

## Chöông 4 Toái ous hoùa truy vaán trong cô sôû döõ lieäu phaân taùn

### Muïc tieâu

Chöông naøy ñeà caäp ñeán vaán ñeà toái ous hoùa trong cô sôû döõ lieäu phaân taùn nghóa laø *giaûm chi phí boä nhòu trung gian, giaûm thôøi gian truy vaán cuõng nhò giaûm thôøi gian truyeàn döõ lieäu* trong caùc truy vaán phaân taùn.

Caùc vaán ñeà ñöôïc ñeà caäp trong chöông naøy nhò sau:

#### 4.1. Truy vaán. Bieáu thòuc chuaån taéc cuûa truy vaán:

Phaân naøy neâu leân khaùï nieäm veà truy vaán vaø theá naøo laø bieáu thòuc chuaån taéc cuûa moät caâu truy vaán. Bieáu thòuc chuaån taéc laø moät bieáu thòuc ñöôïc söû duïng nhieàu trong vieäc truy vaán cô sôû döõ lieäu phaân taùn.

#### 4.2. Toái ous hoùa truy vaán trong cô sôû döõ lieäu taäp trung:

Phaân naøy nhaéc laïï quaù trình toái ous hoùa moät caâu truy vaán cuïc boä, noù goàm caùc böôùc sau:

4.2.1. Böôùc 1- Kieåm tra ngôõ phaùp

4.2.2. Böôùc 2- Kieåm tra söï hôïp leä

4.2.3. Böôùc 3- Dòch truy vaán

4.2.4. Böôùc 4- Toái ous hoùa bieáu thòuc ñaïï soá quan heä

4.2.5. Böôùc 5- Choïn löïa chieán löôïc truy xuaát

4.2.6. Böôùc 6-Taïo sinh maõ

#### 4.3. Toái ous hoùa trong cô sôû döõ lieäu phaân taùn:

Phaân naøy trình baøy quaù trình toái ous hoùa moät caâu truy vaán phaân taùn, noù bao goàm caùc böôùc sau:

4.3.1. Böôùc 1 – Phaân raõ truy vaán

4.3.1.1. Böôùc 1.1- Phaân tích truy vaán

4.3.1.2. Böôùc 1.2- Chuaån hoùa ñieàu kieån cuûa meänh ñeà WHERE

4.3.1.3. Böôùc 1.3- Ñôn giaûn hoùa ñieàu kieån cuûa meänh ñeà WHERE

4.3.1.4. Böôùc 1.4- Bieán ñoãi truy vaán thaønh bieáu thòuc ñaïï soá quan heä hieäu quaù.....

4.3.1.5. Moät giaûi thuaät toái ous hoùa moät bieáu thòuc ñaïï soá quan heä treân löôïc ñoà

toaøn cuïc

4.3.2. Böôùc 2- Ñònh vò döõ lieäu

4.3.2.1. Böôùc 2.1. Bieán ñoãi bieáu thòuc ñaïï soá quan heä treân löôïc ñoà toaøn cuïc

4.3.2.2. Böôùc 2.2. Ñôn giaûn hoùa bieáu thòuc ñaïï soá quan heä treân löôïc ñoà

phaân maûnh

4.3.2.3. Moät giaûi thuaät toái oughoà moät bieâu thöüc ñaïi soá quan heä treân löôic ñoà

phaân maûnh

4.3.3. Böôùc 3- Toái oughoà truy vaán toaøn cuïc

4.3.4. Böôùc 4- Toái oughoà truy vaán cuïc boä

## Môu ñaàu

Chöông naøy trình baøy veà caùc böôùc thöïc hieän trong vieäc toái oughoà truy vaán trong cô sôu döõ lieäu taäp trung vaø trong cô sôu döõ lieäu phaân taùn, caùc tieâu chuaån toái oughoà nhaèm ñeå laøm **giaûm thôøi gian thöïc hieän truy vaán, giaûm vuøng nhôù trung gian** vaø chi phí truy vaán thoâng trong quùa trình thöïc hieän truy vaán, boä suy dieãn duøng trong vieäc ñôn giaûn hoà bieâu thöüc ñaïi soá quan heä cuùa truy vaán.

Chöông naøy sôu duïng moät cô sôu döõ lieäu sau ñaây ñeå minh hoïa cho caùc noãi dung ñöôïc trình baøy trong chöông:

Sinhvien (masv, hoten, ngaysinh, malop)

Lop (malop, tenlop, malt, tenkhoa)

Monhoc(mamh, tenmh)

Hoc (masv, mamh, Diem)

Trong ñoù :

Sinhvien : chöùa thoâng tin veà sinh vieân goàm: mã sinh vieân (masv), hoï teân (hoten),ngaysinh , thuoäc lôùp (malop). Khoùa laø masv.

Lop : chöùa thoâng tin veà lôùp hoïc goàm: mã lôùp (malop), teân lôùp (tenlop), mã lôùp Tröôùng (malt), thuoäc khoa (tenkhoa). Khoùa laø malop.

Monhoc : chöùa thoâng tin veà môân hoïc goàm: mã môân hoïc (mamh), teân môân hoïc (tenmh).

Hoc : chöùa thoâng tin veà sinh vieân (masv) hoïc môân hoïc (mamh) coù ñieäp thi cuoái Kyø (diem). Khoùa laø masv vaø mamh.

## 4.1. Truy vaán. Bieâu thöüc chuaån taéc cuùa truy vaán

### 4.1.1. Truy vaán

**Truy vaán** (query) laø moät bieâu thöüc ñöôïc bieâu dieãn baèng moät ngoân ngữ thích hôïp vaø duøng ñeå xaùc ñònh moät phaàn döõ lieäu ñöôïc chöùa trong cô sôu döõ lieäu.

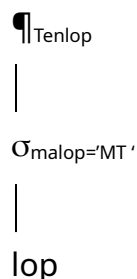
Một truy vấn có thể được thực hiện bằng cách sử dụng các phép toán của một ngôn ngữ, hoặc nó có thể được thực hiện bằng cách sử dụng công việc cần thực hiện để biểu diễn một ngôn ngữ nhằm thực hiện truy vấn cơ sở dữ liệu.

Ví dụ: Xem truy vấn cho biết tên lớp của lớp có mã lớp là 'MT'. Truy vấn này có thể được biểu diễn bằng một biểu thức liên quan sau:

$$\pi_{\text{Tenlop}}(\sigma_{\text{malop}='MT'}(\text{lop}))$$

Một truy vấn có thể được biểu diễn bằng một cây toán tử. Một cây toán tử *operator tree* của một truy vấn, còn được gọi là cây truy vấn (*query tree*) hoặc cây đại số quan hệ (*relational algebra tree*), là một cây mà một nút là một quan hệ trong cơ sở dữ liệu, và một nút khác là (một trung gian hoặc một góc) là một quan hệ trung gian được tạo ra bằng một phép toán liên quan. Chuỗi các phép toán liên quan được thực hiện để tạo ra kết quả truy vấn.

Ví dụ: Truy vấn trên có thể được biểu diễn bằng một cây toán tử như sau:



#### 4.1.3. Biểu thức chuẩn tắc của truy vấn

**Biểu thức chuẩn tắc:** của một biểu thức liên quan là một biểu thức có thể được thay thế bởi một tên quan hệ hoặc xuất hiện trong biểu thức của biểu thức tài liệu của quan hệ hoặc tên của nó.

Tổng thể, chúng ta có thể biến đổi một cây toán tử trên một ngôn ngữ thành một cây toán tử trên một ngôn ngữ phân minh bằng cách thay thế các nút của cây bằng các biểu thức chuẩn tắc của chúng. Một số quan trọng là

biểu thức truy vấn là cây toạ độ của biểu thức chuẩn tắc  
là các mệnh đề thay vì là các quan hệ toán học.

