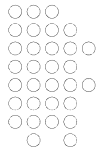


## Chương 6

### Tối ưu hóa câu truy vấn



---

---

---

---

---

---

---

---

### Nội dung chi tiết

- Giới thiệu
- Bộ biên dịch câu truy vấn (query compiler)
- Phân tích cú pháp
  - Cây phân tích (parse tree)
- Chuyển cây phân tích sang ĐSQH
  - Câu truy vấn đơn giản
  - Câu truy vấn lồng - lồng tương quan
- Quy tắc tối ưu cây truy vấn



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGTT

2

---

---

---

---

---

---

---

---

### Giới thiệu

- R(A, B, C)
- S(C, D, E)

```
SELECT  B, D
FROM    R, S
WHERE   R.A='c' AND S.E=2 AND R.C=S.C
```



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGTT

3

---

---

---

---

---

---

---

---

## Giới thiệu (tt)

- Câu truy vấn được thực hiện như thế nào?

R	A	B	C	S	C	D	E
	a	1	10		10	x	2
	b	1	10		20	y	2
	c	2	10		30	z	2
	d	2	10		40	x	1
	e	3	10		50	y	3

Kết quả

B	D
2	x

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

4

## Giới thiệu (tt)

- Cách 1
  - Tích cartesian
  - Phép chọn (selection)
  - Phép chiếu (projection)

$$\Pi_{B,D} [\sigma_{R.A=c' \wedge S.E=2 \wedge R.C=S.C} (R \times S)]$$

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

5

## Giới thiệu (tt)

RxS	A	B	C	C	D	E
	a	1	10	10	x	2
	a	1	10	20	y	2
	⋮					
	c	2	10	10	x	2
	c	2	10	20	y	2
	c	2	10	30	z	2
	⋮					

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

6

## Giới thiệu (tt)

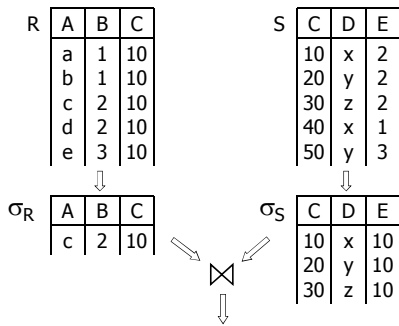
- Cách 2
  - Phép chọn (selection)
  - Phép kết (natural join)
  - Phép chiếu (projection)

$$\Pi_{B,D} [\sigma_{R.A='c'}(R) \bowtie \sigma_{S.E=2}(S)]$$

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

7

## Giới thiệu (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

8

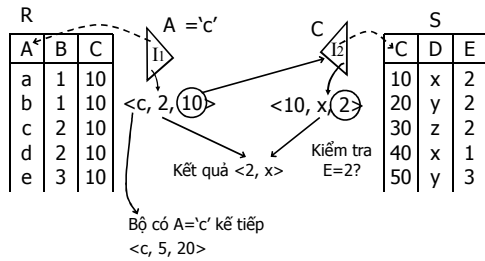
## Giới thiệu (tt)

- Cách 3 - sử dụng chỉ mục trên R.A và S.C
  - Tìm các bộ trong R thỏa R.A='c'
  - Với mỗi bộ tìm thấy, tìm tiếp các bộ trong S thỏa R.C=S.C
  - Bỏ đi những bộ S.E  $\neq$  2
  - Kết các bộ phù hợp của R và S
  - Chiếu trên thuộc tính B và D

