Kết nối tự nhiên:

$$\pi_{\text{MSSV, hoten}}(\text{SINHVIEN}) * \pi_{\text{MSSV, MM}}(\text{DIEM})$$

MSSV	Hoten	MM
1940636	Phạm Thị Ngọc Diễm	CT104
1940647	Trần Thanh Điền	CT165
1940763	Nguyễn Văn Đậm	CT304
1940763	Nguyễn Văn Đậm	CT106

Phép kết nối mở rộng

Mở rộng bên trái

$$\pi_{\text{MSSV, hoten}}(\text{SINHVIEN}) \bowtie \pi_{\text{MSSV, MM}}(\text{DIEM})$$

MSSV	Hoten	MM
1940636	Phạm Thị Ngọc Diễm	CT104
1940647	Trần Thanh Điền	CT165
1940763	Nguyễn Văn Đạm	CT304
1940763	Nguyễn Văn Đạm	CT106
1940836	Lê Thị Thùy Linh	null
1940852	Đặng Thùy Lan	null

Phép kết nối mở rộng

Mở rộng bên phải:

$$\pi_{\text{MSSV, hoten}}(\text{SINHVIEN}) \triangleright < \pi_{\text{MSSV, MM}}(\text{DIEM})$$

MSSV	Hoten	MM
1940636	Phạm Thị Ngọc Diễm	CT104
1940647	Trần Thanh Điền	CT165
1940763	Nguyễn Văn Đậm	CT304
1940763	Nguyễn Văn Đậm	CT106
1940630	null	CT114

Phép kết nối mở rộng

Mở rộng hai bên:

$$\pi_{\text{MSSV, hoten}}(\text{SINHVIEN}) \Join < \pi_{\text{MSSV, MM}}(\text{DIEM})$$

MSSV	Hoten	MM
1940636	Phạm Thị Ngọc Diễm	CT104
1940647	Trần Thanh Điền	CT165
1940763	Nguyễn Văn Đạm	CT304
1940763	Nguyễn Văn Đạm	CT106
1940836	Lê Thị Thùy Linh	null
1940852	Đặng Thùy Lan	null
1940630	null	CT114

Nội dung

- Giới thiệu
- Một số khái niệm
- Các phép toán cơ bản
- Các phép toán khác

Các phép toán khác

- Phép chia
- Tích Descartes
- Các hàm kết tập

Phép chia (Division)

- Ký hiệu /

Định nghĩa

- $R(U)$ là quan hệ bậc n
- $S(V)$ là quan hệ bậc m , $V \neq \emptyset$, $U \cap V \neq \emptyset$
- *Phép chia quan hệ R cho S , ký hiệu R/S , cho kết quả là một quan hệ gồm các bộ t có $n-m$ thuộc tính nhận giá trị trên $U \setminus V$ thuộc tính sao cho mỗi bộ $v \in S$ thì bộ $(t, v) \in R$*

$$R / S = \{ t / t[U \setminus V] \wedge (\forall v \in S, (t, v) \in R) \}$$

Phép chia (Division)

- Ví dụ 1 : Cho hai quan hệ R và S

R	A	B
	a	1
	b	2
	a	2
	c	1
	c	2

S	B
	1
	2

R/S	A
	a
	c

- Ví dụ 2: Tìm MSSV học tất cả các môn học

Phép tích Descartes

- Ký hiệu x
 - $R(U)$ là quan hệ bậc n
 - $S(V)$ là quan hệ bậc m , $U \cap V = \emptyset$, nếu $U \cap V \neq \emptyset$ thì đặt tên lại cho các thuộc tính trùng tên
 - *Phép tích Descartes của quan hệ R và S , ký hiệu $R \times S$, là một quan hệ có $|R| \cdot |S|$ bộ có dạng (u, v) , mỗi bộ có $n+m$ thuộc tính, $u \in R$ và $v \in S$*

$$R \times S = \{ (u, v) / u \in R \wedge v \in S \}$$

Phép tích Descartes

- Ví dụ: Cho 2 quan hệ R và S:

R	A	B
	a	1
	b	2
	a	2
	c	1
	c	2

S	C
	1
	2



R x S	A	B	C
	a	1	1
	a	1	2
	b	2	1
	b	2	2
	a	2	1
	a	2	2
	c	1	1
	c	1	1
	c	2	2
	c	2	2

Các hàm kết tập

(Aggregate Functions)

- Đầu vào : một tập các giá trị của một thuộc tính
- Đầu ra : một giá trị duy nhất
- Cho quan hệ $R(U)$ và thuộc tính $A \in U$
- Các hàm :
 - $AVG_A(R)$: tính giá trị trung bình của thuộc tính A
 - $MIN_A(R)$: tìm giá trị nhỏ nhất của thuộc tính A
 - $MAX_A(R)$: tìm giá trị lớn nhất của thuộc tính A
 - $SUM_A(R)$: tính tổng các giá trị của thuộc tính A
 - $COUNT_A(R)$: đếm số giá trị của thuộc tính A

Các hàm kết tập

(Aggregate Functions)

- Ví dụ:
 - 1) Hãy cho biết điểm trung bình, điểm lớn nhất, điểm nhỏ nhất của môn CSDL
 - 2) Hãy cho biết số môn mà sinh viên Nguyễn Văn An học
 - 3) Tổng số tín chỉ mà sinh viên Nguyễn Văn An học

Các hàm kết tập trên nhóm

- Sử dụng các hàm kết tập trên ***nhóm các tập giá trị***, ***không phải trên tập tất cả các giá trị***
- Cho E là biểu thức quan hệ
- Hàm kết tập trên nhóm có dạng:

$$G_1, G_2, \dots, G_m \text{ G } F_1(A_1), F_2(A_2), \dots, F_p(A_p)(E)$$

- Trong đó:
 - G_i là một thuộc tính mà việc chia nhóm các bộ dựa trên đó
 - F_i là một hàm kết tập
 - A_i là một thuộc tính
- Kết quả sẽ gồm $m+p$ thuộc tính (các thuộc tính phân nhóm và các giá trị của các hàm kết tập)

Các hàm kết tập trên nhóm

- Ví dụ:

Hãy cho biết điểm trung bình, điểm lớn nhất, điểm nhỏ nhất của từng sinh viên