

Chương 5

Biến đổi các truy vấn toàn cục thành các truy vấn mảnh



Nội dung

- ❖ Biểu thức đại số quan hệ.
- ❖ Cây toán tử của truy vấn.
- ❖ Các phép biến đổi tương đương.
- ❖ Tiêu chuẩn 1 và 2.
- ❖ Đồ thị toán tử và biểu thức con chung.
- ❖ Biểu thức chuẩn tắc.
- ❖ Đại số quan hệ định tính.
- ❖ Tiêu chuẩn 3 và 4.
- ❖ Đơn giản hóa các quan hệ được phân mảnh ngang.



Nội dung

- ❖ Đơn giản hóa phép kết giữa các quan hệ được phân mảnh ngang.
- ❖ Tiêu chuẩn 5.
- ❖ Sử dụng phép suy diễn cho các phép đơn giản hóa.
- ❖ Đơn giản hóa phép kết giữa các quan hệ được phân mảnh dọc.
- ❖ Chương trình nửa kết.
- ❖ Phép gom nhóm.
- ❖ Tiêu chuẩn 6.
- ❖ Tính chất của các hàm kết hợp.



Nội dung

- ❖ Đơn giản hóa truy vấn có tham số.
- ❖ Sử dụng vùng nhớ tạm để thực hiện truy vấn có tham số.

Biểu thức đại số quan hệ

- ❖ Biến đổi truy vấn SQL thành các biểu thức đại số quan hệ.
- ❖ Một biểu thức đại số quan hệ (*expression of relational algebra*): **chuỗi các phép toán** (*sequence of operations*).
- ❖ Hai biểu thức có cùng ngữ nghĩa có thể mô tả hai chuỗi phép toán khác nhau.

$\Pi_{\text{name, deptnum}} \sigma_{\text{deptnum} = 15} (\text{emp})$

$\sigma_{\text{deptnum} = 15} \Pi_{\text{name, deptnum}} (\text{emp})$

Cây toán tử của truy vấn

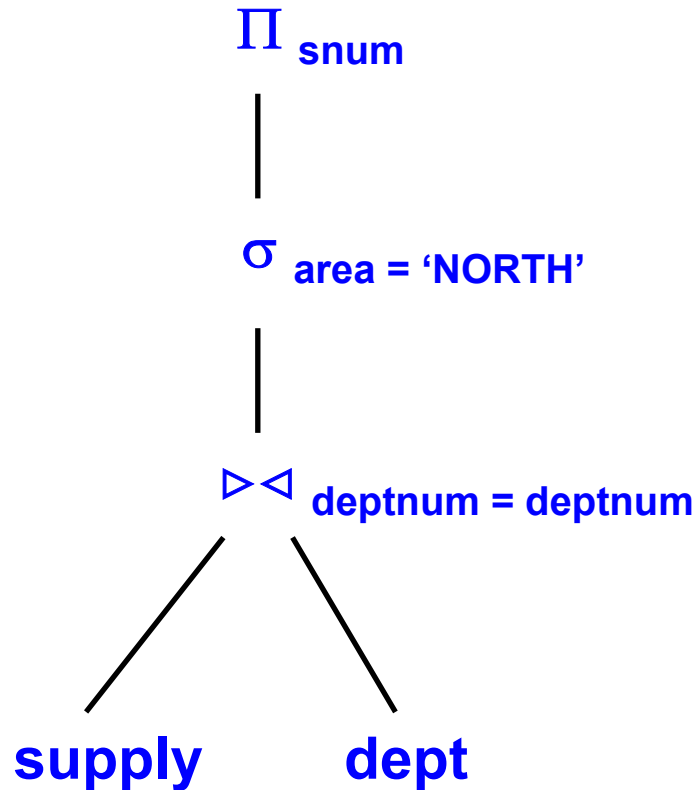
❖ Một truy vấn được biểu diễn bằng **cây toán tử** (*operator tree*).

❖ Ví dụ

▶ Truy vấn Q_1 – Hãy cho biết mã của các nhà cung cấp có đơn hàng cung cấp ở phía Bắc.

$Q_1: \Pi_{snum} \sigma_{area = 'NORTH'} (supply \bowtie_{deptnum = deptnum} dept)$

Cây toán tử của truy vấn



Hình 5.1. Cây toán tử của truy vấn Q_1



Các phép biến đổi tương đương

- ❖ *Hai quan hệ R_1 và R_2 là tương đương* nếu các bộ của chúng biểu diễn cùng ánh xạ từ các tên thuộc tính vào các giá trị, ngay cả khi thứ tự của các thuộc tính là khác nhau.
- ❖ *Hai biểu thức đại số quan hệ E_1 và E_2 là tương đương*, ký hiệu là $E_1 \leftrightarrow E_2$ hoặc $E_1 \equiv E_2$ nếu thay thế cùng các quan hệ cho các tên giống nhau trong hai biểu thức, thì chúng có các kết quả tương đương.

Các phép biến đổi tương đương

❖ Các tính chất

- ▶ **Tính giao hoán (commutativity)** của các phép toán một ngôi:

$$U_1 U_2 R \leftrightarrow U_2 U_1 R$$

- ▶ **Tính giao hoán** của các toán hạng của các phép toán hai ngôi:

$$R B S \leftrightarrow S B R$$

- ▶ **Tính kết hợp (associativity)** của các phép toán hai ngôi:

$$R B (S B T) \leftrightarrow (R B S) B T$$

Các phép biến đổi tương đương

❖ Các tính chất

- ▶ **Tính lũy đẳng** (*idempotence*) của các phép toán một ngôi:

$$U R \leftrightarrow U_1 U_2 R$$

trong đó U, U_1, U_2 thuộc cùng loại phép toán.

- ▶ **Tính phân phối** (*distributivity*) của các phép toán một ngôi đối với các phép toán hai ngôi:

$$U (R B S) \rightarrow U(R) B U(S)$$

- ▶ **Tính rút thừa số** (*factorization*) của các phép toán một ngôi:

$$U(R) B U(S) \rightarrow U(R B S)$$

❖ Một số phép biến đổi tương đương.