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

9

## Giới thiệu (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

10

## Giới thiệu (tt)

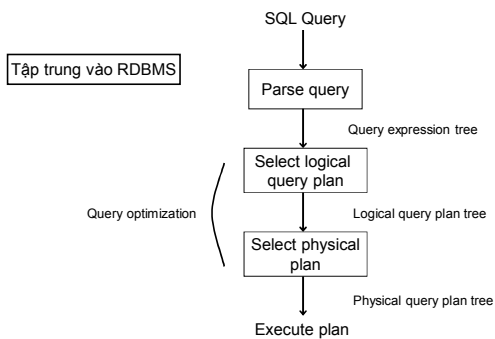
DBMS thực hiện cách nào



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

11

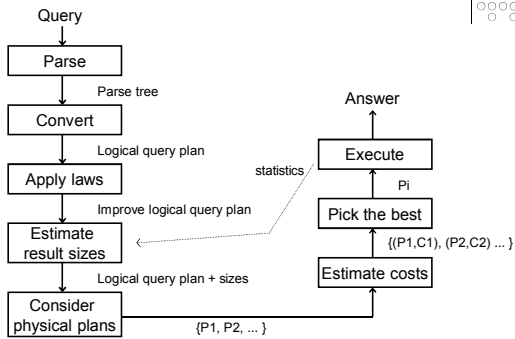
## Bộ biên dịch



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

12

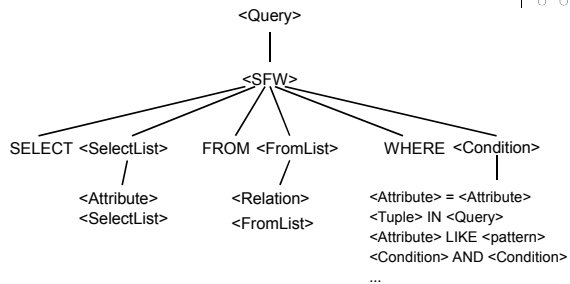
## Quá trình biên dịch



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

13

## Cây phân tích



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

14

## Ví dụ 1

- Customer(cusID, cusNm, cusStreet, cusCity)
- Account(accID, cusID, balance)

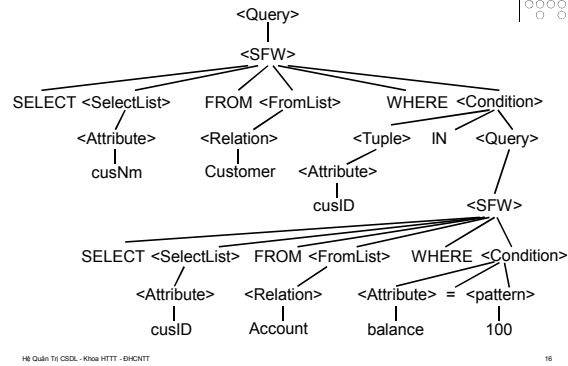
```

SELECT cusNm
FROM Customer
WHERE cusID IN (
  SELECT cusID
  FROM Account
  WHERE balance = 100)
  
```

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

15

## Ví dụ 1 (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

16

## Ví dụ 2

- Customer(cusID, cusNm, cusStreet, cusCity)
- Account(accID, cusID, balance)

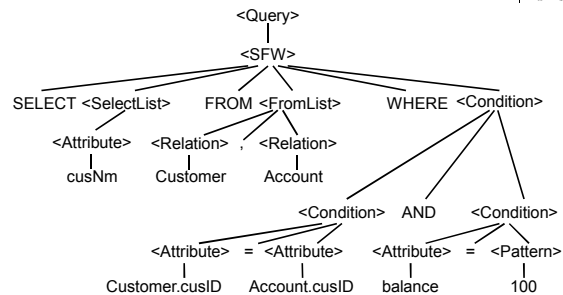
```

SELECT cusNm
FROM Customer, Account
WHERE Customer.cusID = Account.cusID
AND balance = 100
  
```

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

17

## Ví dụ 2 (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

18

## Nhận xét

- Giới hạn
  - GROUP BY
  - HAVING
  - ORDER BY
  - DISTINCT
- Aggregation function (Max, Min, Count, Sum, Avg)
- Alias name

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

19

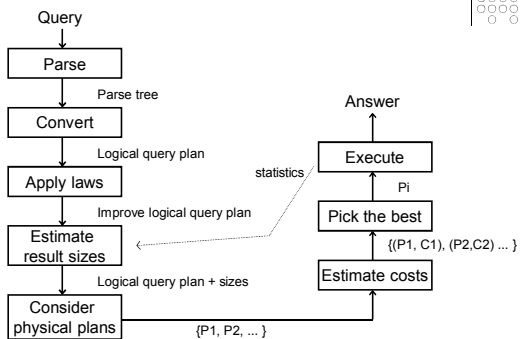
## Tiền xử lý (preprocessing)

- Kiểm tra ngữ nghĩa
  - Quan hệ
  - Thuộc tính
    - Select
    - From
  - Kiểu dữ liệu
    - Where

