

Đại số quan hệ

Lê Thành Văn

14-12-2023

Khoa Hệ thống thông tin quản lý

Giới thiệu

Giới thiệu

Đại số quan hệ

Đại số quan hệ là ngôn ngữ hình thức cho mô hình quan hệ được phát triển trước SQL. Đại số quan hệ còn có thể được hiểu là tập các thao tác trên mô hình quan hệ, được sử dụng như là cơ sở cho việc cài đặt và tối ưu các câu lệnh truy vấn.

Một số khái niệm của đại số quan hệ được tích hợp vào các câu lệnh truy vấn của SQL, do đó việc tìm hiểu về đại số quan hệ là bộ phóng để xây dựng và thực thi các câu lệnh SQL một cách có hiệu quả.

Giới thiệu

Các khái niệm chung

Như tên gọi, các đối tượng chính trong đại số quan hệ gồm:

- Các quan hệ.
- Những thao tác trên quan hệ.

Những thao tác thường được xét đến bao gồm

- Thao tác trên một quan hệ: phép chọn, phép chiếu.
- Thao tác trên nhiều quan hệ: phép hợp, phép giao, phép trừ, phép kết.

Ta gọi những thao tác này là các toán tử.

Sự kết hợp giữa các quan hệ cùng với các thao tác theo đúng cấu trúc được gọi là biểu thức quan hệ.

Ví dụ. $\pi_{A=1}(R \times S)$ (chọn ra những dòng có cột A bằng 1 từ tích Đề-các của R và S).

Kết quả của một biểu thức quan hệ là một quan hệ.

Các toán tử cơ bản

Các toán tử cơ bản

Các ký hiệu cơ bản

Để tiện cho việc theo dõi, chúng ta quy ước ký hiệu như sau:

- Tên quan hệ được viết in hoa toàn bộ: R , S , $CUSTOMER$, ...
- Tên cột được viết dính liền và viết in hoa chữ cái đầu mỗi từ: Age, CustomerId, ...
- Cột X của quan hệ R sẽ được ký hiệu $R.X$ hoặc $R[X]$.
- Một dòng trong quan hệ $\{ \langle t_1, t_2, \dots, t_n \rangle \}$

Các toán tử cơ bản

Các phép toán tập hợp

Đại số quan hệ được xây dựng trên lý thuyết tập hợp, nên ta có một số toán tử sau:

- Phép hợp $R \cup S$.
- Phép giao $R \cap S$.
- Phép trừ $R \setminus S$ hoặc $R - S$.
- Phép tích Đề-các $R \times S$.

Trong đó, phép hợp, phép giao và phép trừ yêu cầu hai quan hệ R và S phải khả hợp.

Định nghĩa. Hai quan hệ $R(A_0, A_1, \dots, A_n)$ và $S(B_0, B_1, \dots, B_m)$ được gọi là khả hợp khi:

- Có cùng số cột, $n = m$.
- Miền xác định của A_i phải giống miền xác định của B_i ,
 $i = 0, 1, \dots, n$.

Ví dụ. Hai quan hệ sau đây là khả hợp:

Tên	NgàySinh
Tùng	12/08/1955
Hằng	19/07/1968

Tên	SinhNhật
Trình	05/07/1985
Khang	29/02/1980
Minh	30/12/1988

Kết quả của $R \cup S$, $R \cap S$ và $R \setminus S$ là một quan hệ với tên cột là tên cột của R .

Định nghĩa. Kết quả của phép hợp $R \cup S$ là tập hợp những dòng có trong R **hoặc** có trong S .

$$R \cup S = \{u \mid u \in R \vee u \in S\}$$

R

A_0	A_1
α	1
α	2
β	1

 S

B_0	B_1
α	2
β	3

 $R \cup S$

A_0	A_1
α	1
α	2
β	1
β	3

Định nghĩa. Kết quả của phép giao $R \cap S$ là tập hợp những dòng có trong R và có trong S .

$$R \cap S = \{u \mid u \in R \wedge u \in S\}$$

R

A_0	A_1
α	1
α	2
β	1

 S

B_0	B_1
α	2
β	3

 $R \cap S$

A_0	A_1
α	2

Định nghĩa. Kết quả của phép trừ $R \setminus S$ là tập hợp những dòng **có** trong R **nhưng không có** trong S .

