CHƯƠNG 4 Biến đổi các truy vấn toàn cục thành 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.
- 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_{NAME, DEPTNUM} \sigma_{DEPTNUM = 15} (EMP)
```

$$\sigma_{DEPTNUM = 15} \Pi_{NAME, DEPTNUM}$$
 (EMP)

Example.DDB

Lược đồ toàn cục:

EMP (EMPNUM, NAME, SAL, TAX, MGRNUM, DEPTNUM)

DEPT (**DEPTNUM**, NAME, AREA, MGRNUM)

SUPPLIER (SNUM, NAME, CITY)

SUPPLY (SNUM, PNUM, DEPTNUM, QUAN)

Example.DDB

LƯỢC ĐỔ PHÂN MẢNH:

```
\begin{split} & EMP1 = \sigma_{DEPTNUM \leq 10} \; \Pi_{EMPNUM, \, NAME, \, MGRNUM, \, DEPTNUM} \; EMP \\ & EMP2 = \sigma_{10 < DEPTNUM \leq 20} \; \Pi_{EMPNUM, \, NAME, \, MGRNUM, \, DEPTNUM} \; EMP \\ & EMP3 = \sigma_{DEPTNUM > 20} \; \Pi_{EMPNUM, \, NAME, \, MGRNUM, \, DEPTNUM} \; EMP \\ & EMP4 = \Pi_{EMPNUM, \, NAME, \, SAL, \, TAX} \; EMP \\ & DEPT1 = \sigma_{DEPTNUM \leq 10} \; DEPT \\ & DEPT2 = \sigma_{10 < DEPTNUM \leq 20} \; DEPT \\ & DEPT3 = \sigma_{DEPTNUM > 20} \; DEPT \end{split}
```

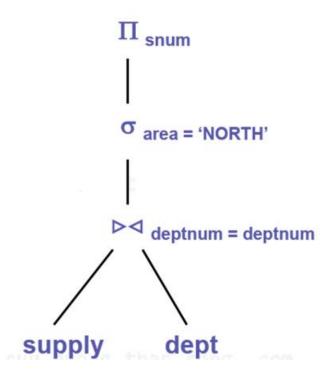
SUPPLIER1 = $\sigma_{CITY = 'SF'}$ SUPPLIER SUPPLIER2 = $\sigma_{CITY = 'IA'}$ SUPPLIER

Cây toán tử truy vấn

- Một truy vấn được biểu diễn bằng cây toán tử (operator tree)
- ❖ Ví dụ: Truy vấn Q1 Hãy cho biết mã của các nhà cung cấp có đơn hàng cung cấp ở phía Bắc.

Q1: $\Pi_{SNUM} \sigma_{AREA = 'NORTH'}$ (SUPPLY $\triangleright \triangleleft_{DEPTNUM = DEPTNUM}$ DEPT)

Cây toán tử truy vấn



Q1:
$$\Pi_{SNUM} \sigma_{AREA = 'NORTH'}$$
 (SUPPLY $\triangleright \triangleleft_{DEPTNUM = DEPTNUM}$ DEPT)

Phép kết được thực hiện trước tiên, tiếp theo là phép chọn và cuối cùng là phép chiếu

Các phép biến đổi tương đương dùng cho đại số quan hệ

- Hai quan hệ R1 và R2 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 E1 và E2 là tương đương, ký hiệu là E1 ↔ E2 hoặc E1 ≡ E2 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 dùng cho đại số quan hệ

Cho U là phép toán đại số một ngôi và B là phép toán đại số hai ngôi (trong đó U, U₁, U₂ thuộc cùng loại phép toán)

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:

$$RBS \leftrightarrow SBR$$

Các phép biến đổi tương đương dùng cho đại số quan hệ

> Tính kết hợp (associativity) của các phép toán hai ngôi:

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

> Tính lũy đẳng (idempotence) của các phép toán một ngôi:

$$U R \leftrightarrow U_1 U_2 R$$

Tiêu chuẩn 1 và tiêu chuẩn 2

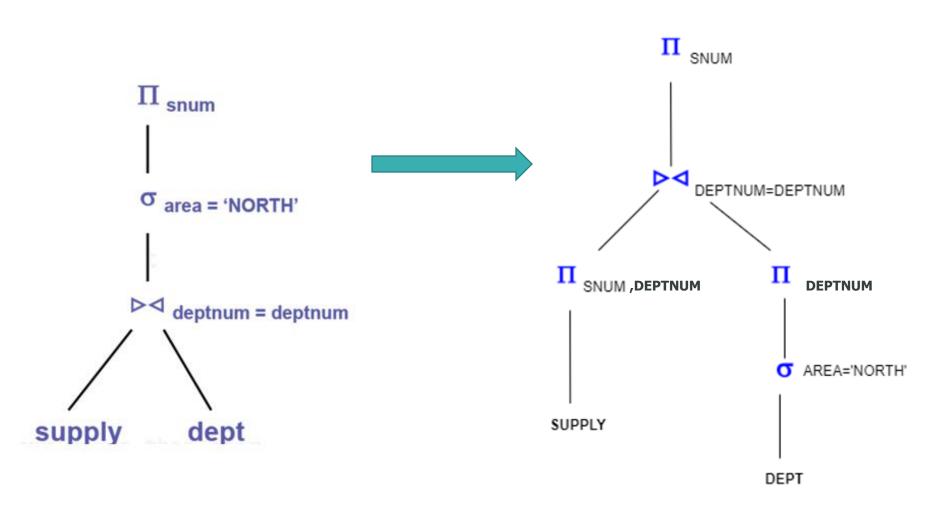
- Mục đích: giảm kích thước của các toán hạng của các phép toán hai ngôi (đặc biệt là các phép kết) trước khi thực hiện chúng.