Ví dụ: Giả sử chúng ta có hai khóa tên là 'CNTT' và 'VT'. Quan hệ  
lop được phân mệnh đề ngang dọc vào tenkhoa thành hai mệnh đề *lop1*  
và *lop2*

$$Lop1 = \sigma_{tenkhoa='CNTT'}(lop)$$

$$Lop2 = \sigma_{tenkhoa='VT'}(lop)$$

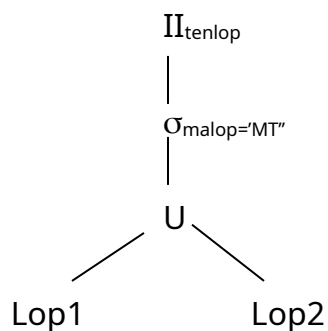
Biểu thức truy vấn của quan hệ toán học lop là :

$$Lop = lop1 \cup lop2$$

Biểu thức chuẩn tắc của biểu thức truy vấn là :

$$\Pi_{tenlop}(\sigma_{malop='MT'}(lop1 \cup lop2))$$

Thay thế quan hệ toán học **lop** trong cây toạ độ của biểu  
thức truy vấn ở trên, chúng ta được cây toạ độ như sau :



#### 4.2. Tối ưu hóa truy vấn trong cơ sở dữ liệu tập trung

Khi một hệ quản trị dữ liệu (DBMS) nhận một truy vấn viết bằng ngôn ngữ cao cấp, chúng ta gọi là SQL, DBMS thực hiện các bước sau đây:

##### 4.2.1. Bước 1- Kiểm tra ngữ pháp (Syntax Checking)

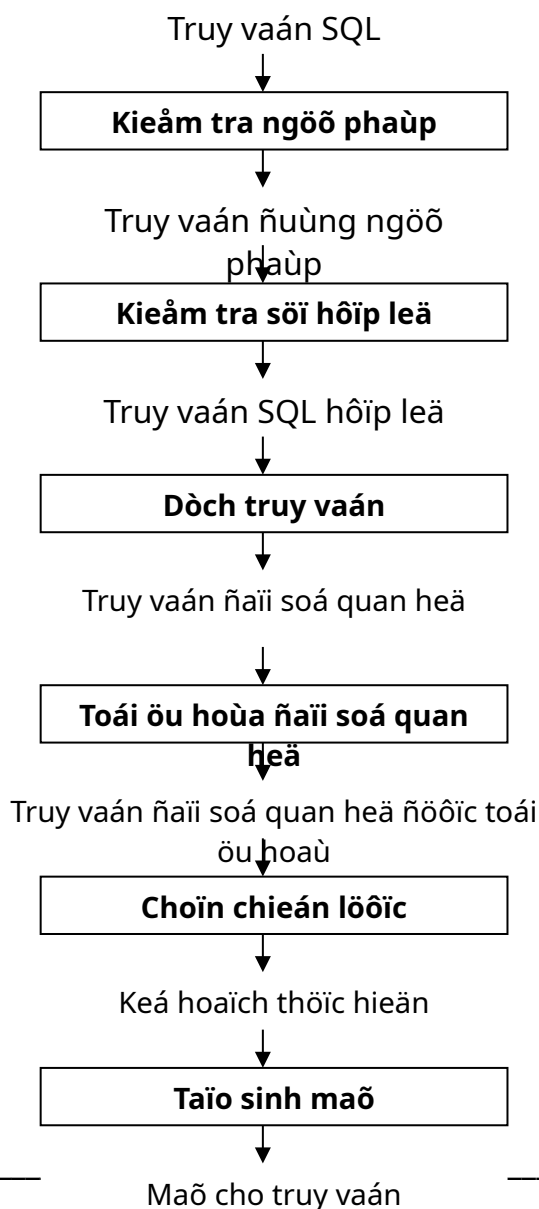
Trong bước này, DBMS sẽ kiểm tra ngữ pháp của truy vấn ban đầu (SQL query). Nếu truy vấn có sai ngữ pháp thì DBMS sẽ thông báo truy vấn có sai ngữ pháp và truy vấn này sẽ không được thực hiện. Nếu truy vấn đúng ngữ pháp (syntactically correct SQL query) thì DBMS sẽ tiếp tục thực hiện bước 2.

Ví dụ: Xét truy vấn Q1

Q1: SELECT masv, hoten FROM sinhvien;

Truy vấn này có sai ngữ pháp (viết sai từ khóa FROM)

Sơ đồ tối ưu hóa truy vấn trong cơ sở dữ liệu tập trung bao gồm các bước sau:



#### 4.2.2. Bước 2- Kiểm tra sự hợp lệ (Validation)

Trong bước này, DBMS sẽ thực hiện các công việc:

- Kiểm tra toàn tài của các đối tượng dữ liệu (các cột, các biểu, các bảng, ...) của truy vấn trong cơ sở dữ liệu.
- Kiểm tra sự hợp lệ về kiểu dữ liệu của các đối tượng dữ liệu (các cột, các biểu, vv...) trong truy vấn.

Ví dụ : Xét truy vấn Q2

Q2: SELECT masv, hoten FROM sinh\_vien ;

Truy vấn này có bảng sinh\_vien không tồn tại trong cơ sở dữ liệu.

Ví dụ: Xét truy vấn Q3

Q3: SELECT masv, hoten FROM sinhvien  
WHERE masv='123';

Truy vấn này không hợp lệ vì có cột masv (thuộc kiểu dữ liệu number) so sánh với một hằng chuỗi '123' trong mệnh đề WHERE.

Nếu truy vấn chứa các đối tượng dữ liệu không tồn tại hoặc truy vấn của các đối tượng dữ liệu không phù hợp kiểu dữ liệu với nhau thì DBMS sẽ thông báo các đối tượng dữ liệu nào không tồn tại hoặc các đối tượng dữ liệu nào không phù hợp kiểu dữ liệu và truy vấn này sẽ không được thực hiện. Nếu các đối tượng dữ liệu này đều tồn tại trong cơ sở dữ liệu (truy vấn hợp lệ – valid SQL query) thì DBMS sẽ tiếp tục thực hiện bước 3.

#### 4.2.3. Bước 3 – Dịch truy vấn (Translation)

Trong bước này, DBMS sẽ biến đổi truy vấn hợp lệ này thành một dạng biểu diễn bên trong hệ thống ô mục tháp hơn mà DBMS có thể xử lý được. Một trong các dạng biểu diễn bên trong này là việc sử dụng các đối tượng quan hệ bởi vì các phép toán các đối tượng quan hệ được biến đổi để được thực hiện các tài vụ của hệ thống : truy vấn ban đầu được biến đổi thành một biểu thức các đối tượng quan hệ hay còn gọi là truy vấn các đối tượng quan hệ (relational algebra query)

Ví dụ: Xét truy vấn Q4 sau đây cho biết các mã môn học mà các sinh viên thuộc lớp có mã 'MT' học.

Q4 : SELECT DISTINCT mamh  
FROM sinhvien, hoc  
WHERE sinhvien.masv=hoc.masv AND malop='MT'

Truy vấn này sẽ được biến đổi thành biểu thức nối số quan hệ như sau :

$\Pi_{mamh}(\sigma_{malop='MT'}(sinhvien \bowtie hoc))$

#### 4.2.4. Böluc 4- Tối ưu hóa biểu thức ñại số quan hệ (relational Algebra Optimization)

Trong böluc này DBMS sẽ dùng các phép biến ñổi tổng ñồng của ñại số quan hệ ñể biến ñổi biểu thức ñại số quan hệ có ñôi ñôi ở böluc 3 thành một biểu thức ñại số quan hệ tổng ñồng (theo nghĩa chúng có cùng một kết quả) ñồng biểu thức sau sẽ hiệu quả hơn: loại bỏ các phép toán không cần thiết và giảm vòng ñầu trung gian. Cuối böluc này, DBMS tạo ra một truy vấn ñại số quan hệ ñã ñôi ñôi tối ưu hóa (optimized relational algebra query).

Ví dụ: Biểu thức quan hệ của truy vấn Q4 ở cuối böluc 3 có thể ñôi ñôi biến ñổi thành biểu thức ñại số quan hệ tổng ñồng ñồng toát ñơn ñó sau:

$$\Pi_{msh}(\Pi_{masv}(\sigma_{malop='MT'}(sinhvien))) \bowtie \Pi_{masv,msh}(hoc)$$
  
 SELECT DISTINCT msh  
 FROM (SELECT MASV FROM sinhvien WHERE MALOP='MT') SV  
 (SELECT MASV, MAMH FROM hoc) HOC  
 WHERE SV.masv=hoc.masv

#### 4.2.5. Böluc 5- Chọn lựa chiến lược truy xuất (strategy selection)

Trong böluc này, DBMS sẽ dùng các thông số về kích thước của các bảng, các đặc điểm v.v... ñể xác ñịnh cách xử lý truy vấn. DBMS sẽ ñánh giá chi phí của các kế hoạch thực hiện khác nhau có thể có ñể chọn ra một kế hoạch thực hiện (execution plan) tốt nhất sao cho tốn ít chi phí nhất (thời gian xử lý và vòng ñầu trung gian). Các thông số dùng ñể ñánh giá chi phí của kế hoạch thực hiện gồm: số lần và loại truy xuất ñó, kích thước của vòng ñầu chính và vòng ñầu ngoài, và thời gian thực hiện của các tài vụ ñể tạo ra kết quả của truy vấn. Cuối böluc này, DBMS tạo ra một kế hoạch thực hiện cho truy vấn.

#### 4.2.6. Böluc 6- Tạo sinh mã (code generation)

Trong böluc này, kế hoạch thực hiện của truy vấn có ñôi ñôi ở cuối böluc 5 sẽ ñôi ñôi mã hóa và ñôi ñôi thực hiện.