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

R

$\mathbf{A_0}$	$\mathbf{A_1}$
α	$\mathbf{1}$
α	$\mathbf{2}$
β	$\mathbf{1}$

 S

$\mathbf{B_0}$	$\mathbf{B_1}$
α	$\mathbf{2}$
β	$\mathbf{3}$

 $R \setminus S$

$\mathbf{A_0}$	$\mathbf{A_1}$
α	$\mathbf{1}$
β	$\mathbf{1}$

Phép tích Đề-các

Định nghĩa. Cho quan hệ R có m cột và quan hệ S có n cột, kết quả của phép tích Đề-các $R \times S$ là tập hợp những dòng có $m + n$ cột, trong đó m cột đầu là một dòng của R và n cột sau là một dòng của S .

$$R \times S = \{\langle u, v \rangle \mid u \in R \wedge v \in S\}$$

R

\mathbf{A}_0	\mathbf{A}_1
α	1
α	2
β	1

 S

\mathbf{B}_0	\mathbf{B}_1
γ	2
δ	3

 $R \times S$

\mathbf{A}_0	\mathbf{A}_1	\mathbf{B}_0	\mathbf{B}_1
α	1	γ	2
α	2	γ	2
β	1	γ	2
α	1	δ	3
α	2	δ	3
β	1	δ	3

Tính chất

Tính giao hoán

- $R \cup S = S \cup R.$
- $R \cap S = S \cap R.$

Tính kết hợp

- $R \cup (S \cap T) = (R \cup S) \cap (R \cup T).$
- $R \cap (S \cup T) = (R \cap S) \cup (R \cap T).$

Tính phân phối

- $R \cup (S \cap T) = (R \cup S) \cap (R \cup T).$
- $R \cap (S \cup T) = (R \cap S) \cup (R \cap T).$

Phép tích Đề-các không có tính giao hoán

$$R \times S \neq S \times R$$

Phép tích Đề-các có tính phân phối đối với phép hợp, phép giao và phép trừ

$$R \times (S \cup P) = (R \times S) \cup (R \times P)$$

$$R \times (S \cap P) = (R \times S) \cap (R \times P)$$

$$R \times (S \setminus P) = (R \times S) \setminus (R \times P)$$

Các toán tử cơ bản

Các toán tử quan hệ

Định nghĩa. Cho quan hệ R và một tập hợp X gồm các cột có trong R , kết quả của phép chiếu $\pi_X(R)$ là một quan hệ chỉ bao gồm những cột có trong X . Nói cách khác, ta loại bỏ khỏi R những cột không có trong X .

R

A	B	C
α	1	a
α	2	a
β	1	b

$\pi_{A, B}(R)$

A	B
α	1
α	2
β	1

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

A	C
α	a
β	b

Tính chất phép chiếu

Phép chiếu có tính lũy đẳng, nghĩa là

$$\pi_X(\pi_X(R)) = \pi_X(R)$$

Tính chất phép chiếu (tiếp)

Ngoài ra, một chuỗi các *phép chiếu hợp lệ* có cùng kết quả với phép chiếu ngoài cùng

$$\pi_Y(\pi_X(R)) = \pi_Y(R)$$

Phép chiếu hợp lệ là phép chiếu chỉ lấy ra những cột đang có trong quan hệ.

Tính chất phép chiếu (tiếp)

Phép chiếu có tính phân phối đối với phép hợp, nghĩa là

$$\pi_X (R \cup S) = \pi_X (R) \cup \pi_X (S)$$

Tuy nhiên, phép chiếu lại không có tính phân phối với phép giao và phép trừ.

Định nghĩa. Cho quan hệ R và điều kiện so sánh θ , kết quả của phép chọn $\sigma_{\theta}(R)$ là một quan hệ gồm **các dòng** mà điều kiện θ là đúng.