- Tiêu chuẩn 1 Sử dụng tính lũy đẳng của phép chọn và phép chiếu để tạo ra các phép chọn và các phép chiếu thích hợp đối với mỗi quan hệ toán hạng.
- Tiêu chuẩn 2 Đẩy các phép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

Tiêu chuẩn 1 và tiêu chuẩn 2

- ❖ Ví dụ: Thay đổi truy vấn Q1
 - Phép chọn được phân phối đối với phép kết. Do đó phép chọn được áp dụng trực tiếp cho quan hệ DEPT
 - > Hai phép chiếu mới được tạo ra và được phân phối đối với phép kết

```
Q1: \Pi_{SNUM} \sigma_{AREA = 'NORTH'} (SUPPLY \triangleright \triangleleft_{DEPTNUM = DEPTNUM} DEPT)
```

Tiêu chuẩn 1 và tiêu chuẩn 2

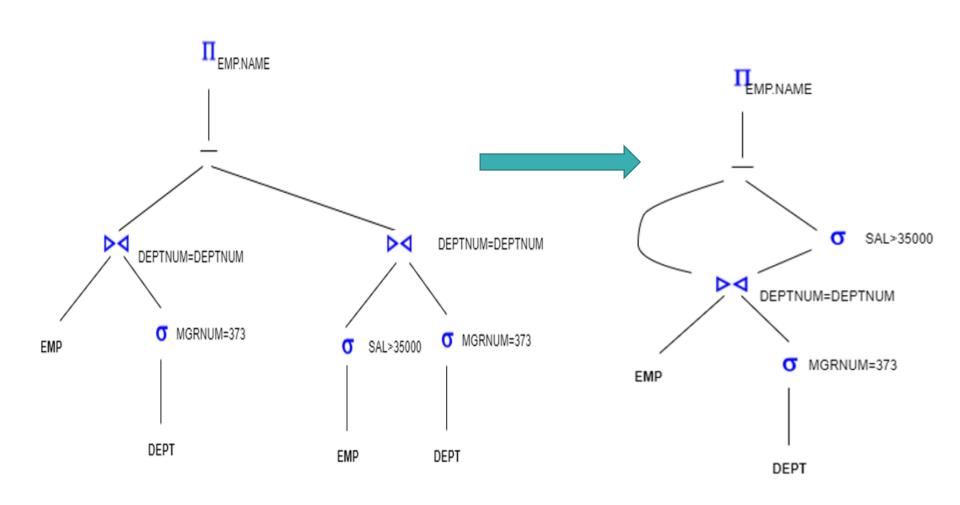


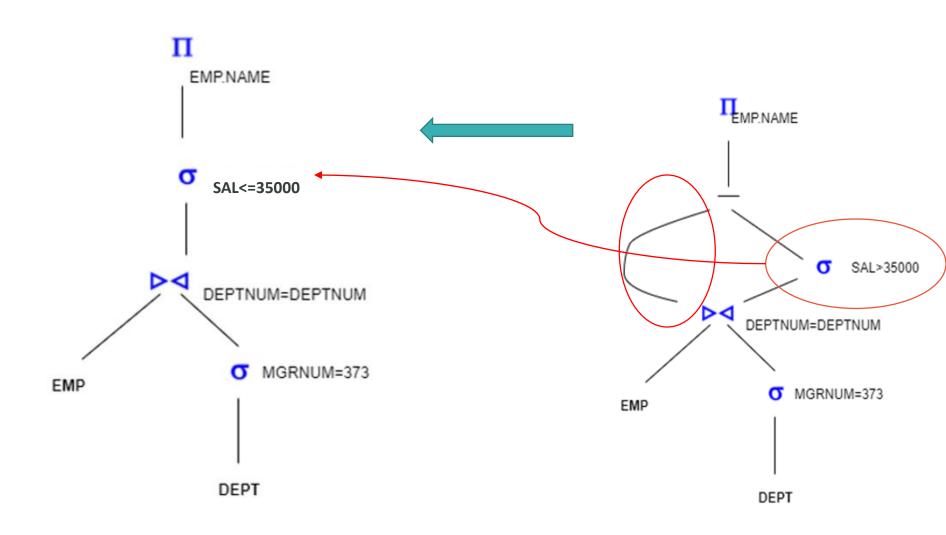
Q1: $\Pi_{SNUM} \sigma_{AREA = 'NORTH'}$ (SUPPLY $\triangleright \triangleleft_{DEPTNUM = DEPTNUM}$ DEPT)

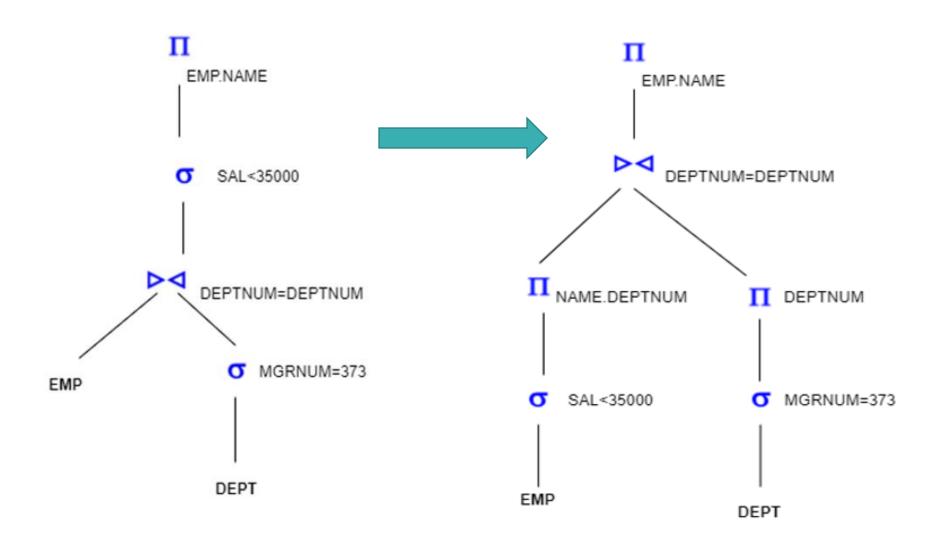
- ❖ Biểu thức con chung (common subexpression) là biểu thức xuất hiện nhiều lần trong truy vấn.
- ❖ Tiết kiệm thời gian thực hiện của truy vấn.