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

20

## Quá trình biên dịch



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

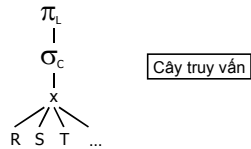
21

## Biến đổi sang ĐSQH

- Truy vấn đơn

- Xét cấu trúc <SFW>

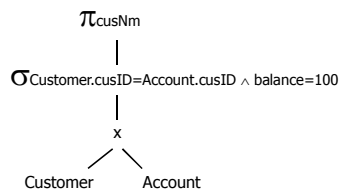
- Thay thế <FromList> thành các biến quan hệ
  - Sử dụng phép tích cartesian cho các biến quan hệ
- Thay thế <Condition> thành phép chọn  $\sigma_C$
- Thay thế <SelectList> thành phép chiếu  $\pi_L$



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

22

## Xét ví dụ 2



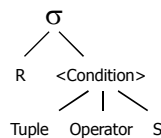
Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

23

## Biến đổi sang ĐSQH (tt)

- Truy vấn lồng

- Tồn tại câu truy vấn con S trong <Condition>
- Áp dụng qui tắc <SFW> cho truy vấn con
- Phép chọn 2 biến (two-argument selection)
  - Nút là phép chọn không có tham số
  - Nhánh con trái là biến quan hệ R
  - Nhánh con phải là <condition> áp dụng cho mỗi bộ trong R

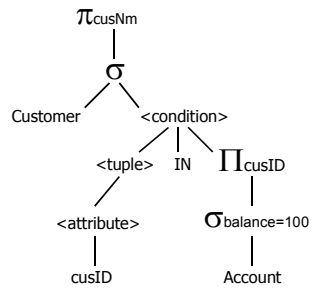


Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

24



## Xét ví dụ 1

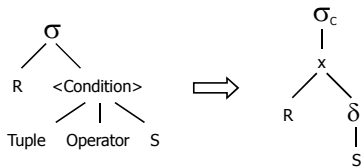


Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

25

## Biến đổi sang ĐSQH (tt)

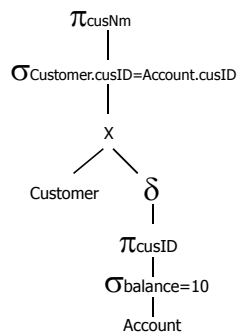
- Truy vấn lồng
- Biến đổi phép chọn 2 biến
  - Thay thế <Condition> bằng 1 cây có gốc là S
    - Nếu S có các bộ trùng nhau thì phải lược bỏ bớt bộ trùng nhau đi
    - Sử dụng phép  $\delta$
  - Thay thế phép chọn 2 biến thành  $\sigma_C$
  - $\sigma_C$  là kết quả của phép cartesian của R và S



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

26

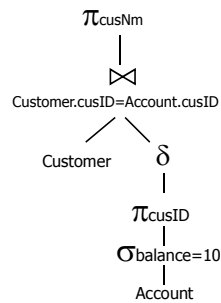
## Xét ví dụ 1 (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

27

## Xét ví dụ 1 (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

28

## Ví dụ 3

- Customer(cusID, cusNm, cusStreet, cusCity)
- Account(accID, cusID, balance)

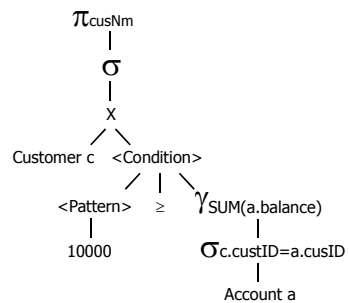
```
SELECT c.cusNm
FROM Customer c
WHERE 10000 >= (
    SELECT SUM(a.balance)
    FROM Account a
    WHERE a.cusID=c.cusID)
```

Truy vấn lồng tương quan

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

29

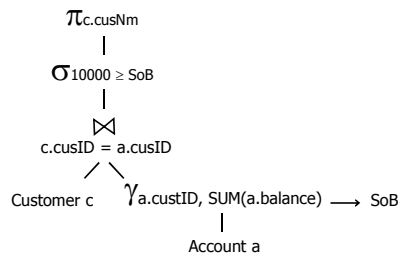
## Ví dụ 3 (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