Điều kiện so sánh

Điều kiện so sánh gồm ba dạng:

- $\langle \text{tên cột 1} \rangle \langle \text{phép so sánh} \rangle \langle \text{tên cột 2} \rangle$.
- $\langle \text{tên cột} \rangle \langle \text{phép so sánh} \rangle \langle \text{giá trị vô hướng} \rangle$.
- $\langle \text{tên cột} \rangle$ có (không) phải trống?

Trong đó:

- $\langle \text{phép so sánh} \rangle$ là các phép $<, >, \leq, \geq, =, \neq$.
- $\langle \text{giá trị vô hướng} \rangle$ là một giá trị thuộc về miền xác định của $\langle \text{tên cột} \rangle$.

Ngoài ra, còn có thể kết hợp nhiều điều kiện so sánh với nhau bằng \wedge (phép hội), \vee (phép tuyển) và \neg (phép phủ định).

- $\varphi \wedge \psi$: cả hai điều kiện được thỏa.
- $\varphi \vee \psi$: một trong hai điều kiện được thỏa.
- $\neg\varphi$: phủ định điều kiện φ (ví dụ: $\neg(A < 5) \Leftrightarrow A \geq 5$).

R

A	B	C
α	1	a
α	2	a
β	1	b

$\sigma_{A=\alpha}(R)$

A	B	C
α	1	a
α	2	a

$\sigma_{A=\alpha \wedge B < 2}(R)$

A	B	C
α	1	a

Tính chất phép chọn

Cho quan hệ R , hai điều kiện so sánh φ và ψ , ta có các tính chất sau:

- $\sigma_{\varphi}(\sigma_{\psi}(R)) = \sigma_{\varphi \wedge \psi}(R) = \sigma_{\varphi}(R) \cap \sigma_{\psi}(R)$
- $\sigma_{\varphi \vee \psi}(R) = \sigma_{\varphi}(R) \cup \sigma_{\psi}(R)$
- $\sigma_{\neg \varphi}(R) = R \setminus \sigma_{\varphi}(R)$

Tính chất phép chọn (tiếp)

Phép chọn có tính phân phối đối với phép hợp, phép giao và phép trừ.

- $\sigma_{\theta}(R \cup S) = \sigma_{\theta}(R) \cup \sigma_{\theta}(S)$
- $\sigma_{\theta}(R \cap S) = \sigma_{\theta}(R) \cap \sigma_{\theta}(S)$
- $\sigma_{\theta}(R \setminus S) = \sigma_{\theta}(R) \setminus \sigma_{\theta}(S)$

Lưu ý. Về cơ bản, giữa phép chọn và phép chiếu **không có tính giao hoán**, nghĩa là

$$\pi_X(\sigma_\theta(R)) \neq \sigma_\theta(\pi_X(R))$$

Dấu bằng chỉ xảy ra nếu như các cột trong điều kiện so sánh θ là một tập con của các cột chiếu X .

Phép gán

Để tiện cho việc theo dõi, người ta thường tách một biểu thức quan hệ ra thành nhiều bước ngắn, vì thế, cần đặt một tên tạm cho kết quả ở mỗi bước.

Phép gán được ký hiệu bằng \leftarrow .

Chẳng hạn, biểu thức $\pi_X (\sigma_\theta (R))$ có thể được ghi thành:

1. $T \leftarrow \sigma_\theta (R).$
2. $\pi_X (T).$

Phép đổi tên

Định nghĩa. Cho quan hệ R và cách đổi tên N , kết quả của phép đổi tên $\rho_N(R)$ vẫn là những thông tin đó, nhưng với tên quan hệ và tên cột được thay đổi dựa theo cách đổi tên N .

Cách đổi tên N có ba dạng sau:

- A/B : đổi tên cột B thành A .
- (B_0, B_1, \dots, B_n) : đổi tên cột đầu tiên thành B_0 , cột thứ hai thành B_1, \dots
- $S(B_0, B_1, \dots, B_n)$: đổi tên quan hệ thành S , đổi tên cột đầu tiên thành B_0 , cột thứ hai thành B_1, \dots

Cho quan hệ $R(A, B, C)$, khi đó:

- $\rho_{D/A}(R) = R(D, B, C)$.
- $\rho_{(D,E)}(R) = R(D, E, C)$.
- $\rho_{S(D,E,F)}(R) = S(D, E, F)$.