### 4.3. Tối ưu hóa truy vấn trong cơ sở dữ liệu phân tán

Tối ưu hóa truy vấn trong cơ sở dữ liệu phân tán bao gồm một số bước như của tối ưu hóa truy vấn trong cơ sở dữ liệu tập trung và một số bước tối ưu hóa có liên quan đến phân tán dữ liệu.

#### 4.3.1. Bước 1- Phân rã truy vấn (Query Decomposition)

Bước này có thể gọi là bước **Tối ưu hóa truy vấn trên lược đồ toàn cục**. Bước này giống với các bước 1, 2, 3 và 4 của tối ưu hóa truy vấn trong cơ sở dữ liệu tập trung, nhằm để biến đổi một truy vấn viết bằng ngôn ngữ cấp cao, chẳng hạn SQL, thành một biểu thức nào đó có quan hệ tương đương (theo nghĩa đúng cho ra cùng một kết quả) và hiệu quả (theo nghĩa loại bỏ các phép toán nào đó có quan hệ không cần thiết, giảm xuống một mức độ nào đó). Bước này chia ra các bước phân tán dữ liệu.

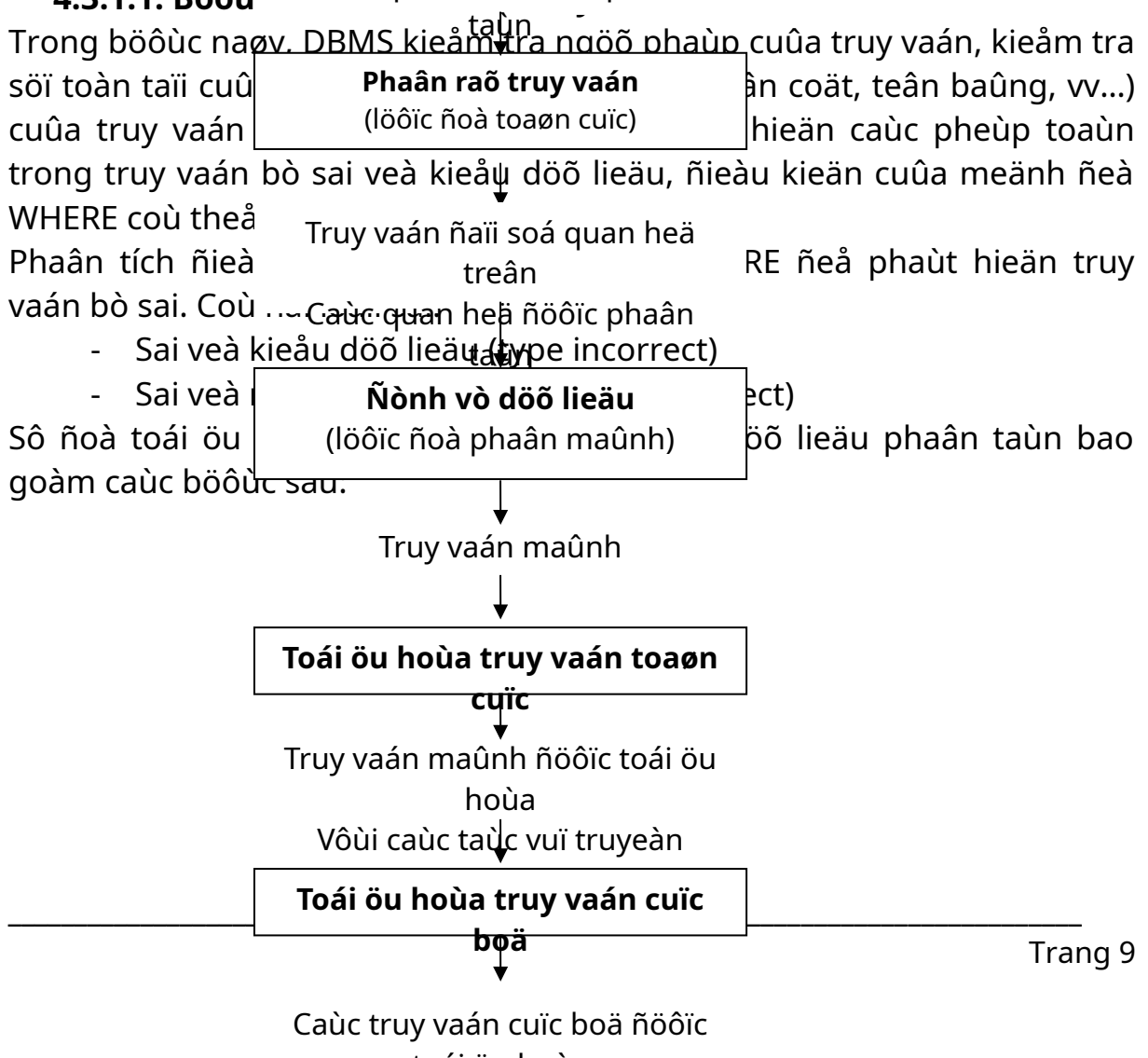
Tối ưu hóa truy vấn trên lược đồ toàn cục bao gồm 4 bước sau:

**4.3.1.1. Bước 1** Các quan hệ lược đồ phân tán

Trong bước này, DBMS kiểm tra ngữ pháp của truy vấn, kiểm tra số toàn thể của truy vấn, phân rã truy vấn (lược đồ toàn cục) thành các phép toán trong truy vấn để sai về kiểu dữ liệu, nhiều kiểu của mệnh đề WHERE có thể phân tích được. Truy vấn nào đó có thể RE để phân rã truy vấn để sai. Có thể có các quan hệ lược đồ phân tán

- Sai về kiểu dữ liệu (type incorrect)
- Sai về lược đồ (schema incorrect)

Số lược đồ tối ưu hóa truy vấn phân tán bao gồm các bước sau.



### **Truy vấn bỏ sai về kiểu dữ liệu**

Một truy vấn bỏ sai về kiểu dữ liệu nếu các thuộc tính của nó hoặc các liên quan hệ không thỏa mãn trong logic của giao diện, hoặc nếu các phép toán thỏa mãn để các thuộc tính bỏ sai về kiểu dữ liệu.

Nếu giả sử quyết định cho vấn đề này, trong logic của toán tử chúng ta phải mô tả kiểu dữ liệu của các thuộc tính của các liên quan hệ.

Ví dụ: Xét truy vấn Q5:

```
Q5: SELECT mssv, hoten FROM sinhvien  
WHERE masv='123';
```

Truy vấn này có hai lỗi sai:

- (1) mssv không tồn tại trong liên quan hệ sinhvien, và
- (2) masv thuộc kiểu number không thể so sánh với hằng chuỗi '123'.

### **Truy vấn bỏ sai về ngữ nghĩa**

Một truy vấn bỏ sai về ngữ nghĩa nếu nó có chứa các thành phần không tham gia vào quá trình tạo ra kết quả của truy vấn.

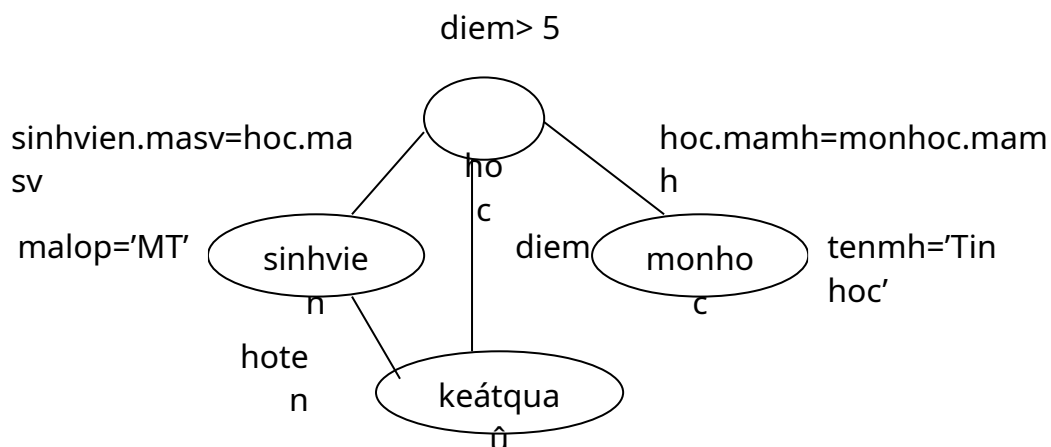
Nếu phát hiện một truy vấn bỏ sai về ngữ nghĩa, chúng ta dùng một số thủ tục truy vấn (query graph) hoặc số kết nối liên quan hệ (relation connection graph) cho các truy vấn có chứa các phép join, phép chiếu và phép kết. Trong một số thủ tục truy vấn, một nút biểu diễn cho một liên quan hệ kết quả (result

relation) và các nút khác biểu diễn cho các quan hệ toán học (operand relation). Một cạnh giữa hai nút quan hệ toán học biểu diễn cho một phép kết, một cạnh giữa một nút quan hệ toán học với một nút quan hệ kết quả biểu diễn cho một phép chiếu. Một nút quan hệ toán học có thể chứa một hoặc nhiều kiến thức. Một node con quan trọng của node này là node kết quả (join graph) được dùng trong bước tối ưu hóa truy vấn.

