

Tối ưu hóa câu hỏi truy vấn

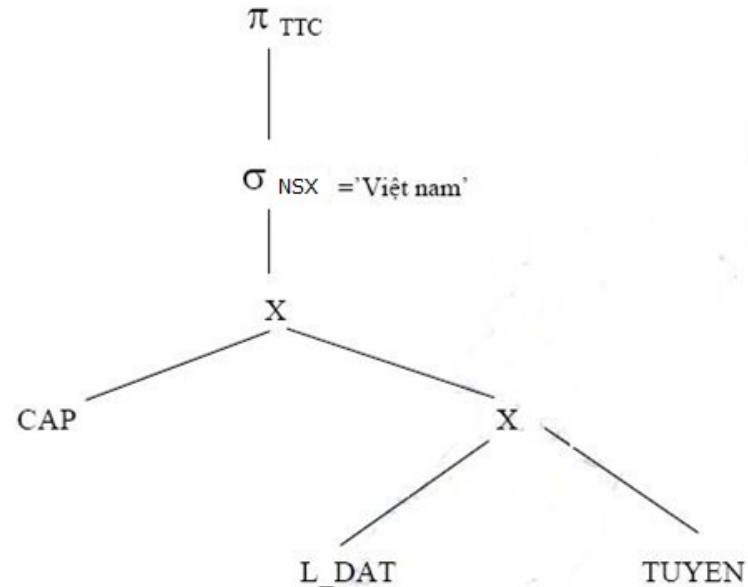
- ❑ Nguyên tắc tối ưu
- ❑ Kỹ thuật phân rã câu hỏi

Tối ưu hóa câu hỏi – Ví dụ

Liệt kê tên tuyến cáp có lắp đặt cáp Việt Nam.

$\pi_{TTC} (\sigma_{NSX = 'Việt nam'} (TUYEN \bowtie L_DAT) \bowtie CAP)$

- TUYEN (TC#, TTC, DAI)
- CAP (MC#, TC, GIA, NSX)
- L_DAT (TC#, MC#, SL, NG)

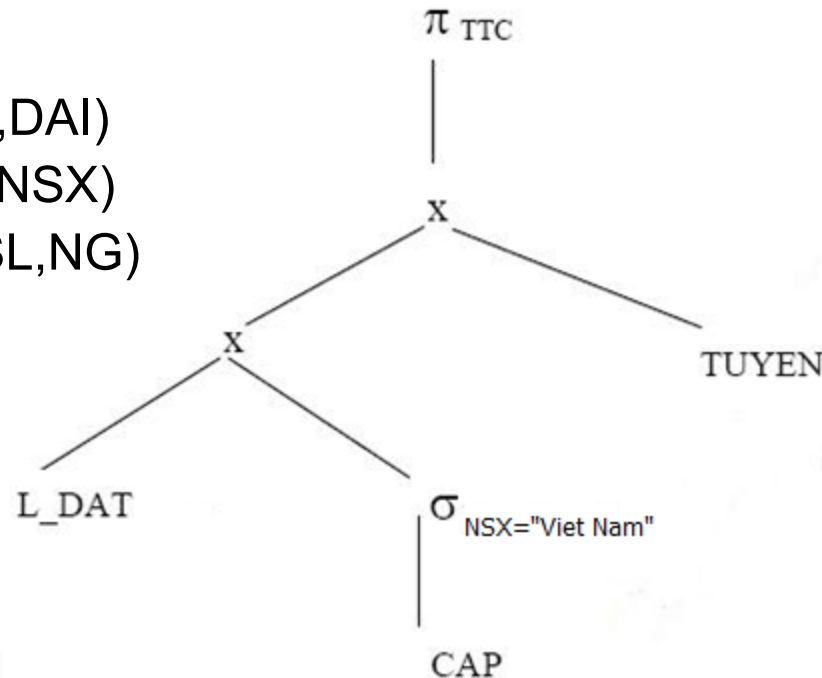


Cây đại số quan hệ chưa tối ưu hoá

Tối ưu hóa câu hỏi – Ví dụ (tt)