Phép đổi tên thường được dùng trong trường hợp sử dụng đến phép kết hoặc có tính toán giữa các cột với nhau.

Các toán tử nâng cao

Các toán tử nâng cao

Phép kết

Phép kết theta

Định nghĩa. Cho hai quan hệ R và S , điều kiện so sánh θ , kết quả của phép kết theta $R \bowtie_{\theta} S$ là một quan hệ gồm những dòng thỏa mãn θ được chọn ra từ tích Đề-các $R \times S$. Nói cách khác,

$$R \bowtie_{\theta} S = \sigma_{\theta}(R \times S)$$

Do phép kết dùng để kết nối hai quan hệ với nhau, nên điều kiện so sánh thường là

⟨tên cột bảng 1⟩⟨phép so sánh⟩⟨tên cột bảng 2⟩

R

A	B	C
•	2	•
•	5	•
•	8	•

 S

D	E
3	•
6	•

 $R \bowtie_{B < D} S$

A	B	C	D	E
•	2	•	3	•
•	2	•	6	•
•	5	•	6	•

Phép kết bằng

Định nghĩa. Phép kết bằng là phép kết có điều kiện so sánh là so sánh bằng.

Phép kết tự nhiên

Định nghĩa. Phép kết tự nhiên là phép kết bằng trên những cột cùng tên giữa hai quan hệ. Ký hiệu $R \bowtie S$. Ngoài ra, trong kết quả của phép kết tự nhiên, các cột cùng tên giữa hai quan hệ chỉ xuất hiện một lần.

Cho hai quan hệ $R(A_1, A_2, \dots, A_n, C_1, C_2, \dots, C_k)$ và $S(B_1, B_2, \dots, B_m, C_1, C_2, \dots, C_k)$, khi đó

$$R \bowtie S = \pi_X \left(\sigma_{R.C_1=S.C_1, \dots, R.C_k=S.C_k} (R \times S) \right)$$

với $X = \{A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m, C_1, C_2, \dots, C_k\}$

R

A	B	C
α	2	a
β	5	a
γ	8	e

 S

A	E
α	x
λ	y

 $R \bowtie S$

A	B	C	E
α	2	a	x

Phép kết nửa

Định nghĩa. Phép kết nửa trái (phải) là phép kết tự nhiên nhưng kết quả không bao gồm những cột của bảng trái (phải). Ký hiệu $R \bowtie S$ (trái), $R \bowtie S$ (phải).

R

A	B	C
α	2	a
β	5	a
γ	8	e

 S

A	E
α	x
λ	y

 $R \bowtie S$

A	B	C
α	2	a

Phép phản kết

Định nghĩa. Kết quả của phép phản kết $R \triangleright S$ là những dòng thuộc R nhưng không tham gia vào phép kết nửa trái $R \ltimes S$. Nói cách khác

$$R \triangleright S = R - R \ltimes S$$

R

A	B	C
α	2	a
β	5	a
γ	8	e

 S

A	E
α	x
λ	y

 $R \triangleright S$

A	B	C
β	5	a
γ	8	e

Phép kết ngoài

Có thể thấy phép kết tự nhiên chỉ giữ lại những bộ thỏa điều kiện kết. Tuy nhiên, trong nhiều trường hợp, cần giữ lại những thông tin không được kết ở bảng trái hoặc bảng phải hoặc cả hai bảng. Vì thế, người ta bổ sung thêm ba phép kết:

- Phép kết (ngoài) trái $R \bowtie S$.
- Phép kết (ngoài) phải $R \ltimes S$.
- Phép kết ngoài đầy đủ $R \ltimes S$

Phép kết ngoài trái

Định nghĩa. Kết quả của phép kết ngoài trái $R \bowtie S$ là một quan hệ bao gồm:

- Tất cả những dòng của phép kết tự nhiên $R \bowtie S$.
- Tích Đề-các giữa những dòng của phép phản kết $R \triangleright S$ và một dòng trống.

$$R \bowtie S = (R \bowtie S) \cup ((R \triangleright S) \times \{\langle \omega, \omega, \dots, \omega \rangle\})$$

Có thể thấy phép kết ngoàì trái $R \bowtie S$ giữ nguyên tất cả các dòng của quan hệ bên trái.