Tiêu chuẩn 1 và 2

- ❖ Mục đích: giảm kích thước của các toán hạng của các phép toán hai ngôi trước khi thực hiện chúng.
- ❖ **Tiêu chuẩn 1** - Sử dụng tính lũy đẳng của phép chọn và phép chiếu để tạo ra các phép chọn và các phép chiếu thích hợp đối với mỗi quan hệ toán hạng.
- ❖ **Tiêu chuẩn 2** - Đẩy các phép chọn và các phép chiếu xuống phía dưới cây nếu có thể được.



Đồ thị toán tử và biểu thức con chung

- ❖ **Biểu thức con chung** (*common subexpression*) là biểu thức xuất hiện nhiều lần trong truy vấn.
- ❖ Tiết kiệm thời gian thực hiện của truy vấn.
- ❖ Biến đổi cây toán tử thành một đồ thị toán tử.

Đồ thị toán tử và biểu thức con chung

❖ Ví dụ

- ▶ Truy vấn Q_2 – Hãy cho biết các tên của các nhân viên làm việc trong phòng ban có mã người quản lý là 373 nhưng tiền lương của họ không lớn hơn \$35.000.

$Q_2: \Pi_{\text{emp.name}} ((\text{emp} \bowtie_{\text{deptnum} = \text{deptnum}} \sigma_{\text{mgrnum} = 373} \text{dept}) - (\sigma_{\text{sal} > 35000} \text{emp} \bowtie_{\text{deptnum} = \text{deptnum}} \sigma_{\text{mgrnum} = 373} \text{dept}))$

- ▶ Biểu thức con chung

$\text{emp} \bowtie_{\text{deptnum} = \text{deptnum}} \sigma_{\text{mgrnum} = 373} \text{dept}$

- ❖ Các phép biến đổi tương đương (liên quan đến một quan hệ R) để đơn giản hóa cây toán tử.

Biểu thức chuẩn tắc

- ❖ **Biểu thức chuẩn tắc** (*canonical expression*) của một biểu thức đại số quan hệ trên lược đồ toàn cục có được bằng cách thay thế mỗi tên quan hệ toàn cục xuất hiện trong nó bởi biểu thức đại số quan hệ tái tạo các quan hệ toàn cục từ các mảnh.
- ❖ Sử dụng tính phân phối của phép chọn và phép chiếu đối với phép hợp và phép kết để phân phối việc xử lý đến các mảnh.

Đại số quan hệ định tính

- ❖ **Quan hệ định tính** (*qualified relation*) là một quan hệ được mở rộng bởi một vị từ định tính.
- ❖ Ký hiệu một quan hệ định tính là một cặp $[R: q_R]$, trong đó R là một quan hệ được gọi là **thân** (*body*) của quan hệ định tính và q_R là một vị từ được gọi là **vị từ định tính** của quan hệ định tính.

Đại số quan hệ định tính

► Quy tắc 1

$$\sigma_F [R : q_R] \Rightarrow [\sigma_F R : F \text{ AND } q_R]$$

► Quy tắc 2

$$\Pi_A [R : q_R] \Rightarrow [\Pi_A R : q_R]$$

► Quy tắc 3

$$[R : q_R] \times [S : q_S] \Rightarrow [R \times S : q_R \text{ AND } q_S]$$

► Quy tắc 4

$$[R : q_R] - [S : q_S] \Rightarrow [R - S : q_R]$$

► Quy tắc 5

$$[R : q_R] \cup [S : q_S] \Rightarrow [R \cup S : q_R \text{ OR } q_S]$$

► Quy tắc 6

$$[R : q_R] \triangleright \triangleleft_F [S : q_S] \Rightarrow [R \triangleright \triangleleft_F S : q_R \text{ AND } q_S \text{ AND } F]$$

► Quy tắc 7

$$[R : q_R] \triangleright \triangleleft_F [S : q_S] \Rightarrow [R \triangleright \triangleleft_F S : q_R \text{ AND } q_S \text{ AND } F]$$

Đại số quan hệ định tính

- ❖ **Hai quan hệ định tính là tương đương** nếu các thân của chúng là các quan hệ tương đương và các vị từ định tính của chúng biểu diễn cùng hàm chân trị (nghĩa là, nếu áp dụng cả hai vị từ định tính cho cùng một bộ thì chúng có cùng một giá trị chân trị).
- ❖ Sử dụng các vị từ định tính để **loại bỏ các mảnh** không dùng để tạo ra kết quả của truy vấn.
- ❖ Các phép biến đổi tương đương (liên quan đến quan hệ rỗng) để đơn giản hóa cây toán tử.

Tiêu chuẩn 3 và 4

- ❖ Mục đích: đơn giản các quan hệ được phân mảnh ngang và các phép kết giữa các quan hệ được phân mảnh ngang.
- ❖ **Tiêu chuẩn 3** - Đẩy các phép chọn xuống phía các nút lá của cây, và sau đó thực hiện chúng bằng cách dùng đại số quan hệ định tính. Thay thế kết quả của phép chọn bởi quan hệ rỗng nếu vị từ định tính của kết quả bị mâu thuẫn.



Tiêu chuẩn 3 và 4

- ❖ **Tiêu chuẩn 4** - Sử dụng đại số quan hệ định tính để định trị vị từ định tính của các toán hạng của các phép kết. Thay thế cây con, bao gồm phép kết và các toán hạng của nó, bởi quan hệ rỗng nếu vị từ định tính của kết quả của phép kết bị mâu thuẫn.



Đơn giản hóa các quan hệ được phân mảnh ngang

❖ Ví dụ

- ▶ Xét truy vấn Q_3 trên quan hệ *dept* được phân mảnh ngang:

$Q_3: \sigma_{\text{deptnum} = 1} \text{ dept}$

Đơn giản hóa các phép kết giữa các quan hệ được phân mảnh ngang

► Giải pháp 1:

$$R \bowtie_F S = (\cup_i R_i) \bowtie_F (\cup_j S_j)$$

► Giải pháp 2: *phép kết phân tán (distributed join)*.

$$R \bowtie_F S = \cup_{ij} (R_i \bowtie_F S_j)$$

❖ Đánh giá:

- Chọn giải pháp 1 nếu có nhiều cặp mảnh được kết với nhau.
- Chọn giải pháp 2 nếu có một số cặp mảnh được kết với nhau.

❖ Đồ thị kết (*join graph*).