- Ví dụ
- Truy vấn Q2 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.

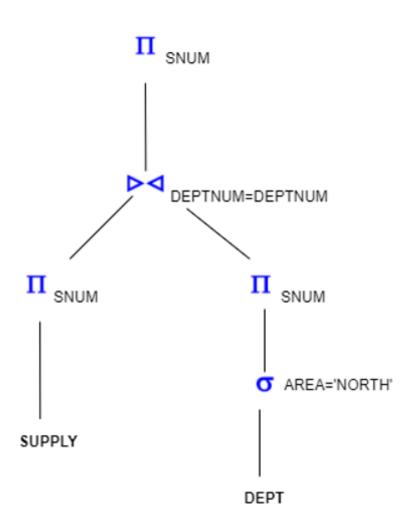
Q2:
$$\Pi_{\text{EMP.NAME}}$$
 ((EMP $\triangleright \triangleleft$ DEPTNUM = DEPTNUM $\sigma_{\text{MGRNUM} = 373}$ DEPT) – ($\sigma_{\text{SAL} > 35000}$ EMP $\triangleright \triangleleft$ DEPTNUM = DEPTNUM $\sigma_{\text{MGRNUM} = 373}$ DEPT))



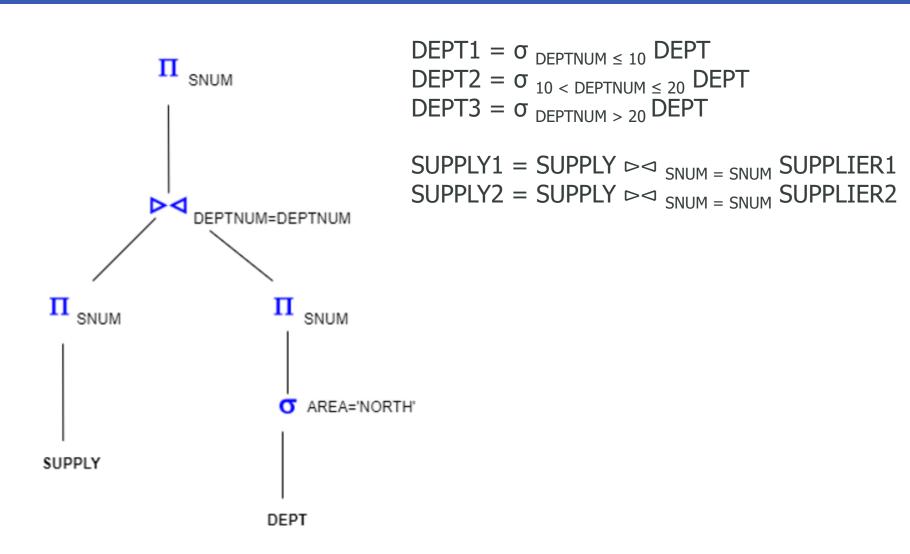




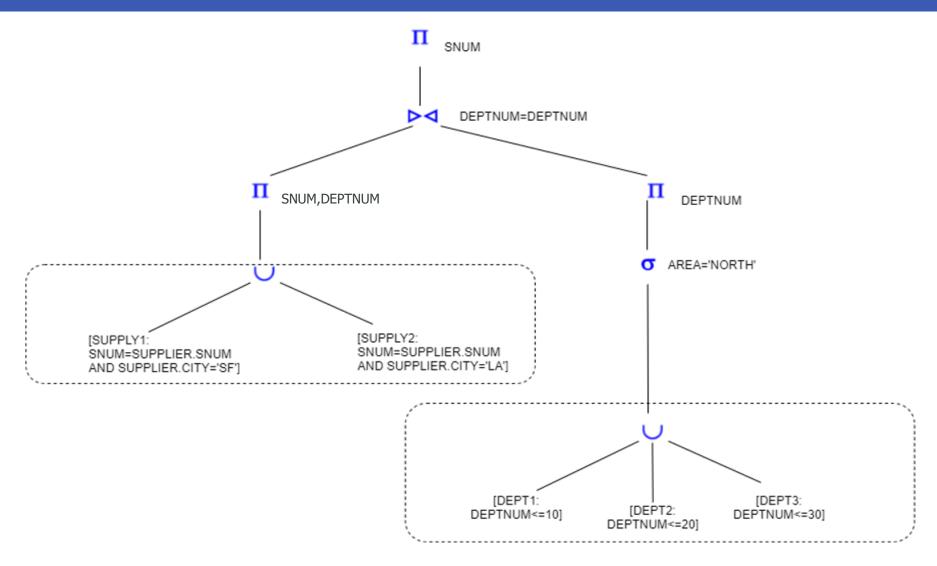
- ❖ Biểu thức chuẩn tắc (canonical expression) của một biểu thức đại số quan hệ trên lược đồ toàn cục có được bằng cách thay thế mỗi tên quan hệ toàn cục xuất hiện trong nó bởi biểu thức đại số quan hệ tái tạo các quan hệ toàn cục từ các mảnh.
- Sử dụng tính phân phối của phép chọn và phép chiếu đối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



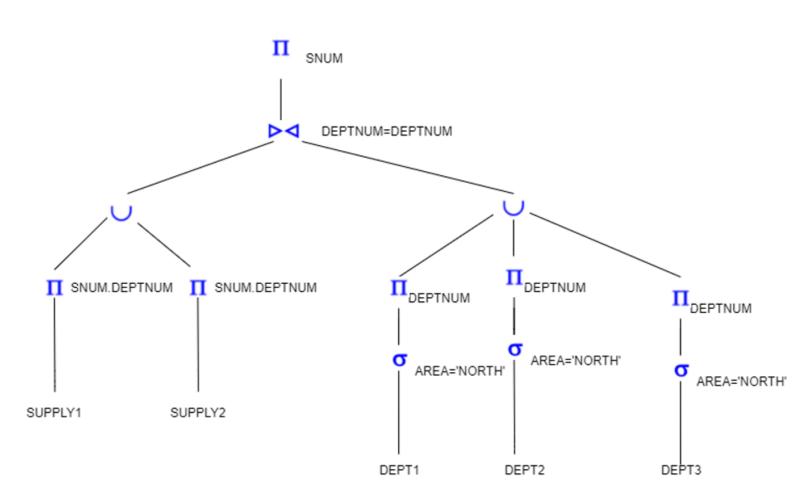
Q1: $\Pi_{SNUM} \sigma_{AREA = 'NORTH'}$ (SUPPLY $\triangleright \triangleleft_{DEPTNUM} = DEPTNUM$ DEPT)



Q1: $\Pi_{SNUM} \sigma_{AREA = 'NORTH'}$ (SUPPLY $\triangleright \triangleleft_{DEPTNUM = DEPTNUM}$ DEPT)



Dạng chuẩn tắc của truy vấn Q1



Đẩy các phép chọn và phép chiếu xuống phía dưới cây toán tử

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.
 SELECT SNUM FROM SUPPLIER1 WHERE CITY = 'LA'
- ❖ Các mảnh ngang là các vị dụ tiêu biểu của các quan hệ định tính, trong đó vị từ định tính tương ứng với vị từ phân mảnh

[SUPPLIER₁:CITY='SF'] [SUPPLIER₂:CITY='LA']

Đại số quan hệ định tính

Quy tắc 1

$$\sigma_{F} [R: q_{R}] \Rightarrow [\sigma_{F}R: F AND q_{R}]$$

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

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

Đại số quan hệ định tính

Quy tắc 5

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

Quy tắc 6

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

Quy tắc 7

$$[R:q_R] \triangleright <_F [S:q_S] \Rightarrow [R \triangleright <_F S:q_R AND q_S AND F]$$

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.