R

A	B	C
α	2	a
β	5	a
γ	8	e

 S

A	E
α	x
λ	y

 $R \bowtie S$

A	B	C	E
α	2	a	x
β	5	a	ω
γ	8	e	ω

Phép kết ngoài phải

Chúng ta có thể định nghĩa phép kết ngoài phải tương tự như phép kết ngoài trái, và ta có công thức:

$$R \bowtie S = S \bowtie R$$

R

A	B	C
α	2	a
β	5	a
γ	8	e

 S

A	E
α	x
λ	y

 $R \bowtie S$

A	B	C	E
α	2	a	x
λ	ω	ω	y

Phép kết ngoài đầy đủ

Cuối cùng, phép kết ngoài đầy đủ là phần hợp giữa phép kết ngoài trái và phép kết ngoài phải.

$$R \bowtie S = (R \ltimes S) \cup (R \rtimes S)$$

R

A	B	C
α	2	a
β	5	a
γ	8	e

 S

A	E
α	x
λ	y

 $R \bowtie S$

A	B	C	E
α	2	a	x
β	5	a	ω
γ	8	e	ω
λ	ω	ω	y

Các toán tử nâng cao

Phép chia

Định nghĩa. Cho hai quan hệ R và S , kết quả của phép chia $R \div S$ là quan hệ T sao cho tất cả các dòng của $T \times S$ (tích Đề-các) đều là dòng của R .

Từ định nghĩa trên có thể thấy, để R có thể chia được cho S thì tất cả các cột của S phải là cột của R . Khi đó

$$R \div S = \pi_{X \setminus Y}(R) \setminus \pi_{X \setminus Y}(\pi_{X \setminus Y}(R) \times S \setminus R)$$

trong đó, X, Y lần lượt là tập hợp tất cả cột của R, S .

R

A	B	C
1	2	2
1	2	6
4	2	2
4	5	6

 S

C
2
6

 $R \div S$

A	B
1	2

Phép chia thường được dùng trả lời cho câu hỏi 'Tìm tất cả X đã tham gia tất cả Y?'

Các toán tử nâng cao

Phép chiếu mở rộng

Người ta còn mở rộng phép chiếu bằng cách cho phép thực hiện tính toán trên cột, hoặc giữa các cột với nhau. *Lưu ý:* cần đặt tên cho những cột là kết quả của việc tính toán.

R

A	B	C
α	1	1
α	2	2
β	1	3

 $\pi_{A, B+10}(R)$

A	B
α	11
α	12
β	11

 $\pi_{A, D/(B \times C)}(R)$

A	D
α	1
α	4
β	3

Các toán tử nâng cao

Phép gom nhóm

Các toán tử vừa được giới thiệu vẫn chưa đủ để trả lời các câu hỏi thống kê, vì thế, người ta bổ sung vào đại số quan hệ hai khái niệm mới là **hàm gom nhóm** và **phép gom nhóm**.

Hàm gom nhóm

Hàm gom nhóm là những hàm có đầu vào là một tập hợp các giá trị và đầu ra là một giá trị duy nhất. Có các hàm gom nhóm cơ bản:

- `sum`: tính tổng tập hợp.
- `count`: đếm số lượng khác trống.
- `avg`: tính giá trị trung bình.
- `max`: tìm giá trị lớn nhất.
- `min`: tìm giá trị nhỏ nhất.
- `count-distinct`: đếm số lượng giá trị khác nhau và khác trống.

Phép gom nhóm cơ bản

Cho quan hệ R , một cột gom nhóm G , và một hàm gom nhóm F , kết quả của phép gom nhóm $_G\mathcal{G}_{F(A)}(R)$ là một quan hệ có hai cột với:

- Cột thứ nhất là một giá trị riêng biệt của G .
- Cột thứ hai là kết quả của hàm F với đầu vào là những giá trị cột A tương ứng với giá trị riêng biệt của cột thứ nhất.