Tiêu chuẩn 5

- ❖ Mục đích: biến đổi một truy vấn không có các phép kết phân tán thành một truy vấn có phép kết phân tán.
- ❖ **Tiêu chuẩn 5** - Để phân phối các phép kết xuất hiện trong một truy vấn toàn cục, các phép hợp (biểu diễn việc tập hợp các mảnh) phải được đẩy lên phía trên các phép kết muốn phân phối.



Tiêu chuẩn 5

❖ Ví dụ

- ▶ Truy vấn Q_4 - Hãy cho biết tên (*name*) của tất cả các nhà cung cấp có đơn hàng cung cấp:

$Q_4: \Pi_{name} (\text{supply} \triangleright \triangleleft \text{supplier})$



Sử dụng phép suy diễn cho các phép đơn giản hóa

- ❖ Mâu thuẫn giữa các điều kiện chọn của các truy vấn và các vị từ định tính của các mảnh.
- ❖ Bộ chứng minh định lý (*theorem prover*).
- ❖ Ví dụ
 - ▶ Xét truy vấn Q_1 – Cho biết mã của các nhà cung cấp có đơn hàng cung cấp ở phía Bắc.
 - ▶ Cây toán tử của Q_1 trong Hình 5.4.



Sử dụng phép suy diễn cho các phép đơn giản hóa

► Giả sử:

- (1) Phía Bắc chỉ bao gồm các phòng ban có mã từ 1 đến 10.
- (2) Tất cả các đơn hàng của các phòng ban có mã từ 1 đến 10 đều gửi đến các nhà cung cấp ở San Francisco.

► Từ (1), có thể viết các điều suy diễn sau đây:

$\text{area} = \text{'NORTH'} \Rightarrow \text{not } (10 < \text{deptnum} \leq 20)$

$\text{area} = \text{'NORTH'} \Rightarrow \text{not } (\text{deptnum} > 20)$

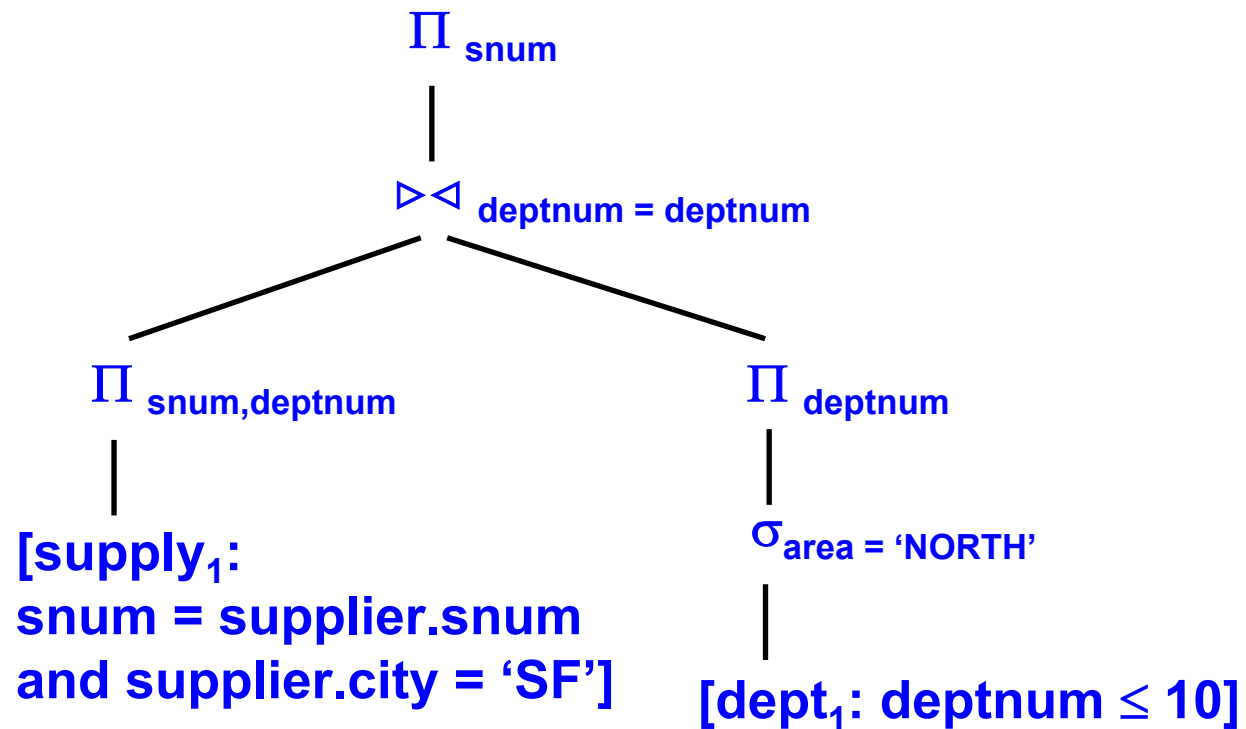
$\text{area} = \text{'NORTH'} \Rightarrow \text{deptnum} \leq 10$

► Từ (2) :

$\text{deptnum} \leq 10 \Rightarrow$

$\text{not } (\text{snum} = \text{supplier.snum and supplier.city} = \text{'LA'})$

Sử dụng phép suy diễn cho các phép đơn giản hóa



Hình 5.7. Đơn giản hóa cây toán tử bằng sự suy diễn.