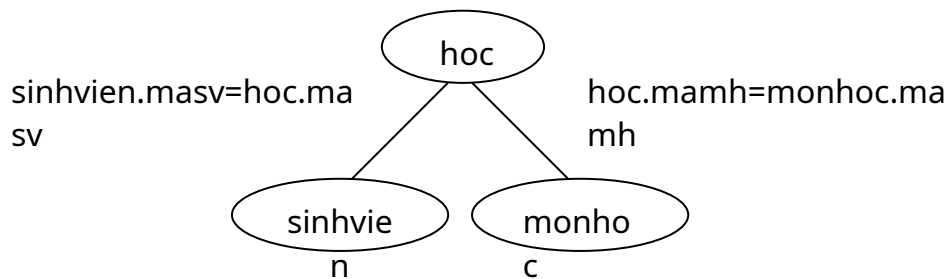
Ví dụ: Xét truy vấn Q6 liệt kê họ tên sinh viên và điểm của môn học 'Tin học' của lớp 'MT' với điểm ít nhất 5.

```
Q6: SELECT hoten, diem
      FROM sinhvien, hoc, monhoc
      WHERE sinhvien.masv=hoc.masv
            AND hoc.mamh=monhoc.mamh
            AND malop='MT' AND diem > 5 AND tenmh = 'Tin hoc';
```

Node truy vấn của truy vấn này như sau:



Vaø ñoà thò keát noái töông öùng laø:

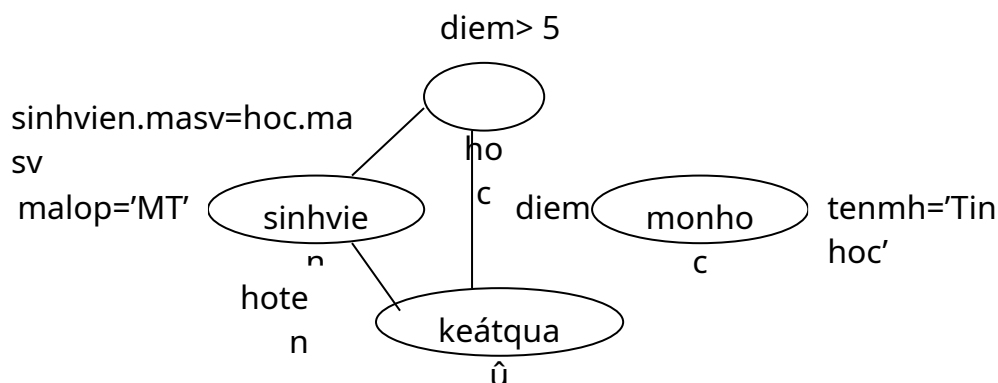


Một truy vấn bỏ sai về ngữ nghĩa nếu ñoà thò truy vấn của nó là không liên thông. Ñoà thò không liên thông là một ñoà thò bao gồm nhiều thành phần liên thông, mỗi thành phần liên thông là một ñoà thò con riêng biệt, hai thành phần liên thông không đối nối với nhau thông qua các cạnh. Trong trường hợp này, một truy vấn ñối xem là ñùng ñến bằng cách chæ giõ ñể thành phần có liên quan ñến quan hệ kết quả và loại bỏ các thành phần còn lại.

Ví dụ: Xét truy vấn Q7

Q7: SELECT hoten, diem  
FROM sinhvien, hoc, monhoc  
WHERE sinhvien.masv=hoc.masv  
AND malop='MT' AND diem > 5 AND tenmh = 'Tin hoc';

Ñoà thò truy vấn của truy vấn này như sau:



Ñoà thò truy vấn của truy vấn này là không liên thông, nên truy vấn bỏ sai về ngữ nghĩa. Có ba giải pháp cho vấn đề này là:

- (1) Huỷ bỏ truy vấn này.
- (2) Huỷ bỏ các bảng không cần thiết trong mệnh đề From và các điều kiện có liên quan ñến các bảng này trong mệnh đề WHERE.

Giaû söû truy xuaát ñeán monhoc laø khoâng caàn thieát, ta huûy boû baûng monhoc trong meänh ñeà From vaø ñieàu kieän tenmh ='Tin hoc' trong meänh ñeà WHERE. Ta coù truy vaán Q8 nhö sau:

Q8: SELECT hoten, diem  
FROM sinhvien, hoc  
WHERE sinhvien.masv = hoc.masv AND malop = 'MT' AND diem > 5;

(3) Bổ sung điều kiện kết sao cho nó là truy vấn nối tiếp liên thông. Một truy vấn nối tiếp liên thông không bị sai ngữ nghĩa nếu nó là truy vấn đơn (có nhiều nhất một phép nối tiếp), liên thông và số phép nối tiếp bằng số phép nối tiếp.

Bổ sung điều kiện kết hoc.mamh = monhoc.mamh vào trong mệnh đề WHERE.

Ta có truy vấn Q9:

Q9: SELECT hoten, diem  
FROM sinhvien, hoc  
WHERE sinhvien.masv = hoc.masv  
AND hoc.mamh = monhoc.mamh AND malop = 'MT' AND diem > 5  
AND tenmh = 'Tin hoc';

#### 4.3.1.2. Bổ túc 1.2- Chuẩn hóa điều kiện của mệnh đề WHERE

Điều kiện ghi trong mệnh đề WHERE là một biểu thức logic logic có thể bao gồm các phép toán logic (not, and, or) nối tiếp viết dưới một dạng bất kỳ. Ký hiệu các phép toán logic logic: not (-), and (^), or (v). Bổ túc này nhằm mục đích chuẩn hóa điều kiện của mệnh đề Where về một trong hai dạng chuẩn:

- Dạng chuẩn giao (conjunctive normal form)  

$$(P_{11} \vee P_{12} \vee \dots \vee P_{1n}) \wedge \dots \wedge (P_{m1} \vee P_{m2} \vee \dots \vee P_{mn})$$
- Dạng chuẩn hợp (disjunctive normal form)  

$$(P_{11} \wedge P_{12} \wedge \dots \wedge P_{1n}) \vee \dots \vee (P_{m1} \wedge P_{m2} \wedge \dots \wedge P_{mn})$$

trong đó  $P_{ij}$  là một biến logic logic (có giá trị là true hoặc false) hoặc là một vị từ đơn giản (simple predicate) có dạng:

$$a R b$$

với  $a, b$  là các biểu thức số học hoặc  $R$  là một trong những phép toán so sánh:

---

=	baèng
< > hoaëc !=	khoâng baèng
<	nhoû hôn
<=	nhoû hôn hoaëc baèng
>	lòùn hôn
>=	lòùn hôn hoaëc baèng

Ñeã bieán ñoãi ñieàu kieän cuûa meänh ñeà WHERE veà moät trong hai daïng chuaån treân, chuùng ta söû duïng caùc pheùp bieán ñoãi töông ñöông cuûa caùc pheùp toaùn luaän lyù.

Kyù hieäu  $\equiv$  laø söï töông ñöông.

Caùc pheùp bieán ñoãi töông ñöông:

- (1)  $P_1 \wedge P_2 \equiv P_2 \wedge P_1$
- (2)  $P_1 \vee P_2 \equiv P_2 \vee P_1$
- (3)  $P_1 \wedge (P_2 \wedge P_3) \equiv (P_1 \wedge P_2) \wedge P_3$
- (4)  $P_1 \vee (P_2 \vee P_3) \equiv (P_1 \vee P_2) \vee P_3$
- (5)  $P_1 \wedge (P_2 \vee P_3) \equiv (P_1 \wedge P_2) \vee (P_1 \wedge P_3)$
- (6)  $P_1 \vee (P_2 \wedge P_3) \equiv (P_1 \vee P_2) \wedge (P_1 \vee P_3)$
- (7)  $\neg(P_1 \wedge P_2) \equiv \neg P_1 \vee \neg P_2$
- (8)  $\neg(P_1 \vee P_2) \equiv \neg P_1 \wedge \neg P_2$
- (9)  $\neg(\neg P) \equiv P$

Ví duï: Xeùt truy vaán Q10

Q10: SELECT malop  
FROM sinhvien  
WHERE ( malop<>'MT1'  
AND (malop='MT1' OR malop='MT2')  
AND malop <>'MT2' ) OR hoten='Nam';

Ñieàu kieän q cuûa meänh ñeà WHERE laø:

(NOT (malop='MT1') AND (malop='MT1' OR malop='MT2')  
AND NOT (malop='MT2')) OR hoten='Nam'

Kyù hieäu:

$P_1$  laø malop='MT1'

$P_2$  laø malop='MT2'

$P_3$  laø hoten='Nam'

Nieàu kieän q seõ laø:

$$(\neg P_1 \wedge (P_1 \vee P_2) \wedge \neg P_2) \vee P_3$$

Baèng caùch àùp düng caùc pheùp bieán ñoãi (3), (5) ñeã ñöa ñieàu kieän q veà daïng chuaån hôïp:

$$(((\neg P_1 \wedge P_1) \vee (\neg P_1 \wedge P_2)) \wedge \neg P_2) \vee P_3$$

$$(\neg P_1 \wedge P_2 \wedge \neg P_2) \vee P_3 = P_3$$

#### 4.3.1.3. Böôùc 1.3- Ñôn giaün hoàu ñieàu kieän cuûa meänh ñeà WHERE

Böôùc naøy söu düng caùc pheùp bieán ñoãi töông ñöông cuûa caùc pheùp toàùn luaän lyù (not, and, or) ñeã ruùt goïn ñieàu kieän cuûa meänh ñeà WHERE.