$\pi_{TTC}((\sigma_{L_DAT.MC\#=CAP.MC\#}(L_DAT) \times \sigma_{NSX='Việt\ nam'}(CAP)) \times TUYEN)$

- TUYEN (TC#, TTC,DAI)
- CAP(MC#,TC,GIA,NSX)
- L_DAT(TC#,MC#,SL,NG)



Cây đại số được tối ưu hóa

Nguyên tắc tối ưu

- Ưu tiên thực hiện các phép chiếu và chọn
 - ▣ Nhằm giới hạn khối lượng dữ liệu trung gian
 - ▣ Giảm chi phí truy cập bộ nhớ
- Trước khi thực hiện phép tích đề các, hãy tìm chiến lược truy cập tốt nhất vào CSDL
 - ▣ Sử dụng các phép sắp xếp ...
- Thực hiện các phép kết cân bằng sẽ rẻ hơn nhiều so với chi phí thực hiện phép tính đề các
- Nhóm các phép chọn và chiếu liên tiếp thành một phép toán duy nhất

Nguyên tắc tối ưu (tt)

- Nhóm các phép tích và chiếu liên tiếp thành một phép toán duy nhất
 - ▣ Trong khi thực hiện phép tích có thể giới hạn chi phí thực hiện bằng phép chiếu
- Tìm biểu thức chung trong một biểu thức.
 - ▣ Nếu kết quả là một quan hệ không lớn lắm nhưng tần suất xuất hiện nhiều lần, nên có biểu thức con chung
- Đánh giá sơ bộ trước khi thực hiện câu hỏi.
 - ▣ Số phép toán thực hiện, tổng chi phí thực hiện: thời gian, bộ nhớ ...

Ví dụ: Tên các loại cáp sử dụng trên tuyến “T03”

■ Ngôn ngữ đại số quan hệ

■ Cách 1:

$T1 := L_DAT \bowtie_{L_DAT.M\#=CAP.M\#} CAP$

$T2 := \sigma_{T\#="T03"}(T1)$

$KQ := \pi_{M\#,MC}(T2)$

$KQ := \pi_{M\#,MC}(\sigma_{T\#="T03"}(L_DAT \bowtie_{L_DAT.M\#=CAP.M\#} CAP))$

■ Cách 2:

$T1 := \sigma_{T\#="T03"}(L_DAT)$

$T2 := CAP \bowtie_{CAP.M\#=T1.M\#} T1$

$KQ := \pi_{M\#,MC}(T2)$

$KQ := \pi_{M\#,MC}(CAP \bowtie_{CAP.M\#=T1.M\#} (\sigma_{T\#="T03"}(L_DAT)))$

■ Ngôn ngữ SQL

```
CREATE VIEW KQ AS
SELECT R.M#,R.MC
FROM CAP R, L_DAT S
WHERE S.T#="T03" and R.M#=S.M#

SELECT * FROM KQ
```