Đơn giản hoá các quan hệ được phân mảnh dọc

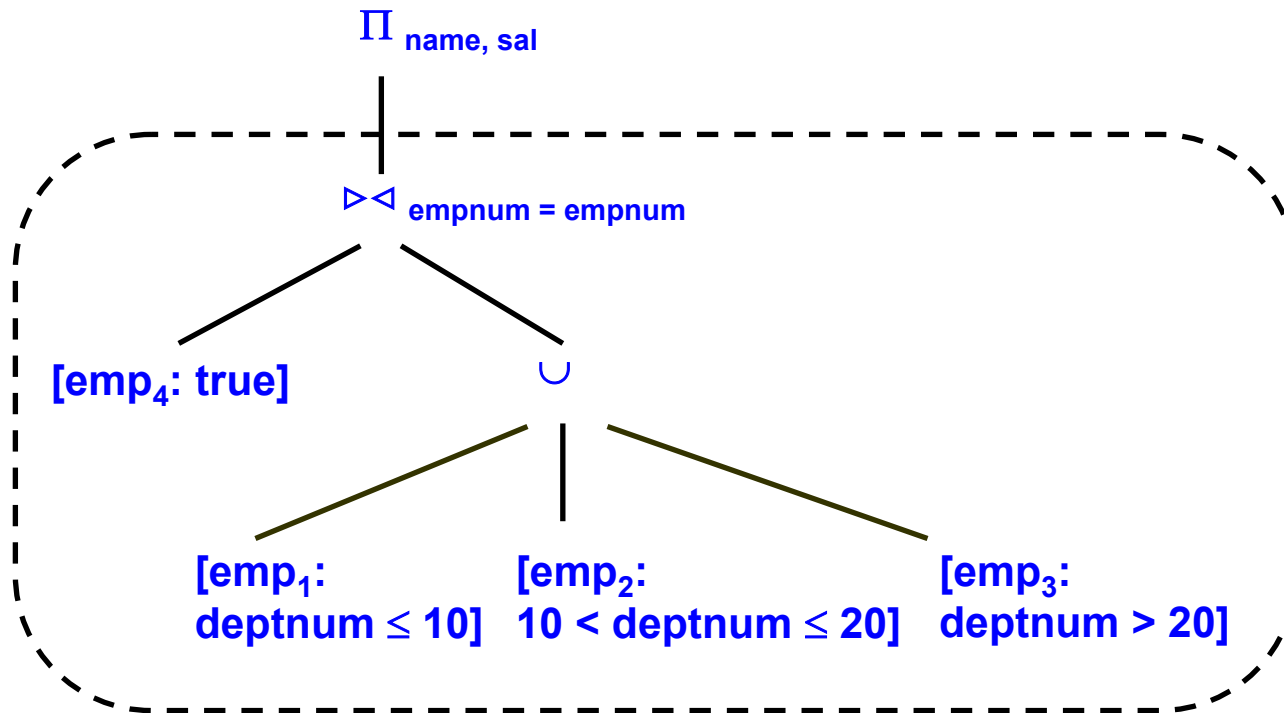
❖ Mục đích: xác định một tập con bao gồm các mảnh đủ để trả lời truy vấn, sau đó loại bỏ tất cả các mảnh khác từ biểu thức truy vấn và các phép kết được dùng trong phép đổi ngược của lược đồ phân mảnh để tái tạo các quan hệ toàn cục.

❖ Ví dụ

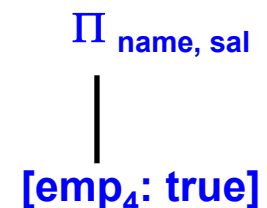
- ▶ Truy vấn Q_5 – Hãy cho biết tên và tiền lương của các nhân viên:

$Q_5: \Pi_{\text{name, sal}} \text{ emp}$

Đơn giản hoá các quan hệ được phân mảnh dọc



Hình 5.8. (a) Dạng chuẩn tắc của truy vấn Q_5



Hình 5.8. (b) Truy vấn Q_5 đã được đơn giản hóa.

Chương trình nửa kết

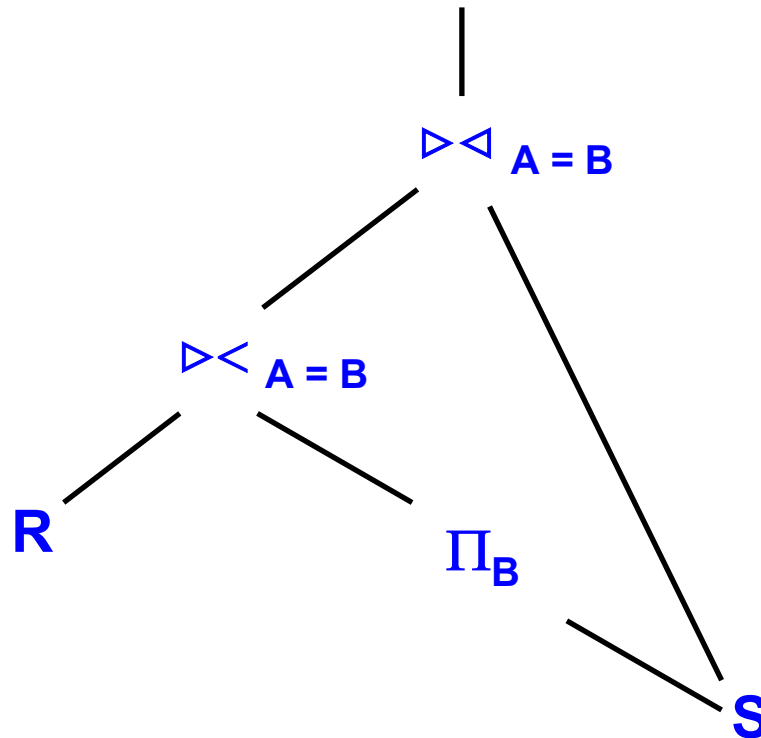
❖ Một phép kết có thể được thực hiện bởi một **chương trình nửa kết** (*semi-join program*) trong đó có các phép nửa kết.

❖ Ví dụ

► Xét phép kết bằng (*equi-join*) $R \bowtie_{A=B} S$, trong đó A và B là các thuộc tính (hoặc tập các thuộc tính) của R và S , chương trình nửa kết ứng với phép kết này là:

$$S \bowtie_{A=B} (R \bowtie_{A=B} \Pi_{S.B})$$

Chương trình nửa kết



Hình 5.9. Đồ thị toán tử của chương trình nửa kết $R \bowtie_{A=B} S$.

Phép gom nhóm

❖ Phép gom nhóm

$$\Psi_{G, AF} R$$

- ▶ G – các thuộc tính dùng để xác định việc gom nhóm của R , được gọi là tập thuộc tính gom nhóm. G tương ứng với mệnh đề *GROUP BY*.
- ▶ AF – các hàm kết hợp được định trị trên mỗi nhóm. AF tương ứng với các hàm kết hợp cần được tính toán.
- ▶ Có thể không có G hoặc AF .

Phép gom nhóm

❖ Phép gom nhóm

- ▶ $\Psi_{G,AF} R$ là một quan hệ có:
 - Lược đồ quan hệ được tạo ra bởi các thuộc tính của G và các hàm kết hợp của AF .
 - Nhiều bộ mà mỗi bộ là một nhóm trong R . Các thuộc tính của G lấy giá trị của nhóm. Các thuộc tính của AF lấy giá trị của các hàm kết hợp được định trị trên nhóm.