Caùc pheùp bieán ñoãi töông ñöông goàm coù :

- (10)  $P \wedge P \equiv P$
- (11)  $P \vee P \equiv P$
- (12)  $P \wedge \text{true} \equiv P$
- (13)  $P \vee \text{false} \equiv P$
- (14)  $P \wedge \text{false} \equiv \text{false}$
- (15)  $P \vee \text{true} \equiv \text{true}$
- (16)  $P \wedge \neg P \equiv \text{false}$
- (17)  $P \vee \neg P \equiv \text{true}$
- (18)  $P_1 \wedge (P_1 \vee P_2 \vee P_3) \equiv P_1$
- (19)  $P_1 \vee (P_1 \wedge P_2 \wedge P_3) \equiv P_1$

Víduï: Xeùt truy vaán Q10 ôû treân, ñieàu kieän q ôû daïng chuaån hôïp laø:

$$(\neg P_1 \wedge P_1 \wedge \neg P_2) \vee (\neg P_1 \wedge P_2 \wedge \neg P_2) \vee P_3$$

Baèng caùch àùp düng pheùp bieán ñoãi (16), chuùng ta ñöôïc:

$$(\text{false} \wedge \neg P_2) \vee (\neg P_1 \wedge \text{false}) \vee P_3$$

Aùp düng pheùp bieán ñoãi (14), cuùng ta ñöôïc:

$$\text{False} \vee \text{False} \vee P_3$$



Aùp dũng pheùp bieán ñoái (15), chuùng ta ñöôïc dieàu kieän q cuoái cuøng laø  $P_3$ , töùc laø  $hoten = \text{'Nam'}$ . Vaây truy vaán Q10 trôø thaønh truy vaán Q11 nhö sau:

```
Q11: SELECT malop
      FROM sinhvien
      WHERE hoten='Nam';
```

#### **4.3.1.4. Böôùc 1.4- Bieán ñoái truy vaán thaønh moät bieáu thöùc ñaïi soá quan heä hieäu quaù**

Böôùc naøy söù dũng caùc pheùp bieán ñoái töông ñöông cuûa caùc pheùp toaùn ñaïi soá quan heä nhaèm ñeå loaï boû caùc pheùp toaùn ñaïi soá quan heä khoâng caàn thieát vaø giaûm vuøng nhôø trung gian ñöôïc söù dũng trong quaù trình thöïc hieän caùc pheùp toaùn ñaïi soá quan heä caàn thieát cho truy vaán.

Böôùc naøy bao goàm hai böôùc sau ñaây:

**Böôùc 1.4.1** – Bieán ñoái truy vaán thaønh moät bieáu thöùc ñaïi soá quan heä, bieáu dieãn bieáu thöùc ñaïi soá quan heä naøy baèng moät caây toaùn töù.

**Böôùc 1.4.2** – Ñôn giaûn hoùa caây toaùn töù ñeå coù ñöôïc moät bieáu thöùc ñaïi soá quan heä hieäu quaù.

##### **Böôùc 1.4.1. Bieáu dieãn truy vaán baèng caây toaùn töù**

Quaù trình bieán ñoái moät truy vaán ñöôïc vieát baèng leänh SELECT thaønh moät caây toaùn töù bao goàm caùc böôùc sau:

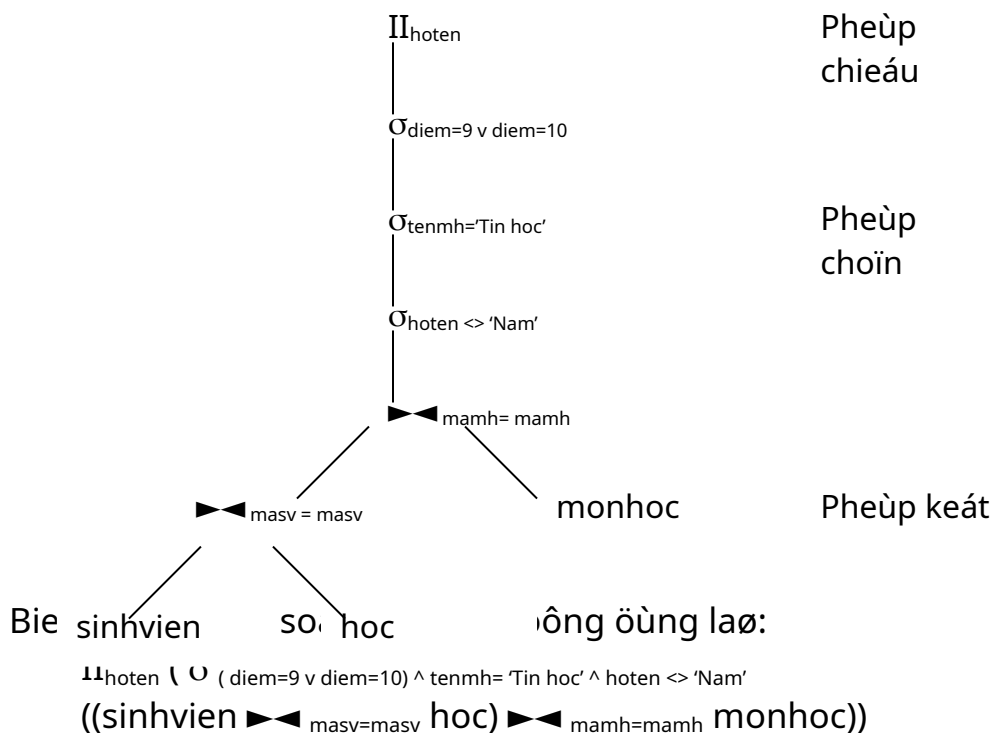
- (1) Caùc nuùt laø ñöôïc taïo laäp töø caùc quan heä ghi trong meänh ñeå From
- (2) Nuùt goác ñöôïc taïo laäp baèng pheùp chieáu treân caùc thuoäc tính ghi trong meänh ñeå SELECT.
- (3) Ñieàu kieän ghi trong meänh ñeå WHERE ñöôïc bieán ñoái thaønh moät chuoãi thích hôïp caùc pheùp toaùn ñaïi soá quan heä (pheùp choïn, pheùp keát, pheùp hôïp...) ñi töø caùc nuùt laø ñeán nuùt goác. Chuoãi caùc pheùp toaùn naøy coù theå ñöôïc cho tröïc tieáp böôù thòu töï cuûa caùc vò töø ñôn giaûn vaø caùc pheùp toaùn luaän lyù.

Moät caây toaùn töù töông öùng vôùi moät bieáu thöùc ñaïi soá quan heä.

Ví dụ: Xét truy vấn Q12 cho biết họ tên của các sinh viên không phải là 'Nam' hoặc môn học 'Tin học' năm 9 hoặc 10.

```
Q12: SELECT hoten
      FROM sinhvien, hoc, monhoc
      WHERE sinhvien.masv= hoc.masv
            AND hoc.mamh= monhoc.mamh
            AND hoten <> 'Nam'
            AND tenmh= 'Tin hoc'
            AND (diem= 9 OR diem = 10);
```

Truy vấn này có thể được biểu diễn thành một cây toàn bộ, các vòng tròn bên dưới biểu diễn các điều kiện theo thứ tự xuất hiện tổng cộng với các phép kết nối nên các phép chọn.



#### **Böôùc 1.4.2. Ñôn giaûn hoà caây toàùn töû**

Ñôn giaûn hoà caây toàùn töû nhaèm muïc ních ñeã ñaët hieäu quaû (loaïi boû caùc pheùp toàùn dö thöøa treân caùc quan heä, giaûm vuøng nhöù trung gian, giaûm thôøi gian xöû lý truy vaán) baèng caùch söû döùng caùc pheùp bieán ñoái töông ñöông cuûa caùc pheùp toàùn ñaïi soá quan heä.