Kỹ thuật tối ưu hóa các biểu thức đại số quan hệ - Quy tắc tương đương

$$E_1 \bowtie_F E_2 = E_2 \bowtie_F E_1$$

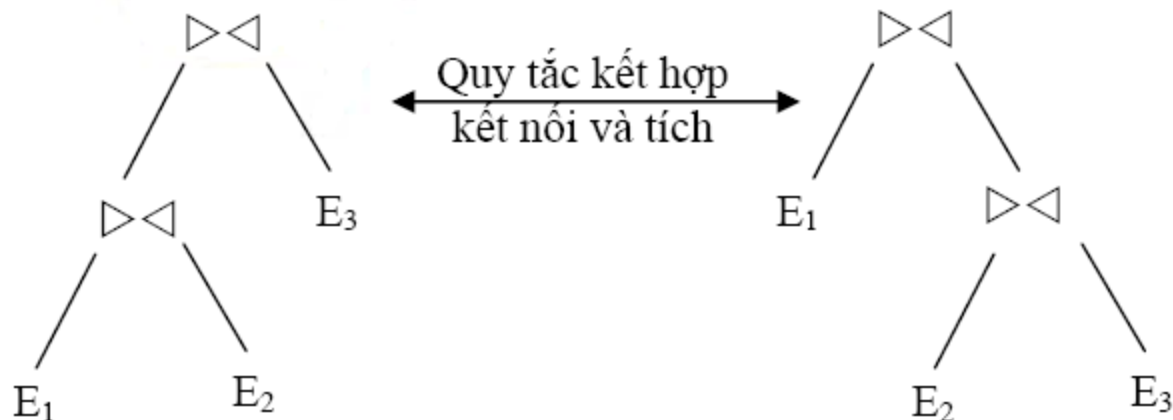
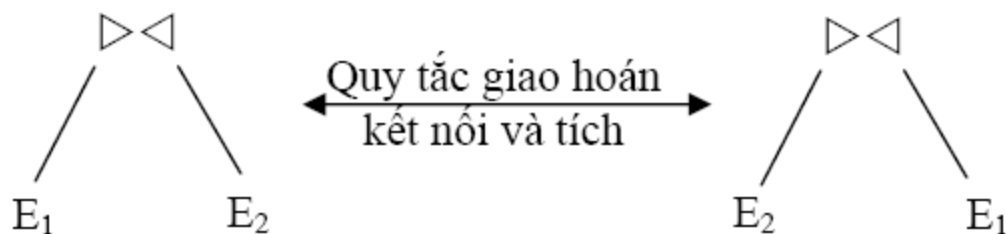
$$E_1 \bowtie E_2 = E_2 \bowtie E_1$$

$$E_1 \times E_2 = E_2 \times E_1$$

$$(E_1 \bowtie_{F1} E_2) \bowtie_{F2} E_3 = E_1 \bowtie_{F1} (E_2 \bowtie_{F2} E_3)$$

$$(E_1 \bowtie E_2) \bowtie E_3 = E_1 \bowtie (E_2 \bowtie E_3)$$

$$(E_1 \times E_2) \times E_3 = E_1 \times (E_2 \times E_3)$$



Các quy tắc cho phép chọn và phép chiếu

- Nhóm các phép chiếu thành một phép chiếu duy nhất
 - ▣ Nếu E là một biểu thức QH và A_1, A_2, \dots, A_n là các thuộc tính có mặt trong B_1, B_2, \dots, B_k , khi đó:
 - ▣ $\pi_{A_1, A_2, \dots, A_n}(\pi_{B_1, B_2, \dots, B_k}(E)) = \pi_{A_1, A_2, \dots, A_n}(E)$
- Nhóm các phép chọn thành một chuỗi phép chọn
 - ▣ Nếu E là một biểu thức QH và một điều kiện $F = F_1 \wedge F_2 \wedge \dots \wedge F_n$, khi đó:
 - $$\sigma_{F_1 \wedge F_2 \wedge \dots \wedge F_n}(E) = \sigma_{F_1}(\sigma_{F_2 \dots}(\sigma_{F_n}(E)) \dots)$$
- Giao hoán các phép chọn
 - $$\sigma_{F_1}(\sigma_{F_2}(E)) = \sigma_{F_2}(\sigma_{F_1}(E))$$

Các quy tắc cho phép chọn và phép chiếu (tt)

□ Giao hoán các phép chiếu và phép chọn:

- Nếu điều kiện F chỉ chứa các thuộc tính A_1, A_2, \dots, A_n , khi đó:

$$\pi_{A_1, A_2, \dots, A_n}(\sigma_F(E)) = \sigma_F(\pi_{A_1, A_2, \dots, A_n}(E))$$