Phép gom nhóm

❖ Ví dụ

Q_6 :
 select AVG(*quan*)
 from *supply*
 where *pnum* = 'P1';

Q_7 :
 $\Psi_{\text{AVG}(\text{quan})} \sigma_{\text{pnum} = \text{'P1'}} \text{ supply}$
 select *snum*, *pnum*, SUM(*quan*)
 from *supply*
 group by *snum*, *pnum*;

Q_8 :
 $\Psi_{\text{snum, pnum, SUM}(\text{quan})} \text{ supply}$
 select *snum*, *pnum*, SUM(*quan*)
 from *supply*
 group by *snum*, *pnum*
 having SUM(*quan*) > 300;

$\sigma_{\text{SUM}(\text{quan}) > 300} \Psi_{\text{snum, pnum, SUM}(\text{quan})} \text{ supply}$

Tính chất của phép gom nhóm

- ❖ Tính phân phối của phép gom nhóm đối với phép hợp:

$$\Psi_{G,AF}(R_1 \cup R_2) \rightarrow (\Psi_{G,AF} R_1) \cup (\Psi_{G,AF} R_2)$$

- ▶ **Điều kiện cần và đủ:** mỗi nhóm G_i hoặc được chứa hoặc không được giao nhau với mọi toán hạng R_j .

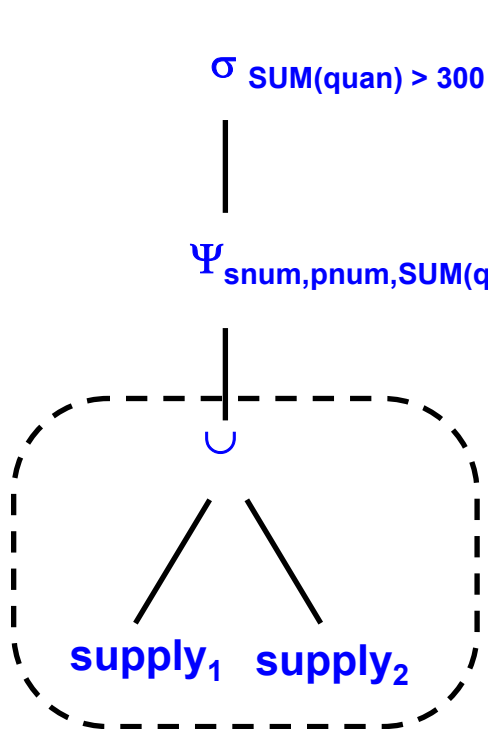
$$\forall i, j : (G_i \subseteq R_j) \text{ hoặc } (G_i \cap R_j = \emptyset)$$

- ▶ Mỗi nhóm phải được chứa hoàn toàn trong một mảnh.
- ▶ Thực hiện phép gom nhóm trên các toán hạng của phép hợp và sau đó hợp các kết quả này.

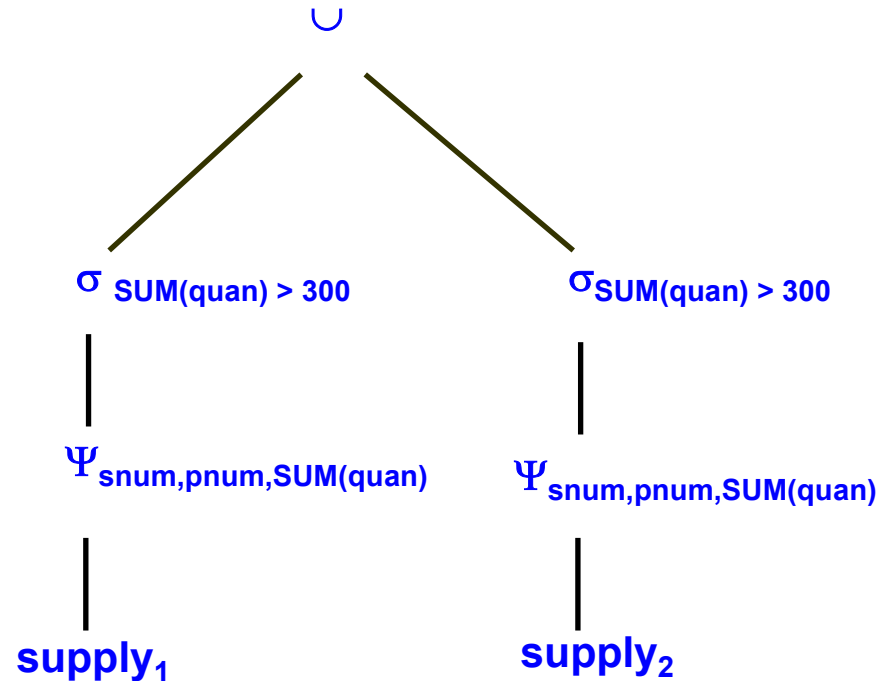
Tiêu chuẩn 6

- ❖ Mục đích: tập hợp các kết quả (nhỏ) của các phép gom nhóm thay vì tập hợp các quan hệ toàn cục (lớn).
- ❖ **Tiêu chuẩn 6** - Để phân tán việc gom nhóm và định trị hàm kết hợp xuất hiện trong một truy vấn toàn cục, các phép hợp (biểu diễn việc tập hợp các mảnh) phải được đẩy lên phía trên phép gom nhóm tương ứng.

Tiêu chuẩn 6



(a) Dạng chuẩn tắc của truy vấn Q₈



(b) Bản phân tán của truy vấn Q₈

Hình 5.10. Một truy vấn với việc gom nhóm và các hàm kết hợp.