Trong böôùc ñôn giaûn hoà caây toàùn töû, *moät ñieàu quan troïng trong vieäc aùp döùng caùc pheùp bieán ñoái töông ñöông cho moät bieäu thöùc truy vaán laø vieäc phaùt hieän caùc bieäu thöùc con chung (common*

*subexpression*) có trong biểu thức truy vấn, nghĩa là các biểu thức con xuất hiện nhiều lần trong biểu thức truy vấn. Nhiều máy có ý nghĩa là tiết kiệm thời gian thực hiện truy vấn vì các biểu thức con nào chưa được tính toán thì chỉ duy nhất một lần. Một phương pháp để nhận biết chuỗi các ô cho việc biến đổi cây toàn bộ tổng cộng thành một mô hình đơn giản bằng cách trích xuất tiên quyết các nút là giao nhau của cây (nghĩa là các quan hệ giao nhau), và sau đó ghép các nút trung gian khác của cây tổng cộng vào cùng các phép toán và có cùng các toán hạng.

Khi các biểu thức con nào được xác định, chúng ta có thể dùng các phép biến đổi tổng cộng sau đây để biến đổi cây toàn bộ:

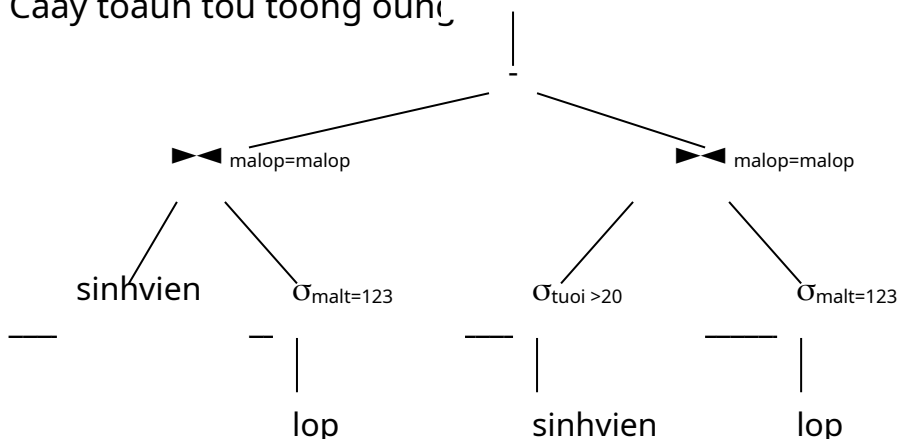
- (1)  $R \bowtie R \equiv R$
- (2)  $R \cup R \equiv R$
- (3)  $R - R \equiv \emptyset$
- (4)  $R \bowtie \sigma_F(R) \equiv \sigma_F(R)$
- (5)  $R \cup \sigma_F(R) \equiv R$
- (6)  $R - \sigma_F(R) \equiv \sigma_{\neg F}(R)$
- (7)  $\sigma_{F_1}(R) \bowtie \sigma_{F_2}(R) \equiv \sigma_{F_1 \wedge F_2}(R)$
- (8)  $\sigma_{F_1}(R) \cup \sigma_{F_2}(R) \equiv \sigma_{F_1 \vee F_2}(R)$
- (9)  $\sigma_{F_1}(R) - \sigma_{F_2}(R) \equiv \sigma_{F_1 \wedge \neg F_2}(R)$
- (10)  $R \cap R \equiv R$
- (11)  $R \cap \sigma_F(R) \equiv \sigma_F(R)$
- (12)  $\sigma_{F_1}(R) \cap \sigma_{F_2}(R) \equiv \sigma_{F_1 \wedge F_2}(R)$
- (13)  $\sigma_F(R) - R \equiv \emptyset$

Ý nghĩa của các phép biến đổi này là loại bỏ các phép toán thừa.

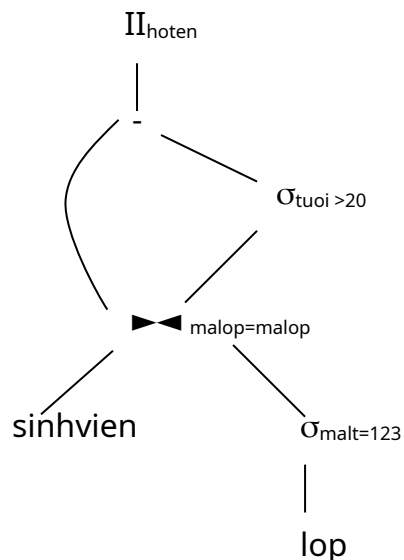
Ví dụ: Xét truy vấn Q13 cho biết các học sinh của các sinh viên thuộc lớp có mã lớp tương ứng là 123 và các sinh viên này có tuổi không lớn hơn 20 tuổi. Một biểu thức cho truy vấn này là:

$$\Pi_{\text{hoten}}((\text{sinhvien} \bowtie_{\text{malop=malop}} \sigma_{\text{malt}=123}(\text{lop})) - (\sigma_{\text{tuoi} > 20}(\text{sinhvien} \bowtie_{\text{malop=malop}} \sigma_{\text{malt}=123}(\text{lop})))$$

Cây toàn bộ tổng cộng



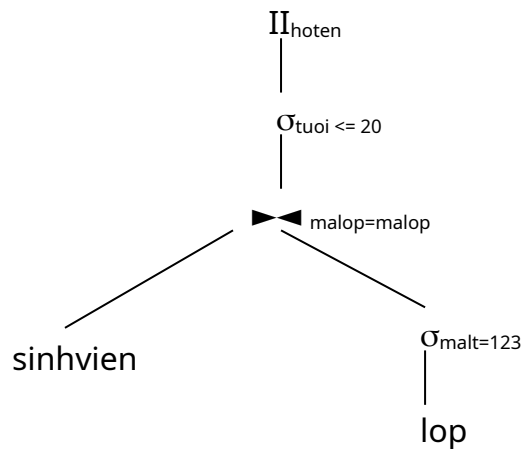
Nếu phát hiện ra biểu thức con chung, chúng ta biết rằng bằng cách ghép các nút là tổng cộng với các quan hệ sinh viên và lớp. Sau đó chúng ta sẽ thực hiện phép chọn trên tuổi hoặc với phép kết (trong cách làm này, chúng ta di chuyển phép chọn lên phía trên). Bây giờ chúng ta có thể triển khai các nút tổng cộng với phép chọn trên malt và cuối cùng các nút tổng cộng với phép kết, chúng ta sẽ có cây toàn bộ sau:



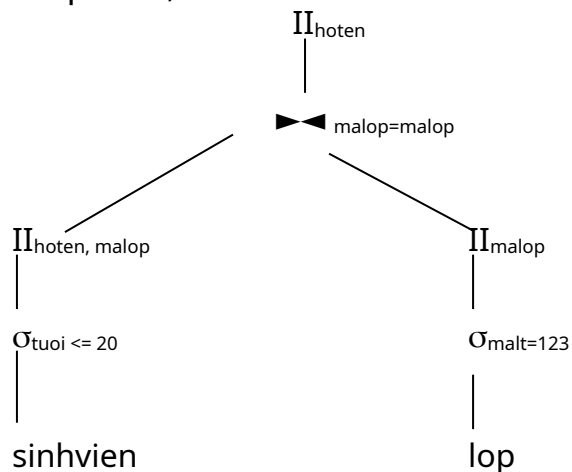
Áp dụng phép biến đổi tổng cộng (6) với R là biểu thức :

$\text{sinhvien} \bowtie_{\text{malop=malop}} \sigma_{\text{malt}=123} \text{lop}$

chúng ta thiết kế cây toàn cục sau:



Sau đó áp dụng tính phân phối của phép chiếu và phép chọn vào cây kết quả, ta thiết kế cây toàn cục:



Và biểu thức đại số quan hệ sau khi đã vận dụng:

$$II_{hoten}(II_{hoten, malop}(\sigma_{tuoi \leq 20}(sinhvien)) \bowtie_{malop=malop} II_{malop}(\sigma_{malt=123}(lop)))$$

Vận dụng một biểu thức đại số quan hệ để thể hiện rõ ràng các tiêu chuẩn sau đây:

**Tiêu chuẩn 1.** Dạng tính idempotence (tính lũy thừa) của phép chọn và phép chiếu nên ta có các phép chọn và phép chiếu thích hợp cho mọi quan hệ toàn cục.

---

**Tieâu chuaån 2.** Thöïc hieän caùc pheùp choïn vaø caùc pheùp chieáu caøng sôùm caøng toát, töùc laø ñaây caùc pheùp choïn vaø caùc pheùp chieáu xuoáng phía döôùi caây caøng xa caøng toát.