- Nếu điều kiện F có các thuộc tính B_1, B_2, \dots, B_k không chứa các thuộc tính A_1, A_2, \dots, A_n , khi đó

$$\pi_{A_1, A_2, \dots, A_n}(\sigma_F(E)) = \pi_{A_1, A_2, \dots, A_n}(\sigma_F(\pi_{A_1, \dots, A_n, B_1, \dots, B_k}(E)))$$

□ Giao hoán phép chọn và tích Đề các

- Nếu các thuộc tính có mặt trong điều kiện F là các thuộc tính của E_1 , khi đó:

$$\sigma_F(E_1 \times E_2) = \sigma_F(E_1) \times E_2$$

- Nếu điều kiện $F = F_1 \wedge F_2$, F_1 chứa các thuộc tính của E_1 và F_2 chứa các thuộc tính của E_2 , khi đó

$$\sigma_F(E_1 \times E_2) = \sigma_{F_1}(E_1) \times \sigma_{F_2} E_2$$

- Nếu $F = F_1 \wedge F_2$, F_1 chỉ chứa các thuộc tính của E_1 và F_2 chứa các thuộc tính của E_1 và E_2 , khi đó

$$\sigma_F(E_1 \times E_2) = \sigma_{F_2}(\sigma_{F_1}(E_1) \times E_2)$$

Các quy tắc cho phép chọn và phép chiếu (tt)

- Giao hoán phép chọn và phép hợp:
 - ▣ Nếu biểu thức có dạng $E=E1 \cup E2$ và giả sử các thuộc tính của $E1$ và $E2$ cùng tên với các thuộc tính của E , F là một điều kiện, khi đó:

$$\sigma_F(E1 \cup E2) = \sigma_F(E1) \cup \sigma_F(E2)$$
- Giao hoán phép chọn và phép trừ

$$\sigma_F(E1 - E2) = \sigma_F(E1) - \sigma_F(E2)$$
- Giao hoán phép chọn và phép kết nối tự nhiên
 - ▣ Nếu điều kiện F chỉ chứa các thuộc tính chung biểu thức $E1$ và $E2$, khi đó:

$$\sigma_F(E1 \bowtie E2) = \sigma_F(E1) \bowtie \sigma_F(E2)$$

Các quy tắc cho phép chọn và phép chiếu (tt)

- Giao hoán phép chiếu và phép tích Đề các:
 - ▣ Nếu $E1$ và $E2$ là các biểu thức; $A1, A2, \dots, An$ là các thuộc tính; $B1, B2, \dots, Bk$ là các thuộc tính của biểu thức $E1$; $C1, C2, \dots, Cj$ là các thuộc tính của biểu thức $E2$, khi đó:

$$\pi_{A1, A2, \dots, An}(E1 \times E2) = \pi_{B1, B2, \dots, Bk}(E1) \times \pi_{C1, C2, \dots, Cj}(E2)$$

- Giao hoán phép chiếu và phép hợp

$$\pi_{A1, A2, \dots, An}(E1 \cup E2) = \pi_{A1, A2, \dots, An}(E1) \cup \pi_{A1, A2, \dots, An}(E2)$$

- Kết hợp phép giao và phép hợp

$$(E1 \cup E2) \cup E3 = E1 \cup (E2 \cup E3)$$

$$(E1 \cap E2) \cap E3 = E1 \cap (E2 \cap E3)$$

Các bước trong một thuật toán tối ưu hóa

- Tách phép chọn liên kết thành một chuỗi phép chọn riêng lẻ
- Cho thực hiện các phép chọn trước
 - ▣ Giao hoán, phân phối ...
- Xác định phép chọn và phép kết
 - ▣ Cho biểu thức quan hệ nhỏ nhất -> số bộ dữ liệu ít nhất
- Thay phép kết của kết quả theo tích đề các bằng điều kiện chọn.
- Tách các phép chiếu và cho thực hiện trước