Tính chất của hàm kết hợp

► Hàm tìm giá trị nhỏ nhất

$$\text{MIN}(S) = \text{MIN}(\text{MIN}(S_1), \text{MIN}(S_2), \dots, \text{MIN}(S_n))$$

► Hàm tìm giá trị lớn nhất

$$\text{MAX}(S) = \text{MAX}(\text{MAX}(S_1), \text{MAX}(S_2), \dots, \text{MAX}(S_n))$$

► Hàm đếm

$$\text{COUNT}(S) = \text{SUM}(\text{COUNT}(S_1), \text{COUNT}(S_2), \dots, \text{COUNT}(S_n))$$

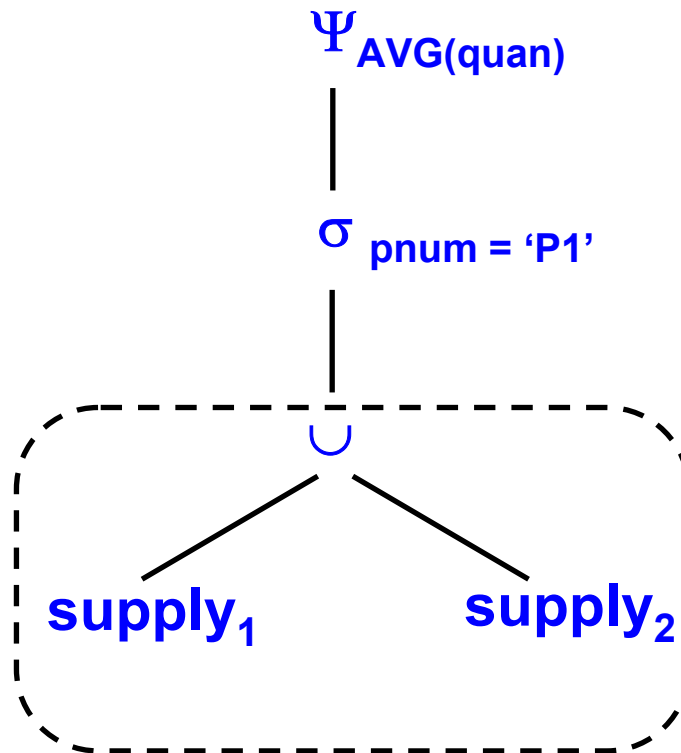
► Hàm tính giá trị tổng cộng

$$\text{SUM}(S) = \text{SUM}(\text{SUM}(S_1), \text{SUM}(S_2), \dots, \text{SUM}(S_n))$$

► Hàm tính giá trị trung bình

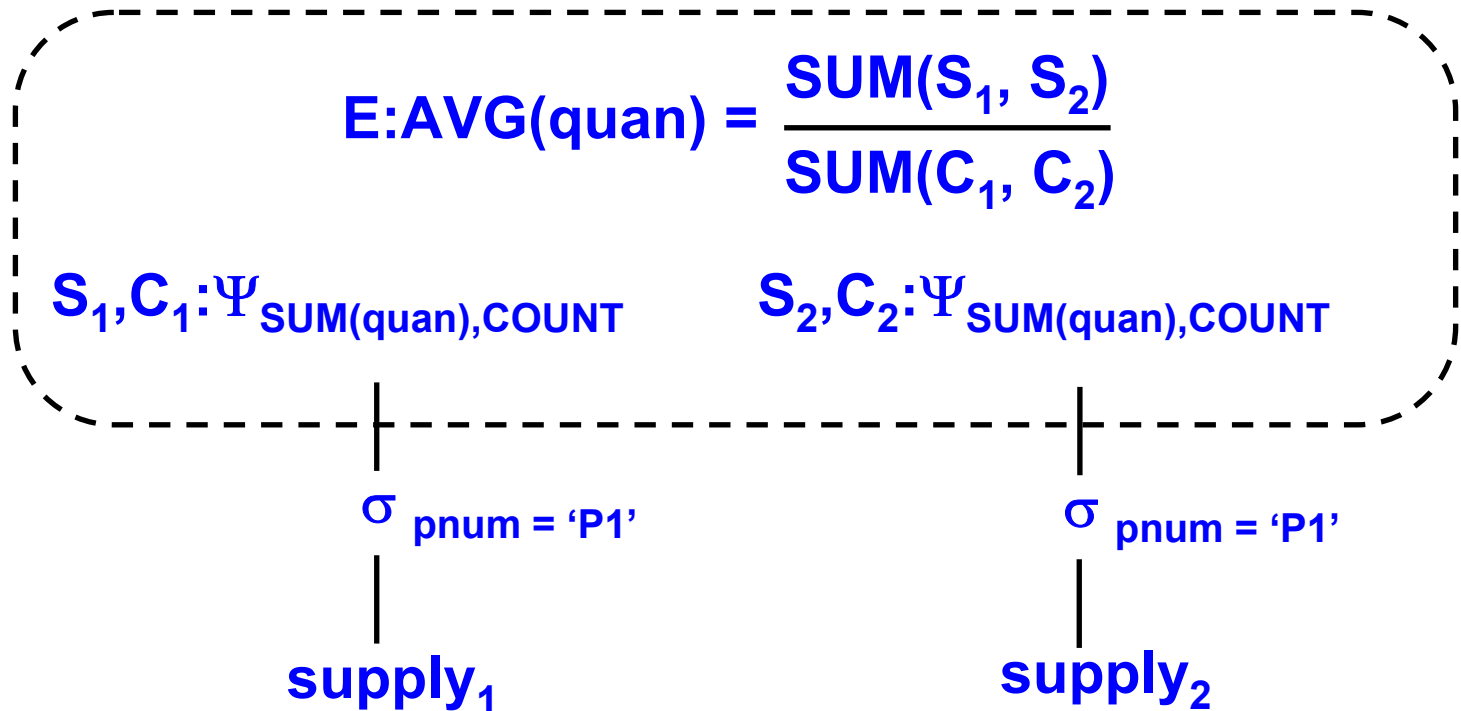
$$\text{AVG}(S) = \frac{\text{SUM}(\text{SUM}(S_1), \text{SUM}(S_2), \dots, \text{SUM}(S_n))}{\text{SUM}(\text{COUNT}(S_1), \text{COUNT}(S_2), \dots, \text{COUNT}(S_n))}$$

Tính chất của hàm kết hợp



Hình 5.11. (a) Định trị phân tán của các hàm kết hợp.

Tính chất của hàm kết hợp



Hình 5.11. (b) Định trị phân tán của các hàm kết hợp.

Truy vấn có tham số

- ❖ **Truy vấn có tham số** (*parametric query*) là truy vấn mà trong đó các công thức trong các điều kiện chọn của truy vấn bao gồm các tham số mà các giá trị của chúng chưa được biết khi biên dịch truy vấn.
- ❖ Truy vấn có tham số cho phép thực hiện truy vấn nhiều lần với nhiều giá trị khác nhau của các tham số; ở mỗi lần thực hiện sẽ trả về kết quả khác nhau.