**Tieâu chuaån 3.** Khi caùc pheùp choïn ñöôïc thöïc hieän sau moät pheùp tích thì keát hôïp caùc pheùp toaùn naøy ñeå taïo thaønh moät pheùp keát.

**Tieâu chuaån 4.** Keát hôïp chuoãi caùc pheùp toaùn moät ngoâi lieân tieáp nhau aùp duïng cho moät quan heä toaùn haïng. Moät chuoãi caùc pheùp choïn lieân tieáp nhau (hoaëc moät chuoãi caùc pheùp lieân keát lieân tieáp nhau) coù theå ñöôïc keát hôïp thaønh moät pheùp choïn (hoaëc moät pheùp keát).

**Tieâu chuaån 5.** Khi phaùt hieän caùc bieäu thöùc con chung trong bieäu thöùc truy vaán, aùp duïng caùc pheùp bieán ñoái töông ñöông ñeå ñôn giaûn hoùa bieäu thöùc truy vaán.

#### **4.3.1.5. Moät giaûi thuaät toái ous hoùa moät bieäu thöùc ñaïi soá quan heä treân löôïc ñoà toaøn cuïc**

**Vaøo:** Moät bieäu thöùc ñaïi soá quan heä treân löôïc ñoà toaøn cuïc

**Ra:** Moät bieäu thöùc ñaïi soá quan heä ñaõ ñöôïc toái ous hoùa

Giaûi thuaät toái ous hoùa moät bieäu thöùc ñaïi soá quan heä treân löôïc ñoà toaøn cuïc bao goàm caùc böôùc sau ñaây:

**Böôùc 1.** Phaùt hieän caùc bieäu thöùc con chung coù trong caây toaùn töù, bieán ñoái caây toaùn töù döïa treân bieäu thöùc con chung

**Böôùc 2.** Thöïc hieän pheùp choïn caøng sôùm caøng toát. Sôu duïng tính idempotence cuûa pheùp choïn, tính giao hoùa cuûa pheùp choïn vôùi pheùp chieáu, vaø tính phaân phóái cuûa pheùp choïn ñoái vôùi pheùp hôïp, pheùp giao, pheùp hieäu, pheùp keát vaø pheùp tích ñeå di chuyeån pheùp choïn caøng xuoáng phía döôùi caây caøng toát.

Sôu duïng caùc pheùp bieán ñoái töông ñöông:

$$\sigma_{F_1}(\sigma_{F_2}(R)) \equiv \sigma_{F_2}(\sigma_{F_1}(R))$$

$$\sigma_{F_1}(\sigma_{F_2}(R)) \equiv \sigma_{F_1 \wedge F_2}(R)$$

$$\Pi_X(\sigma_F(R)) \leftrightarrow \sigma_F(\Pi_X(R))$$

$$(\rightarrow \text{nếu Attr}(F) \subseteq X)$$

$$\Pi_X(\sigma_F(R)) \equiv \Pi_X(\sigma_F(\Pi_{X \cup \text{Attr}(F)}(R)))$$

$$\sigma_F(R \cup S) \equiv \sigma_F(R) \cup \sigma_F(S)$$

$$\sigma_F(R \cap S) \equiv \sigma_F(R) \cap \sigma_F(S)$$

$$\sigma_{F_1 \wedge F_2}(R \cap S) \leftrightarrow \sigma_{F_1}(R) \cap \sigma_{F_2}(S)$$

$$(\rightarrow \text{nếu Attr}(F_1) \subseteq \text{Attr}(R) \text{ và } \text{Attr}(F_2) \subseteq \text{Attr}(S))$$

$$\sigma_F(R - S) \equiv \sigma_F(R) - \sigma_F(S)$$

$$\sigma_F(R \bowtie_{F_1} S) \leftrightarrow \sigma_F(R) \bowtie_{F_1} S$$

$$(\rightarrow \text{nếu Attr}(F) \subseteq \text{Attr}(R))$$

$$\sigma_{F_1 \wedge F_2}(R \bowtie_{F_3} S) \leftrightarrow \sigma_{F_1}(R) \bowtie_{F_3} \sigma_{F_2}(S)$$

$$(\rightarrow \text{nếu Attr}(F_1) \subseteq \text{Attr}(R) \text{ và } \text{Attr}(F_2) \subseteq \text{Attr}(S))$$

$$\sigma_F(R \bowtie_{F_3} S) \leftrightarrow \sigma_{F_2}(\sigma_{F_1}(R) \bowtie_{F_3} S)$$

$$(\rightarrow \text{nếu } F=F_1 \wedge F_2 \text{ và } \text{Attr}(F_1) \subseteq \text{Attr}(R) \text{ và } \text{Attr}(F_2) \subseteq \text{Attr}(R) \cup$$

Attr(S))

$$\sigma_F(R \times S) \leftrightarrow \sigma_F(R) \times S$$

$$(\rightarrow \text{nếu Attr}(F) \subseteq \text{Attr}(R))$$

$$\sigma_{F_1 \wedge F_2}(R \times S) \leftrightarrow \sigma_{F_1}(R) \times \sigma_{F_2}(S)$$

$$(\rightarrow \text{nếu Attr}(F_1) \subseteq \text{Attr}(R) \text{ và } \text{Attr}(F_2) \subseteq \text{Attr}(S))$$

$$\sigma_F(R \times S) \leftrightarrow \sigma_{F_2}(\sigma_{F_1}(R) \times S)$$

$$(\rightarrow \text{nếu } F=F_1 \wedge F_2 \text{ và } \text{Attr}(F_1) \subseteq \text{Attr}(R) \text{ và } \text{Attr}(F_2) \subseteq \text{Attr}(R) \cup$$

Attr(S))

**Bổ đề 3.** Thử kiểm tra tính phân phối của phép chiếu, tính idempotence của phép chiếu, tính phân phối của phép chiếu đối với phép nối, phép kết và phép tích nếu di chuyển phép chiếu sang trước hoặc sau phép nối, phép kết và phép tích. Kiểm tra tất cả các phép chiếu là cần thiết, loại bỏ phép chiếu không cần thiết nếu phép nối phép chiếu trên tất cả các thuộc tính của quan hệ toàn bộ.

Sử dụng phép biến đổi:

$$\Pi_{X_1}(\Pi_{X_2}(R)) \equiv \Pi_{X_1}(R) \quad \text{với } X_1 \subseteq X_2$$

$$\Pi_X(R \cup S) \equiv \Pi_X(R) \cup \Pi_X(S)$$

$$\Pi_X(R \bowtie_F S) \leftrightarrow \Pi_X(R) \bowtie_F (S)$$

$$\begin{aligned}
 & (\rightarrow \text{nếu Attr}(F_R) \subseteq X \text{ và } X \subseteq \text{Attr}(R)) \\
 \Pi_{X_1 \cup X_2}(R \bowtie_F S) & \leftrightarrow \Pi_{X_1}(R) \bowtie_F \Pi_{X_2}(S) \\
 & (\rightarrow \text{nếu Attr}(F) \subseteq X_1 \cup X_2 \text{ và } X_1 \subseteq \text{Attr}(R) \text{ và } X_2 \subseteq \text{Attr}(S)) \\
 \Pi_{X_1 \cup X_2}(R \times S) & \leftrightarrow \Pi_{X_1}(R) \times \Pi_{X_2}(S) \\
 & (\rightarrow \text{nếu } X_1 \subseteq \text{Attr}(R) \text{ và } X_2 \subseteq \text{Attr}(S))
 \end{aligned}$$

**Bổ đề 4.** Nếu một phép chọn nào đó thực hiện ngay sau một phép tích, mà phép chọn bao gồm các thuộc tính của các quan hệ trong phép tích, thì biến đổi phép tích thành phép kết. Nếu phép chọn chỉ bao gồm các thuộc tính của một quan hệ trong phép tích, thì thực hiện phép chọn cho quan hệ này trước khi thực hiện phép tích.

Sơ đồ thực hiện phép biến đổi:

$$\begin{aligned}
 \sigma_F(R \times S) & \leftrightarrow \sigma_F(R) \times S \\
 & (\rightarrow \text{nếu Attr}(F) \subseteq \text{Attr}(R)) \\
 \sigma_{F_1 \wedge F_2}(R \times S) & \leftrightarrow \sigma_{F_1}(R) \times \sigma_{F_2}(S) \\
 & (\rightarrow \text{nếu Attr}(F_1) \subseteq \text{Attr}(R) \text{ và } \text{Attr}(F_2) \subseteq \text{Attr}(S)) \\
 \sigma_F(R \times S) & \leftrightarrow \sigma_{F_2}(\sigma_{F_1}(R) \times S) \\
 & (\rightarrow \text{nếu } F=F_1 \wedge F_2 \text{ và } \text{Attr}(F_1) \subseteq \text{Attr}(R) \text{ và } \text{Attr}(F_2) \subseteq \text{Attr}(R) \cup \text{Attr}(S))
 \end{aligned}$$

**Bổ đề 5.** Nếu có một chuỗi các phép chọn và/ hoặc các phép chiếu, sơ đồ thực hiện giao hoán hoặc tính idempotence nếu kết hợp chúng thành một phép chọn, một phép chiếu hoặc một phép chọn rồi trước một phép chiếu và sắp xếp chúng cho mỗi bộ của quan hệ toàn hàng. Nếu một phép kết hoặc phép tích rồi trước một chuỗi các phép chọn hoặc các phép chiếu, thì sắp xếp chúng cho mỗi bộ của phép kết hoặc phép chiếu ngay khi tạo ra kết quả.