30

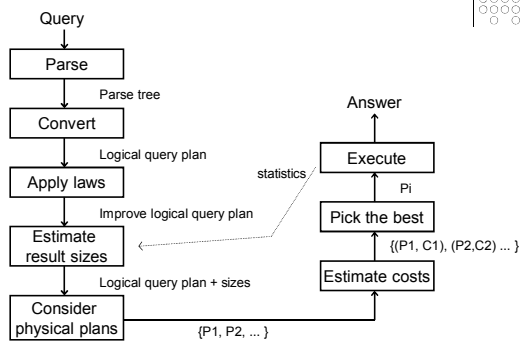
### Ví dụ 3 (tt)



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

31

### Quá trình biên dịch



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

32

### Qui tắc: Kết tự nhiên, tích cartesian, hội

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

$$(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$$

$$R \times S = S \times R$$

$$(R \times S) \times T = R \times (S \times T)$$

$$R \cup S = S \cup R$$

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

Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

33

## Qui tắc: Phép chọn $\sigma$

- Cho
  - $p$  là vị từ chỉ có các thuộc tính của  $R$
  - $q$  là vị từ chỉ có các thuộc tính của  $S$
  - $m$  là vị từ có các thuộc tính của  $R$  và  $S$

Pushing selections

$$\sigma_{p1 \wedge p2}(R) = \sigma_{p1} [\sigma_{p2}(R)]$$

$$\sigma_{p1 \vee p2}(R) = [\sigma_{p1}(R)] \cup [\sigma_{p2}(R)]$$

Quan hệ  $R$  là tập hợp  
 $\cup_s$  là phép hội trên tập hợp

Hệ Quản Trị CSDB - Khoa CNTT - ĐHQGNTT

34

---

---

---

---

---

---

---

---

## Qui tắc: $\sigma, \bowtie$

$$\sigma_p(R \bowtie S) = [\sigma_p(R)] \bowtie S$$

$$\sigma_q(R \bowtie S) = R \bowtie [\sigma_q(S)]$$

Hệ Quản Trị CSDB - Khoa CNTT - ĐHQGNTT

35

---

---

---

---

---

---

---

---

## Qui tắc: $\sigma, \bowtie$ (tt)

$$\sigma_{p \wedge q}(R \bowtie S) = [\sigma_p(R)] \bowtie [\sigma_q(S)]$$

$$\sigma_{p \wedge q \wedge m}(R \bowtie S) = \sigma_m[\sigma_p(R) \bowtie \sigma_q(S)]$$

$$\sigma_{p \vee q}(R \bowtie S) = [\sigma_p(R) \bowtie S] \cup [R \bowtie \sigma_q(S)]$$

Hệ Quản Trị CSDB - Khoa CNTT - ĐHQGNTT

36

---

---

---

---

---

---

---

---

### Qui tắc: $\sigma$ , $\cup$ và $\sigma$ , $-$



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

$$\sigma_c(R - S) = \sigma_c(R) - S = \sigma_c(R) - \sigma_c(S)$$

---

---

---

---

---

---

---

---

### Qui tắc: Phép chiếu $\pi$



- Cho
  - $X$  = tập thuộc tính con của  $R$
  - $Y$  = tập thuộc tính con của  $R$
- Ta có
  - $XY = X \cup Y$

$$\pi_{XY}(R) = \pi_X[\pi_Y(R)]$$

---

---

---

---

---

---

---

---

### Qui tắc: $\pi$ , $\bowtie$



- Cho
  - $X$  = tập thuộc tính con của  $R$
  - $Y$  = tập thuộc tính con của  $S$
  - $Z$  = tập giao thuộc tính của  $R$  và  $S$

Pushing projections

$$\pi_{XY}(R \bowtie S) = \pi_{XY}[\pi_{XZ}(R) \bowtie \pi_{YZ}(S)]$$

Except intersection and difference

---

---

---

---

---

---

---

---

## Qui tắc: $\sigma, \pi$

- Cho
  - $X$  = tập thuộc tính con của  $R$
  - $Z$  = tập thuộc tính con của  $R$  xuất hiện trong vị từ  $p$

$$\pi_X[\sigma_p(R)] = \pi_X\{\sigma_p[\pi_{XZ}(R)]\}$$

## Qui tắc: $\sigma, \pi, \bowtie$

- Cho
  - $X$  = tập thuộc tính con của  $R$
  - $Y$  = tập thuộc tính con của  $S$
  - $Z$  = tập giao thuộc tính của  $R$  và  $S$
  - $Z' = Z \cup \{\text{các thuộc tính xuất hiện trong vị từ } p\}$

$$\pi_{XY}[\sigma_p(R \bowtie S)] = \pi_{XY}\{\sigma_p[\pi_{XZ'}(R) \bowtie \pi_{YZ'}(S)]\}$$