Đơn giản hóa truy vấn có tham số

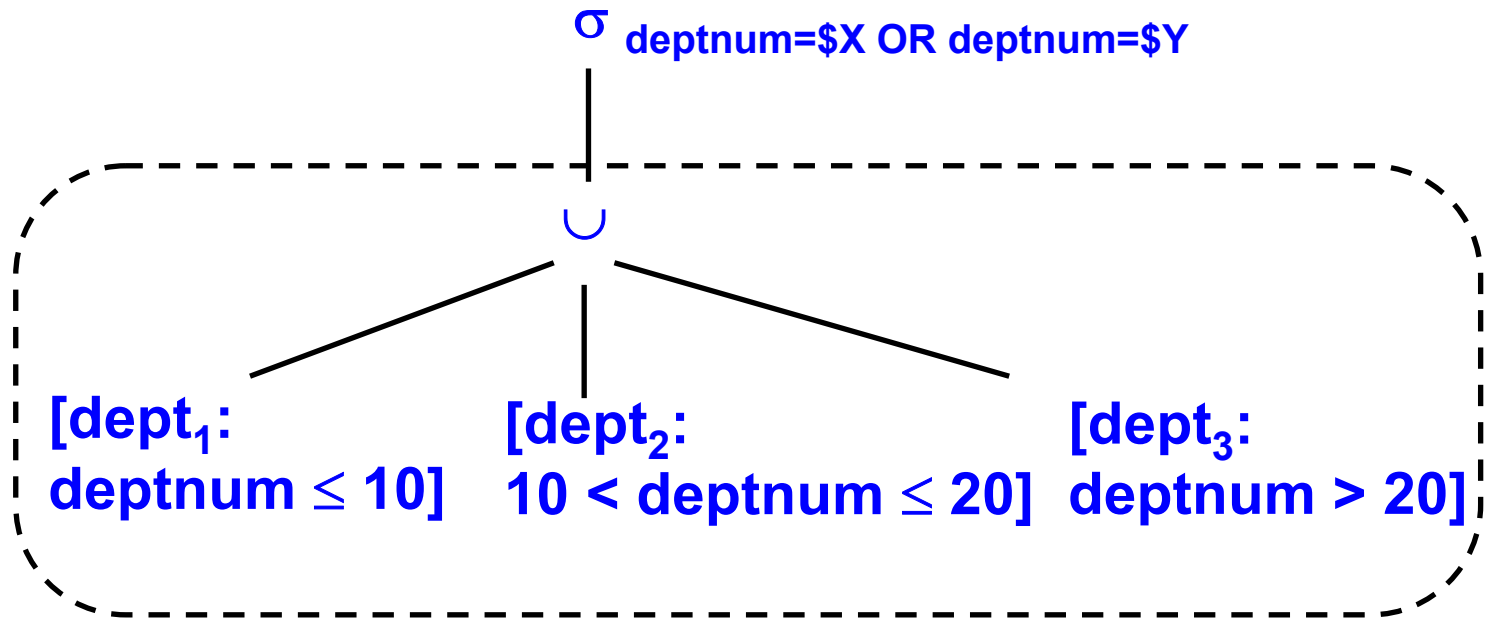
❖ Ví dụ

- ▶ Xét truy vấn Q_9 - Chọn các bộ của quan hệ toàn cục *dept* có các mã phòng ban cho trước. Phép chọn trên *deptnum* có tham số:

$$Q_9: \sigma_{\text{deptnum} = \$X \text{ OR } \text{deptnum} = \$Y} \text{ dept}$$

- ▶ Ở thời gian biên dịch: không biết các mảnh nào của quan hệ toàn cục *dept* sẽ được sử dụng.
- ▶ Ở thời gian chạy: các giá trị thực sự được gán cho các tham số $\$X$ và $\$Y$ và xác định được các mảnh nào có liên quan đến truy vấn.