Tiêu chuẩn 3 và 4

❖ Ví dụ các biểu thức con dẫn đến quan hệ rỗng là:

$$\sigma_{\text{CITY='LA'}}[\text{SUPPLIER}_1:\text{CITY='SF'}]$$

Hoặc

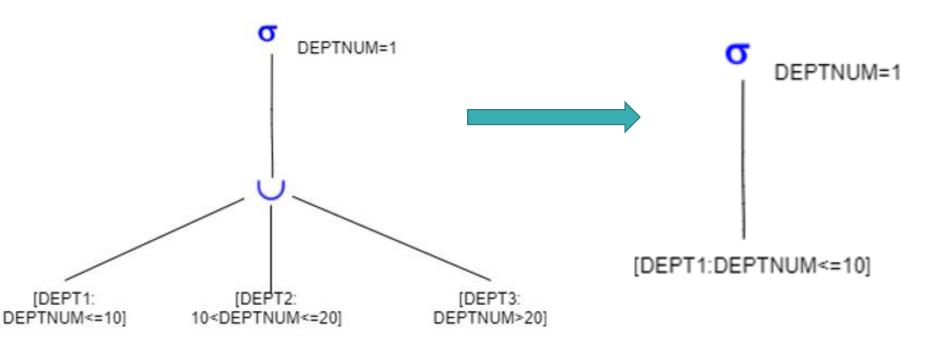
[DEPT₁:DEPTNUM ≤10] $\triangleright \triangleleft$ DEPTNUM=DEPTNUM[EMP3: DEPTNUM>20]

Đơn giản hóa các quan hệ được phân mảnh ngang

* Ví dụ

Xét truy vấn Q3 trên quan hệ dept được phân mảnh ngang:

Q3:
$$\sigma_{DEPTNUM = 1}$$
 DEPT



Dạng chuẩn tắc của truy vấn Q3

Đơn giản hóa của truy vấn Q3

❖ Ví dụ: Truy vấn Q4 - Hãy cho biết tên (name) của tất cả các nhà cung cấp có đơn hàng cung cấp:

EMP (EMPNUM, NAME, SAL, TAX, MGRNUM, DEPTNUM)

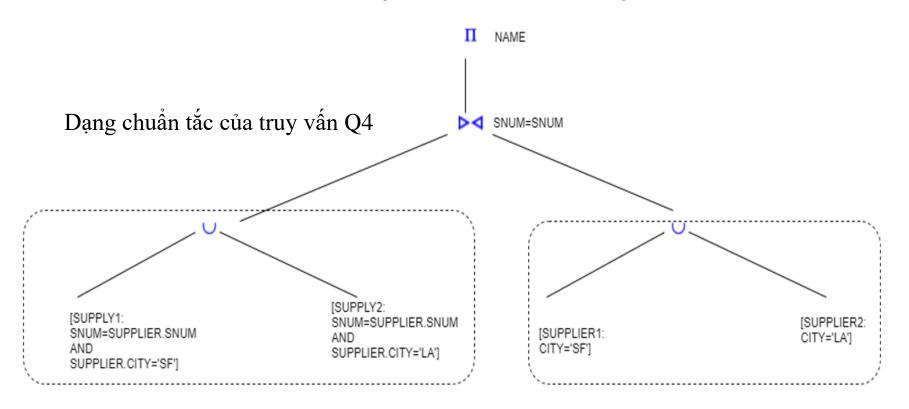
DEPT (**DEPTNUM**, NAME, AREA, MGRNUM)

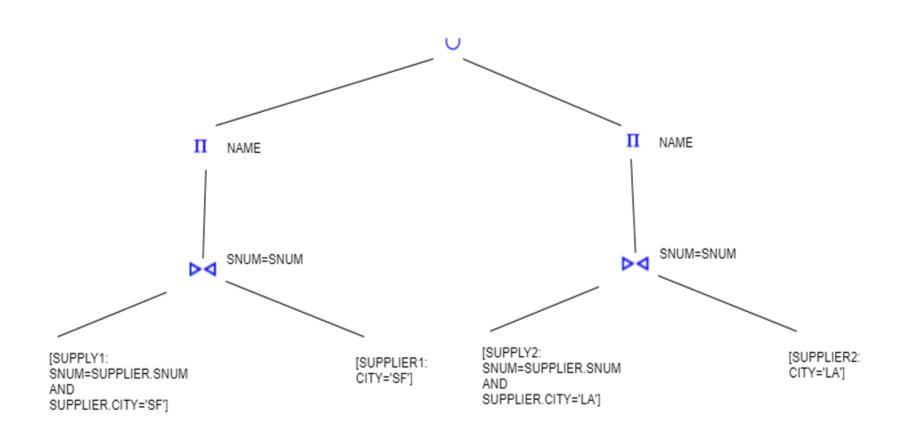
SUPPLIER (SNUM, NAME, CITY)

SUPPLY (SNUM, PNUM, DEPTNUM, QUAN)

❖ Ví dụ: Truy vấn Q4 - Hãy cho biết tên (name) của tất cả các nhà cung cấp có đơn hàng cung cấp:

Q4: ∏ NAME (SUPPLY ⊳⊲ SUPPLIER)





Phép kết phân bố dùng cho Q4

- Xét phép kết giữa hai quan hệ được phân mảnh ngang là R và S. Có 2 giải pháp
- Giải pháp 1: tập hợp tất cả các mảnh của R và S trước khi thực hiện phép kết

$$R \bowtie_F S = (U_i R_i) \bowtie_F (U_i S_i)$$

Giải pháp 2: phép kết phân tán (distrubuted join): kết các mảnh sau đó tập hợp các kết quả lại.