## Nhận xét: $\sigma, \pi$

- Ví dụ
  - $R(A, B, C, D, E)$
  - $X = \{E\}$
  - $p: A=3 \wedge B='a'$

$$\pi_X[\sigma_p(R)]$$



$$\pi_E\{\sigma_p[\pi_{ABE}(R)]\}$$

Chọn trước  
tốt hơn???

Chiều trước  
tốt hơn???

## Nhận xét: $\sigma$ , $\pi$ (tt)

- Bình thường
- Chiều trước
- Nhưng
  - Giả sử A và B được cài đặt chỉ mục (index)
  - Physical query plan dùng chỉ mục để chọn ra những bộ có A=3 và B='a' trước
  - Nếu thực hiện chiều trước  $\pi_{AB}(R)$  thì chỉ mục trên A và B là vô ích
  - Chọn trước

→ Thông thường chọn trước tốt hơn

## Qui tắc: $\times$ , $\bowtie$

$$\sigma_C(R \bowtie S) = R \bowtie_C S$$

$$R \times S = \pi_L[\sigma_C(R \times S)]$$

## Qui tắc: $\delta$

$$\delta(R \bowtie S) = \delta(R) \bowtie \delta(S)$$

$$\delta(R \times S) = \delta(R) \times \delta(S)$$

$$\delta[\sigma_C(R)] = \sigma_C[\delta(R)]$$

$$\begin{aligned} \delta(R \cap_B S) &= \delta(R) \cap_B S = R \cap_B \delta(S) \\ &= \delta(R) \cap_B \delta(S) \end{aligned}$$

Except:  $\cup_B$ ,  $\neg_B$ ,  $\pi$

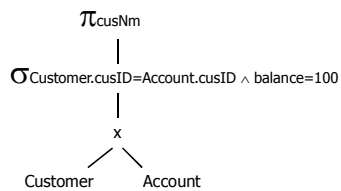
## Qui tắc: $\gamma$

- Cho
  - $X$  = tập thuộc tính trong  $R$  được gom nhóm
  - $Y = X \cup \{\text{một số thuộc tính khác của } R\}$

$$\delta[\gamma_X(R)] = \gamma_X(R)$$

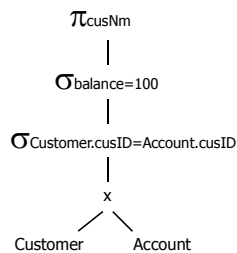
$$\gamma_X(R) = \gamma_X[\pi_Y(R)]$$

## Xét ví dụ 2



## Xét ví dụ 2

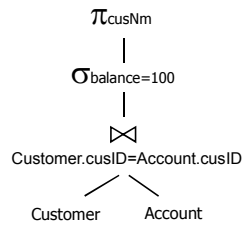
Qui tắc  $\sigma$





## Xét ví dụ 2 (tt)

Qui tắc  $\sigma$ ,  $\bowtie$



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

49

---

---

---

---

---

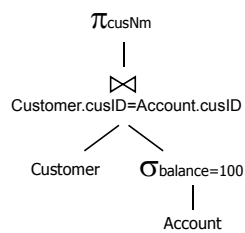
---

---

---

## Xét ví dụ 2 (tt)

Pushing  $\sigma$



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

50

---

---

---

---

---

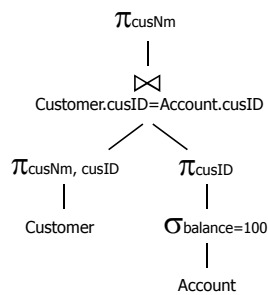
---

---

---

## Xét ví dụ 2 (tt)

Pushing  $\pi$



Hệ Quản Trị CSDL - Khoa CNTT - ĐHQGNTT

51

---

---

---

---

---

---

---

---