Đơn giản hóa truy vấn có tham số

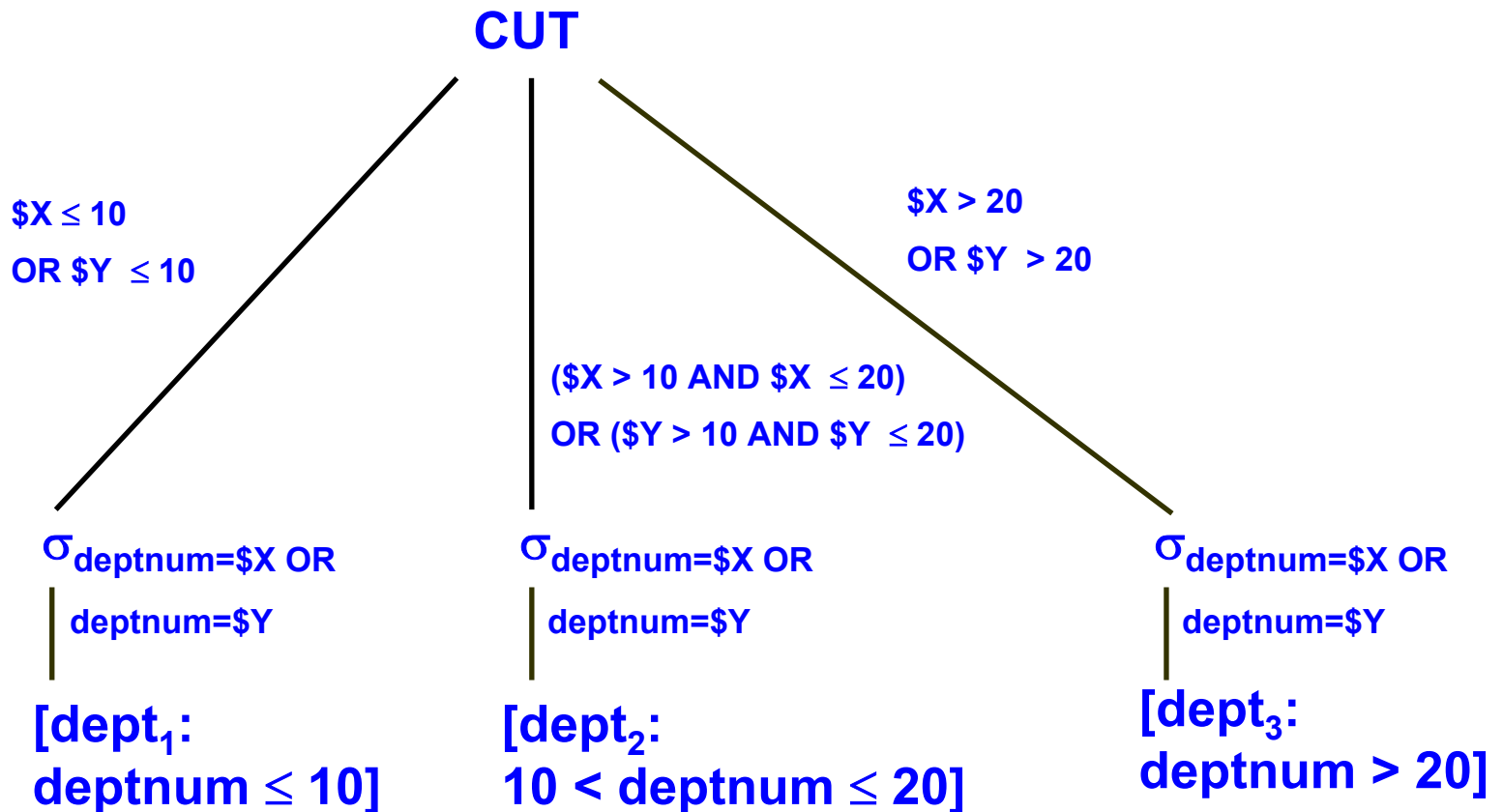


Hình 5.12. (a) Dạng chuẩn tắc của truy vấn Q_9

Đơn giản hóa truy vấn có tham số

- ❖ Đơn giản hóa truy vấn có tham số: áp dụng đại số quan hệ định tính để xác định các vị từ định tính của các biểu thức con là mâu thuẫn với nhau.
- ❖ Biểu diễn phép đơn giản hóa ở thời gian chạy:
 - ▶ Thay thế các phép hợp bởi một phép toán mới n -ngôi, được gọi là *CUT*.
 - ▶ Phép toán *CUT* thực hiện phép hợp của chỉ một số toán hạng của nó.

Đơn giản hóa truy vấn có tham số



Hình 5.12. (b) Cây truy vấn với phép CUT.

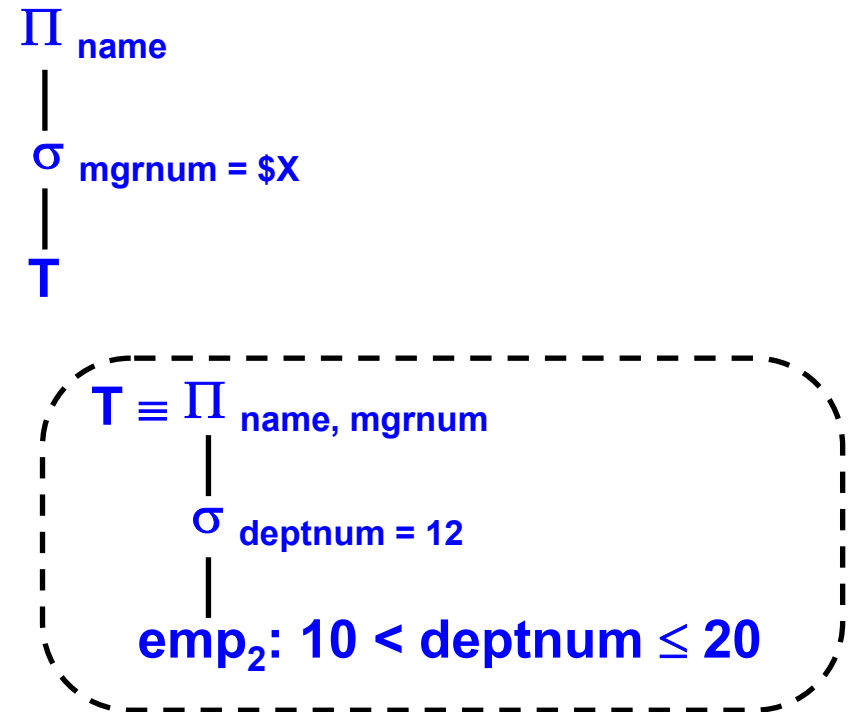
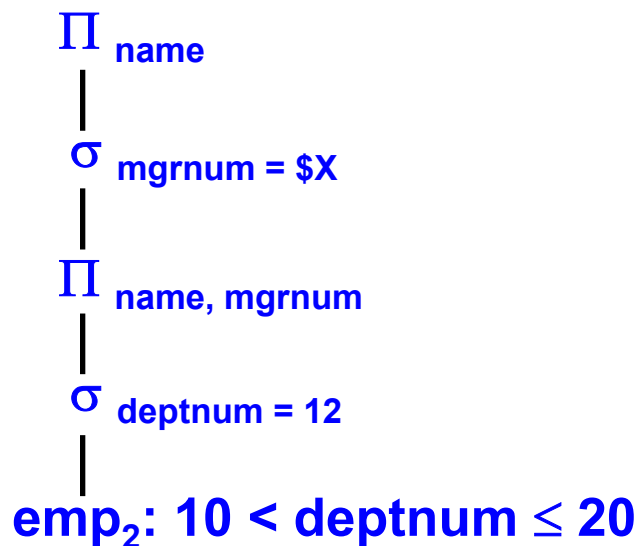


Sử dụng vùng nhớ tạm khi thực hiện nhiều lần truy vấn có tham số

- ❖ Giảm chi phí thực hiện: sử dụng các quan hệ tạm thời ở nơi gốc của truy vấn.
- ❖ Ví dụ
 - ▶ Xét truy vấn Q_{10} - Hãy cho biết tên của các nhân viên đang làm việc ở phòng ban 12 mà có mã sếp là \$X (tham số của truy vấn):

$Q_{10}: \Pi_{\text{name}} \sigma_{\text{mgrnum} = \$X \text{ AND deptnum} = 12} \text{emp}$

Sử dụng vùng nhớ tạm khi thực hiện nhiều lần truy vấn có tham số



Hình 5.13. Sử dụng các quan hệ tạm thời cho các truy vấn có tham số.