$$R \bowtie_F S = \bigcup_{ij} (R_i \bowtie_F S_i)$$

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

Tiêu chuẩn 5

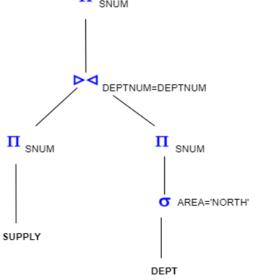
- Mục đích: biến đổi một truy vấn không có các phép kết phân tán thành một truy vấn có phép kết phân tán
- Tiêu chuẩn 5 Để phân phối các phép kết xuất hiện trong một truy vấn toàn cục, các phép hợp (biểu diễn việc tập hợp các mảnh) phải được đẩy lên phía trên các phép kết muốn phân phối

Mục đích: Mâu thuẫn giữa các điều kiện chọn của các truy vấn và các vị từ định tính của các mảnh

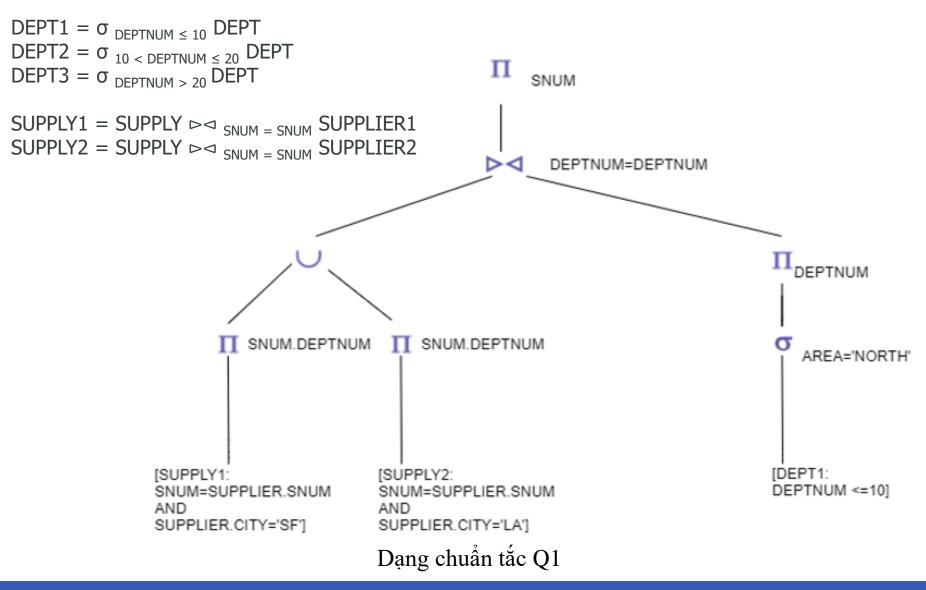
Ví dụ

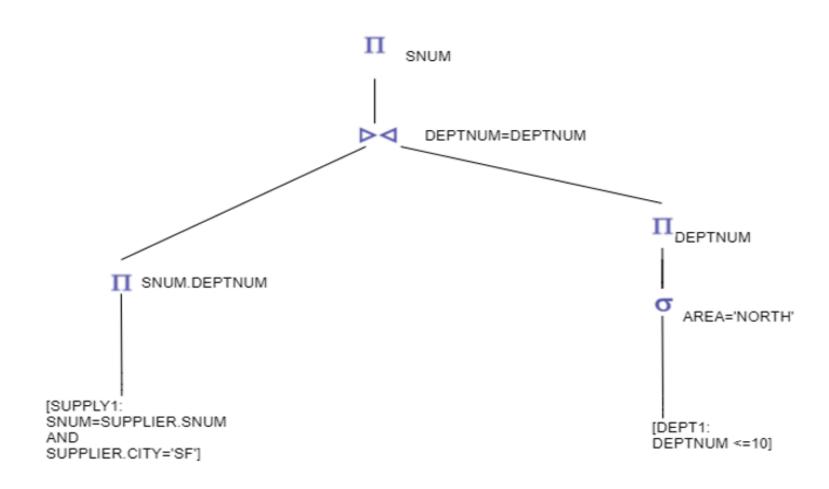
Xét truy vấn Q1 – Cho biết mã của các nhà cung cấp có đơn hàng cung cấp ở phía Bắc (NORTH)

Cây toán tử của Q1



- ❖ 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 (tương đương) sau đây: area = 'NORTH' ⇒ not (10 < deptnum ≤ 20) area = 'NORTH' ⇒ not (deptnum > 20) area = 'NORTH' ⇒ deptnum ≤ 10
- Từ (2):
 deptnum ≤ 10 ⇒ not (snum = supplier.snum and supplier.city = 'LA')





Đơn giản hóa cây toán tử bằng cách sử dụng sự suy diễn

Đơn giản hóa 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 Q5 – Hãy cho biết tên và tiền lương của các nhân viên:

Q5: Π_{NAME, SAL} EMP

Đơn giản hóa các quan hệ được phân mảnh dọc

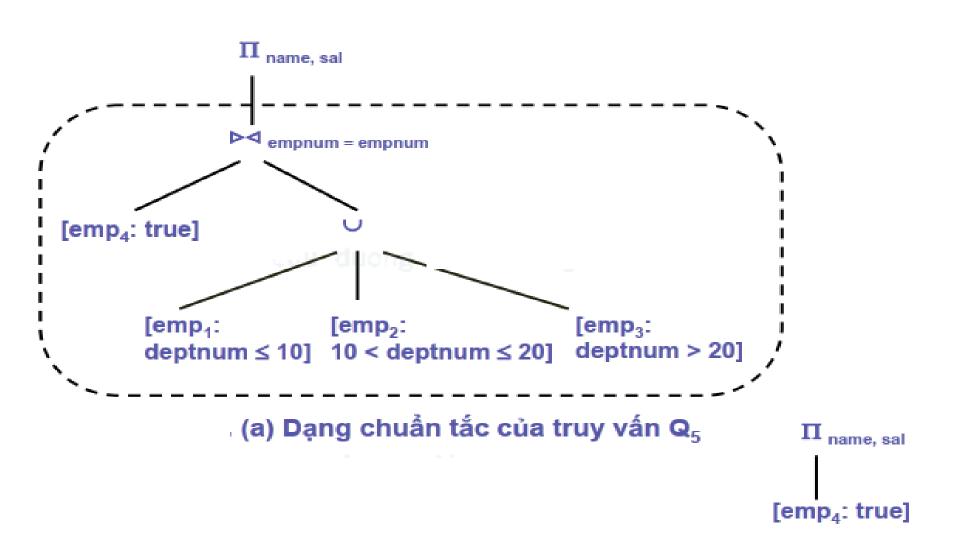
```
EMP1 = \sigma_{\text{DEPTNUM} \le 10} \Pi_{\text{EMPNUM, NAME, MGRNUM, DEPTNUM}} EMP

EMP2 = \sigma_{10 < \text{DEPTNUM} \le 20} \Pi_{\text{EMPNUM, NAME, MGRNUM, DEPTNUM}} EMP

EMP3 = \sigma_{\text{DEPTNUM} > 20} \Pi_{\text{EMPNUM, NAME, MGRNUM, DEPTNUM}} EMP

EMP4 = \Pi_{\text{EMPNUM, NAME, SAL, TAX}} EMP
```

Đơn giản hóa các quan hệ được phân mảnh dọc



$\Psi_{G,AF}R$

- 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

```
* Q6:
```

```
select AVG(quan)
from supply
where pnum = 'P1';
```

$$\Psi_{AVG(QUAN)}\,\sigma_{PNUM\,=\,{}^{'}\!P1^{''}}\,SUPPLY$$

* Q7:

SELECT SNUM, PNUM, SUM(QUAN)

FROM SUPPLY

GROUP BY SNUM, PNUM;

 Ψ SNUM, PNUM, SUM(QUAN) SUPPLY

Q8

SELECT SNUM, PNUM, SUM(QUAN)

FROM SUPPLY

GROUP BY SNUM, PNUM

HAVING SUM(QUAN) > 300;

 $\sigma_{SUM(QUAN)>300}~\Psi_{SNUM,~PNUM,~SUM(QUAN)}~SUPPLY$

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

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

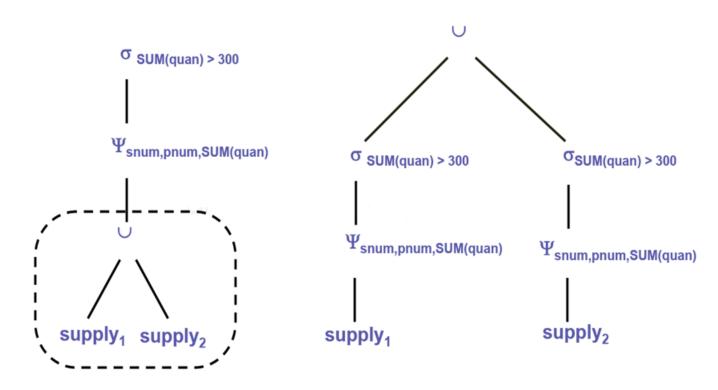
- Mỗi nhóm phải được chứa hoàn toàn trong một mảnh.