Tối ưu hóa bằng khung nhìn

```
CREATE VIEW SO_LUONG (Loai_cap,SLG) AS  
SELECT MC#,SUM(SL)  
FROM L_DAT  
GROUP BY MC#
```

- Biến đổi câu hỏi để sử dụng khung nhìn
 - ▣ View : $V = R \bowtie S$
 - ▣ Câu hỏi: $R \bowtie S \bowtie T$
 - ▣ Khi đó câu hỏi có thể biến đổi và được sử dụng View:
 $V \bowtie T$

Tối ưu hóa bằng khung nhìn – Phép kết nối

□ Cho $V := R \triangleright \triangleleft S$

$$R^{\text{new}} = R^{\text{old}} \cup I_R$$

□ Chèn thêm vào R_{I_R}

$$(R^{\text{old}} \triangleright \triangleleft S) \cup (I_R \triangleright \triangleleft S). \text{ Tức là } V^{\text{new}} = V^{\text{old}} \cup (I_R \triangleright \triangleleft S).$$

□ Xóa bớt một số bộ D_R ra khỏi R

$$V^{\text{new}} = V^{\text{old}} - (D_R \triangleright \triangleleft S)$$

Tối ưu hóa bằng khung nhìn –Phép Chọn và phép Chiếu

- Xét: $V = \sigma_F(R)$:
- Thêm các bộ I_R vào R $V^{new} = V^{old} \cup \sigma_F(I_R)$
- Xóa các bộ D_R khỏi F $V^{new} = V^{old} - \sigma_F(D_R)$

- Xét: $V = \Pi_A(R)$
- Và $R=(A,B)$ có 2 bộ $(a,2)$ và $(a,3)$
- Nếu xóa bộ $(a,2)$:
 - Xóa trực tiếp từ $\Pi_A(R)$: khi đó quan hệ $\Pi_A(R)$ là quan hệ rỗng.
 - Xóa từ quan hệ R trước khi chiếu, quan hệ $\Pi_A(R)$ sẽ chứa bộ đơn (a) .

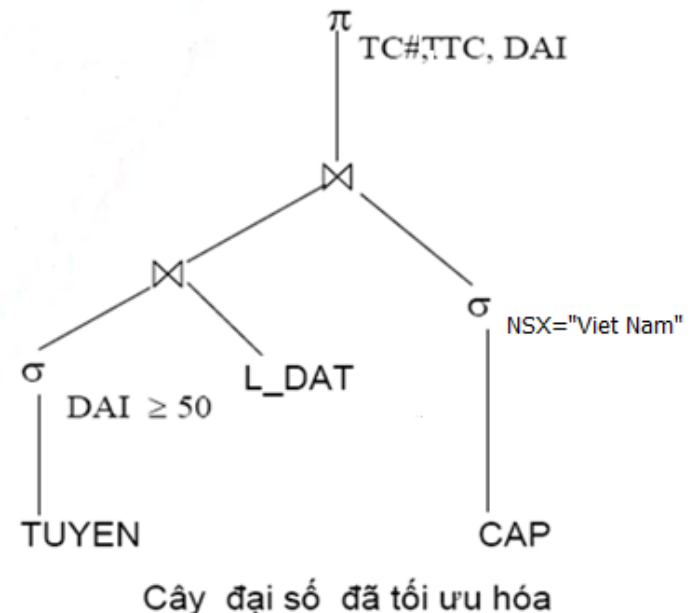
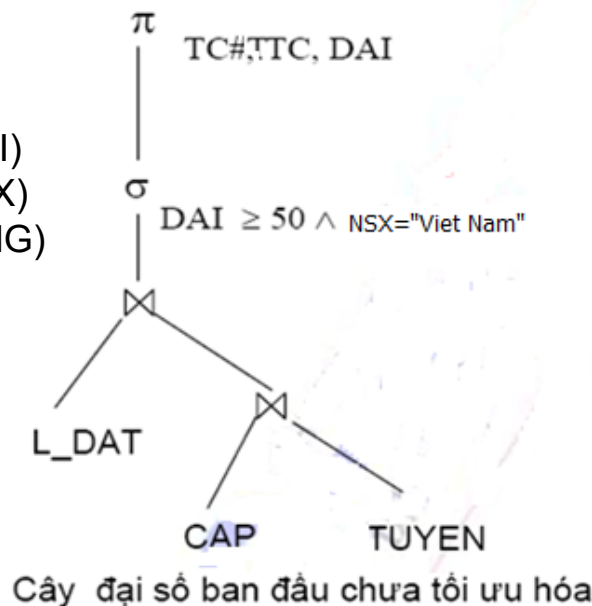
Quan trọng