R

A	B
α	1
α	2
β	1

 $A\mathcal{G}_{\text{count}(B)}(R)$

A	count
α	2
β	1

Phép gom nhóm mở rộng

Có thể mở rộng phép gom nhóm cho nhiều cột G_1, G_2, \dots, G_n và nhiều hàm gom nhóm F_1, F_2, \dots, F_m , ký hiệu

$$G_1, G_2, \dots, G_n \mathcal{G}_{F_1(A_1), F_2(A_2), \dots, F_m(A_m)}(R)$$

Lưu ý:

- Các cột A_1, A_2, \dots, A_m không nhất thiết phải khác nhau.
- Cần đặt tên cho $F_1(A_1), F_2(A_2), \dots, F_m(A_m)$.

R

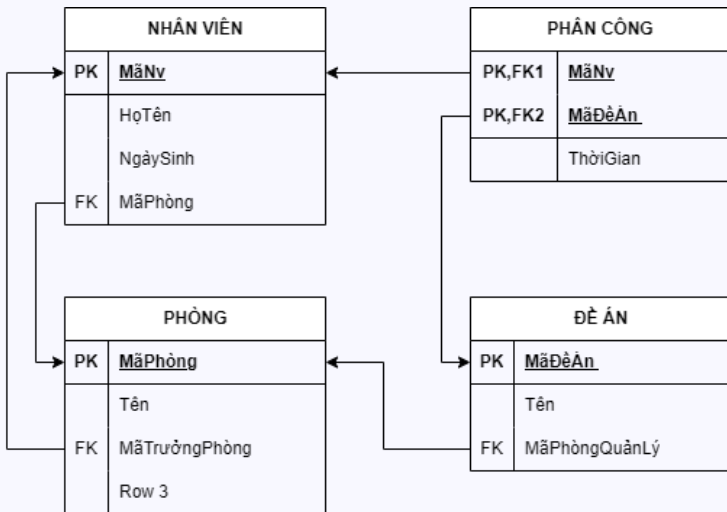
A	B	C
α	1	7
α	2	4
β	1	2
β	1	5

 $A, B \mathcal{G}_{\text{count}(C), \text{sum}(C)}(R)$

A	B	count	sum
α	1	1	7
α	2	1	4
β	1	2	7

Ví dụ

Cho cơ sở dữ liệu như hình:



Tìm **mã nhân viên** không tham gia đề án nào?

1. $T \leftarrow \pi_{\text{MãNv}}(\text{NHÂN VIÊN})$
2. $U \leftarrow \pi_{\text{MãNv}}(\text{PHÂN CÔNG})$
3. $T \setminus U$

Tìm **mã nhân viên** được phân công tham gia đề án có mã số 'THo1' hoặc đề án có mã số 'THo2'?

1. $T \leftarrow \sigma_{\text{MãĐềÁn} = \text{'THo1'} \vee \text{MãĐềÁn} = \text{'THo2'}} (\text{PHÂN CÔNG})$
2. $\pi_{\text{MãNV}}(T)$

Tìm **tên nhân viên** được phân công tham gia đề án có mã số 'TH03'?

1. $T \leftarrow \sigma_{\text{MãĐềÁn} = \text{'TH03'}} (\text{PHÂN CÔNG})$
2. $U \leftarrow \pi_{\text{MãNV}} (T)$
3. $\pi_{\text{HọTên}} (U \bowtie \text{NHÂN VIÊN})$

Tìm **mã nhân viên** tham gia tất cả các đề án?

1. $T \leftarrow \pi_{\text{MãNv, MãĐềÁn}} (\text{PHÂN CÔNG})$
2. $U \leftarrow \pi_{\text{MãĐềÁn}} (\text{ĐỀ ÁN})$
3. $T \div U$

Tìm **mã đề án** có nhiều nhân viên tham gia nhất?

1. $T \leftarrow \text{MãĐềÁn} \mathcal{G}_{\text{count}(\text{MãNv})}(\text{PHÂN CÔNG})$
2. $\rho_{\text{SốLượng/count}}(T)$
3. $U \leftarrow \sigma_{\text{SốLượng} = \max(\text{SốLượng})}(T)$
4. $\pi_{\text{MãĐềÁn}}(U)$