- Thực hiện phép gom nhóm trên các toán hạng của phép hợp và sau đó hợp các kết quả này.

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

 $\sigma_{\text{ SUM(QUAN)} \ > \ 300} \ \Psi_{\text{ SNUM, PNUM, SUM(QUAN)}} \ \text{SUPPLY}$



- (a) Dạng chuẩn tắc của truy vấn Q₈
- (b) Bản phân tán của truy vấn Q8

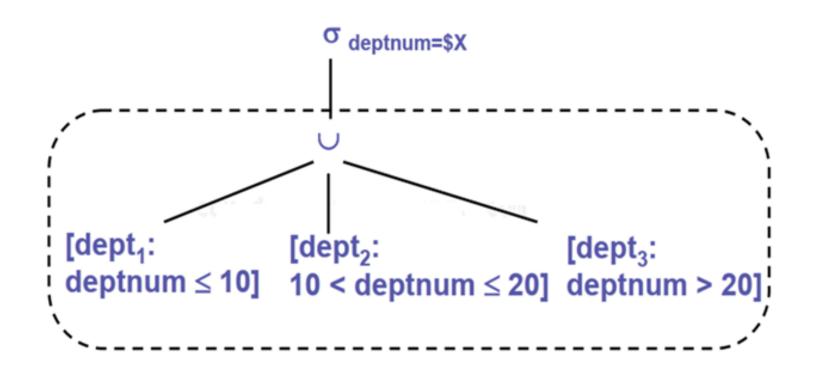
- Truy vấn có tham số (parametric query) là truy vấn mà trong đó các công thức trong các điều kiện chọn của truy vấn bao gồm các tham số mà các giá trị của chúng chưa được biết khi biên dịch truy vấn.
- Truy vấn có tham số cho phép thực hiện truy vấn nhiều lần với nhiều giá trị khác nhau của các tham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

Ví dụ Xét truy vấn Q9 - Chọn các bộ của quan hệ toàn cục dept có các mã phòng ban cho trước. Phép chọn trên deptnum có

Q9: $\sigma_{DEPTNUM = $X}$ DEPT

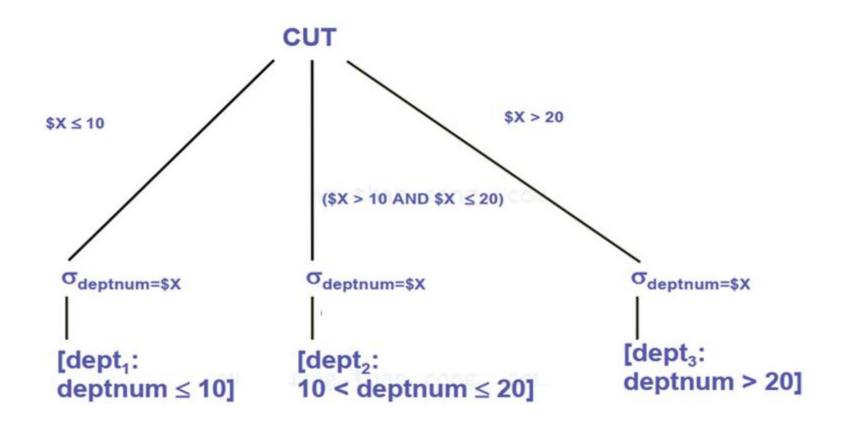
tham số:

- 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 xác định được các mảnh nào có liên quan đến truy vấn.



Dạng chuẩn tắc của cây truy vấn Q9

- 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.



Cây truy vấn với phép CUT

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

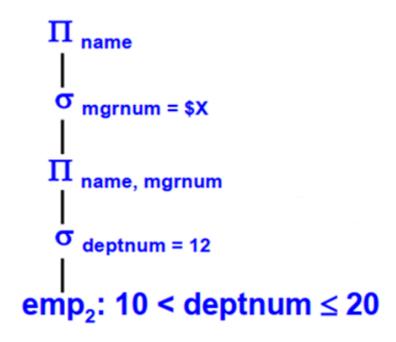
Giảm chi phí thực hiện: sử dụng các quan hệ tạm thời ở nơi gốc của truy vấn.

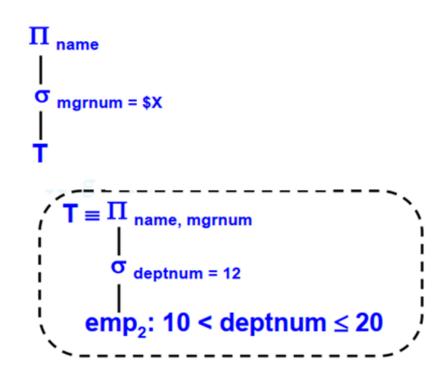
Ví dụ

Xét truy vấn Q10 - 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):

Q10: $\Pi_{NAME} \sigma_{MGRNUM = $X \text{ and } DEPTNUM = 12} EMP$

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





Sử dụng quan hệ tạm thời cho các truy vấn tham số