- Các phép chọn sẽ làm giảm bớt số bộ dữ liệu và phép chiếu sớm làm giảm số các thuộc tính.

Ví dụ tối ưu hóa câu hỏi truy vấn

Ví dụ: Hãy liệt kê tất cả thông tin về tuyến cáp có độ dài trên 50km và lắp đặt các loại cáp do Việt nam sản xuất.

- TUYEN (TC#, TTC,DAI)
- CAP(MC#,TC,GIA,NSX)
- L_DAT(TC#,MC#,SL,NG)



Biểu thức ĐSQH ban đầu:

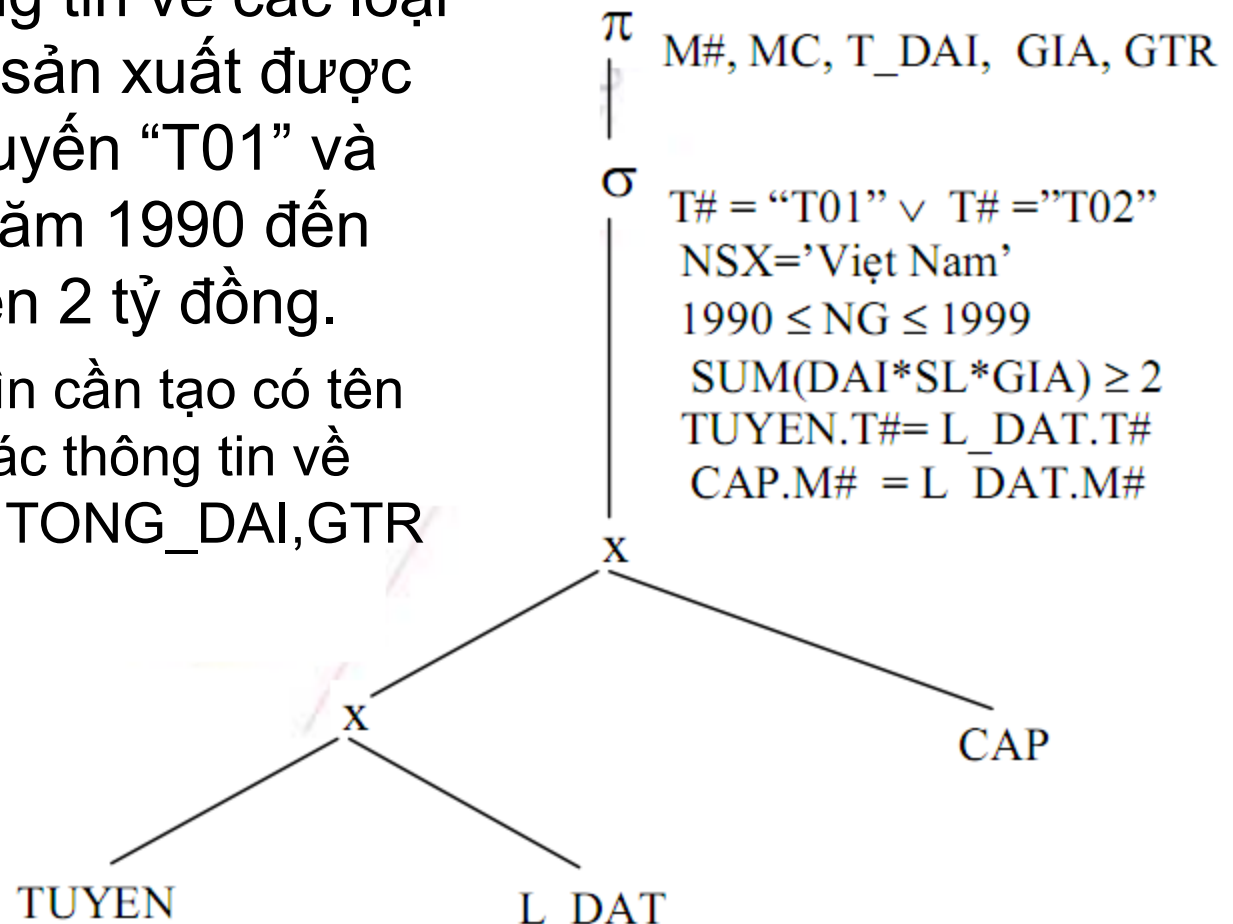
$$E = \pi_{TC\#,TTC,DAI} (\sigma_{NSX='Việt\ nam' \text{ AND } (DAI>100)} (TUYEN \bowtie CAP \bowtie L_DAT))$$

Biểu thức ĐSQH tối ưu:

$$E' = \pi_{TC\#,TTC,DAI} ((\sigma_{NSX='Việt\ nam'}(CAP) \bowtie (\sigma_{DAI>100}(TUYEN) \bowtie L_DAT))$$

Ví dụ: Phân tích tối ưu truy vấn

- Hãy cho biết thông tin về các loại cáp do Việt Nam sản xuất được lắp đặt trên các tuyến “T01” và “T02” trong các năm 1990 đến 1999 có giá trị trên 2 tỷ đồng.
- ▣ Giả sử khung nhìn cần tạo có tên CAP_VN chứa các thông tin về MC#, TMC, GIA, TONG_DAI, GTR



Cây đại số chưa tối ưu hóa

Ví dụ: Phân tích tối ưu truy vấn (tt)

$$F_1 = \{1990 \leq NG \leq 1999\}$$

$$F_2 = \{NSX = 'Việt Nam' \}$$

$$F_3 = \{T\# = "T01" \vee T\# = "T02" \}$$

$$F_4 = \{TUYEN.T\# = L_DAT.T\# \}$$

$$F_5 = \{CAP.M\# = L_DAT.M\# \}$$

$$F_6 = \{SUM(DAI * SL * GIA) \geq 2\}$$

$$T1 = \pi_{T\#,DAI}(\sigma_{F3}(TUYEN)) = \pi_{T\#,DAI}(\sigma_{T\# = "T01" \vee T\# = "T02"}(TUYEN))$$

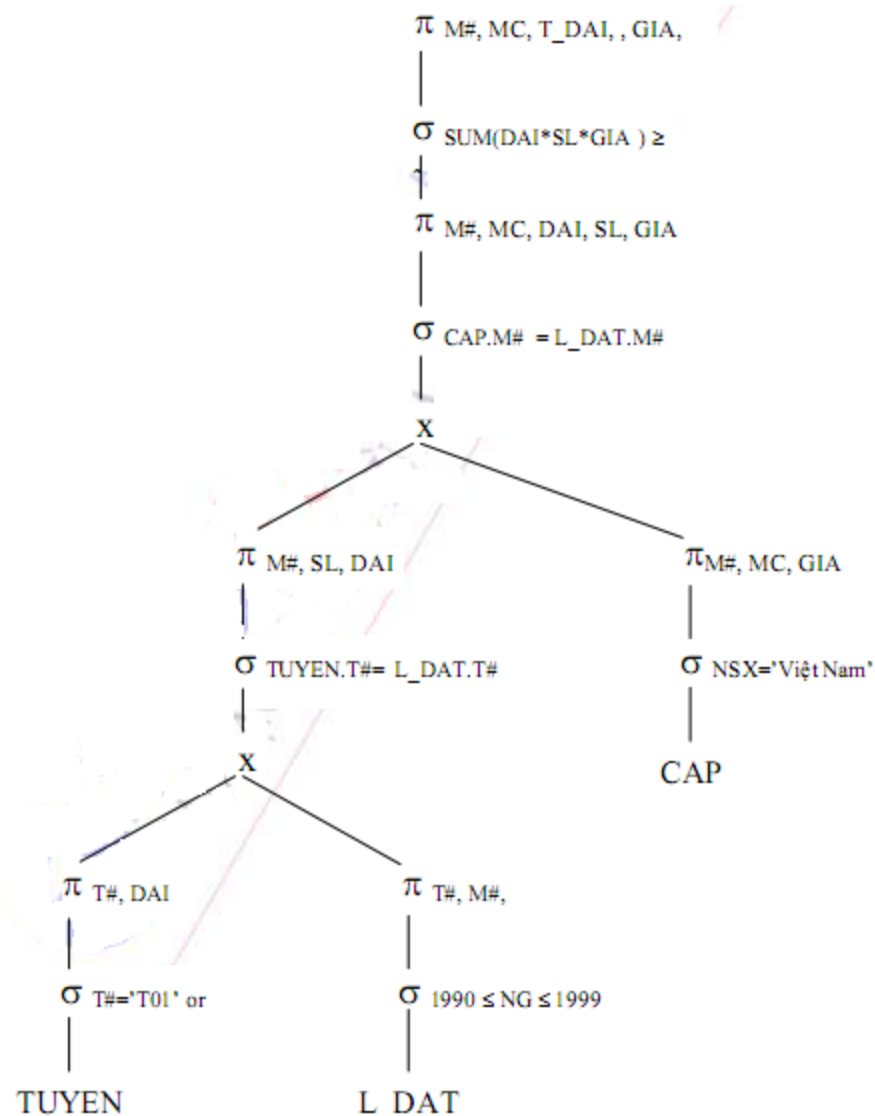
$$T2 = \pi_{T\#,M\#,SL}(\sigma_{F1}(L_DAT)) = \pi_{T\#,M\#,SL}(\sigma_{1990 \leq YEAR \leq 1999}(L_DAT))$$

$$T3 = \pi_{M\#,MC,GIA}(\sigma_{F2}(CAP)) = \pi_{M\#,MC,GIA}(\sigma_{NSX = 'Việt Nam'}(CAP))$$

$$T4 = \sigma_{F6}((T1 \bowtie_{F4} T2) \bowtie_{F5} T3)$$

$$T5 = \pi_{M\#,MC}(T4) .$$

Ví dụ: Phân tích tối ưu truy vấn (tt)



Ví dụ: Phân tích tối ưu truy vấn (tt)

```
CREATE VIEW    T1 AS
    SELECT  T#, DAI
    FROM    TUYEN
    WHERE   T#="T01" AND T#="T03"
CREATE VIEW    T2 AS
    SELECT  T#, M#, SL
    FROM    L_DAT
    WHERE   YEAR FROM NG ) BETWEEN 1990 AND 1999.
CREAT VIEW    T3 AS
    SELECT  M#, MC, GIA
    FROM    CAP
    WHERE   NSX='Việt Nam'
CREAT VIEW    T4 AS
    SELECT  M#, MC#, DAI*SL*GIA GT
    FROM    T1, T2, T3
    WHERE   (T1.T#=T2.T# and T3.M#=T2.M# ) and DAI*SL*GIA ≥ 2
SELECT  *
FROM    T4
```