**Bổ đề 6.** Sơ đồ thực hiện kết hợp của phép giao, phép tích và phép kết nếu sắp xếp lại các quan hệ trong cây toàn bộ, sao cho phép toàn bộ nào mà nó tạo ra kết quả ít nhất sẽ nào đó thực hiện trước tiên.

Sơ đồ thực hiện phép biến đổi:



$$(R \cap S) \cap T \equiv (R \cap T) \cap S$$

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

$$(R \bowtie_{F1} S) \bowtie_{F2} T \leftrightarrow (R \bowtie_{F2} T) \bowtie_{F1} S$$

$$(\rightarrow \text{ nếu } \text{Attr}(F2) \subseteq \text{Attr}(R) \cup \text{Attr}(T))$$

$$(\leftarrow \text{ nếu } \text{Attr}(F1) \subseteq \text{Attr}(R) \cup \text{Attr}(S))$$

### 4.3.2. Bôôùc 2 – Ñònh vò döô lieäu

Bôôùc ñònh vò döô lieäu (Data Localization) coøn ñöôïc goïi laø bôôùc **toái öu hoùa truy vaán trên lược đồ phân miền**. Bôôùc naøy bieán ñoài truy vaán toaøn cuïc (keát quaû cuûa Bôôùc 1) thaønh caùc truy vaán miền hiệu quả: *loại bỏ các phép toán unnecessary không cần thiết trên các miền và giảm vòng nhôu trung gian*. Toán tử hoán truy vấn trên lược đồ phân miền bao gồm 2 bước sau:

**Bôôùc 2.1.** Bieán ñoài biểu thức ñoài số quan hệ trên lược đồ toaøn cuïc (chöùa caùc quan hệ toaøn cuïc) thaønh biểu thức ñoài số quan hệ trên lược đồ phân miền (chöùa caùc miền của quan hệ toaøn cuïc) bằng cách thay thế các quan hệ toaøn cuïc bởi biểu thức tài liệu của chúng.

**Bôôùc 2.2.** Nôn giảm hóa biểu thức ñoài số quan hệ trên lược đồ phân miền để có ñöôïc một biểu thức hiệu quả (loại bỏ các phép toán không cần thiết giảm vòng nhôu trung gian) bằng cách sử dụng các phép biến đổi tổng cộng của ñoài số quan hệ và các ñoài số quan hệ ñöôïc tuyeån chọn.

#### 4.3.2.1. Böôùc 2.1 – Bieán ñoái bieáu thöùc ñaïi soá quan heä treân löôïc ñoà toaøn cuïc

Böôùc naøy seõ bieán ñoái bieáu thöùc ñaïi soá quan heä treân löôïc ñoà toaøn cuïc (chöùa caùc quan heä toaøn cucĩ) thaønh bieáu thöùc ñaïi soá quan heä treân löôïc ñoà phaân maûnh (chöùa caùc maûnh cuùa quan heä toaøn cucĩ) baèng cach thay theá moãi quan heä toaøn cuïc trong caây toaùn töû böôûi bieáu thöùc taùi laäp cuùa noù. Bieáu thöùc taùi laäp cuùa moät quan heä toaøn cuïc laø moät bieáu thöùc ñaïi soá quan heä bao goàm caùc maûnh cuùa quan heä naøy maø bieáu thöùc naøy cho pheùp taïo laïi quan heä toaøn cuïc naøy. Bieáu thöùc taùi laäp cuõng ñöôïc bieáu dieãn baèng moät caây toaùn töû.

Xeùt löôïc ñoà quan heä sinhvien vaø lop sau ñaây:

Sinhvien (masv, hoten, tuoi, malop)

Lop (malop, tenlop, malt, tenkhoa)

Giaû söû chuùng ta coù hai khoa teân laø 'CNTT' vaø 'DIEN'. Quan heä *lop* ñöôïc phaân maûnh ngang döïa vaøo *tenkhoa* thaønh hai maûnh *lop1* vaø *lop2*. Quan heä *sinhvien* ñöôïc phaân maûnh ngang suy daãn theo *lop* döïa vaøo *malop* thaønh hai maûnh *sinhvien1* vaø *sinhvien2*. Löôïc ñoà phaân maûnh nhö sau:

Lop1 (malop, tenlop, malt, tenkhoa)

Lop2 (malop, tenlop, malt, tenkhoa)

Sinhvien1 (masv, hoten, tuoi, malop)

Sinhvien2 (masv, hoten, tuoi, malop)

Caùc bieáu thöùc taùi laäp cuùa quan heä *lop* vaø *sinhvien* laø:

Lop = Lop1 U Lop2

Sinhvien = sinhvien1 U sinhvien2

Trong ñoù:

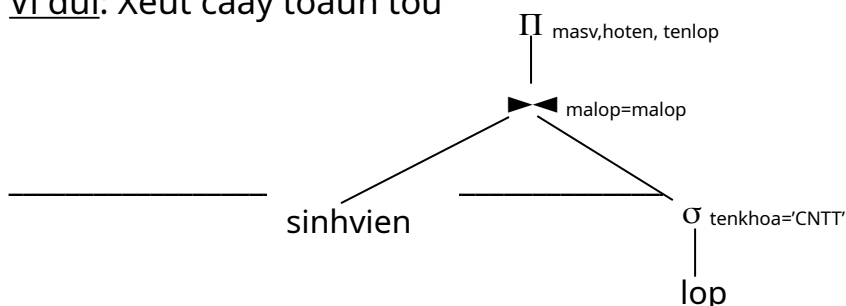
Lop1 =  $\sigma_{tenkhoa = 'CNTT'}(lop)$

Lop2 =  $\sigma_{tenkhoa = 'DIEN'}(lop)$

Sinhvien1 = sinhvien  $\bowtie$  (Lop1)

Sinhvien2 = sinhvien  $\bowtie$  (Lop2)

Ví dụi: Xeùt caây toaùn töû







$$(3) R \times \emptyset \equiv \emptyset$$

$$(4) R \cup \emptyset \equiv R$$

$$(5) R \cap \emptyset \equiv \emptyset$$

$$(6) R - \emptyset \equiv R$$

$$(7) \emptyset - R \equiv \emptyset$$

$$(8) R \blacktriangleright \blacktriangleleft \emptyset \equiv \emptyset$$

$$(9) R \blacktriangleright < \emptyset \equiv \emptyset$$

$$(10) \emptyset \blacktriangleright < R \equiv \emptyset$$

Ñôn giaûn hoùa möät bieåu thöüc ñaïi soá quan heä treân löôic ñoà phaân maûnh ñöôic thöïc hieän döïa treân caùc tieâu chuaån sau:

**Tieâu chuaån 6:** Di chuyeån caùc pheùp choïn xuoáng caùc nuùt laù cuûa caây, vaø sau ñoù aùp duïng chuùng baèng caùch duøng ñaïi soá quan heä ñöôic tuyeån choïn; thay theá caùc keát quaû choïn löïa böûi quan heä roãng neáu ñieàu kieän choïn cuûa keát quaû bò maâu thuaãn.

**Tieâu chuaån 7:** Ñeå phaân phóái caùc pheùp keát xuaát hieän trong möät truy vaán toaøn cuïc, caùc pheùp hôïp (bieåu dieãn taäp hôïp cuûa caùc phaân maûnh) phaûi ñöôic di chuyeån leân phía treân caùc pheùp keát maø chuùng ta muoán phaân phóái ñeå loaïi böû caùc pheùp keát khoâng caàn thieát.

**Tieâu chuaån 8:** Duøng ñaïi soá quan heä ñöôic tuyeån choïn ñeå ñònh trò ñieàu kieän choïn cuûa caùc toaøn haïng cuûa caùc pheùp keát; thay theá caây con, bao goàm pheùp keát vaø caùc toaøn haïng cuûa noù, baèng quan heä roãng neáu ñieàu kieän choïn cuûa keát quaû cuûa pheùp keát bò maâu thuaãn.

Ví duï : Xeùt caây toaøn töû treân löôic ñoà phaân maûnh treân  
Ñaây pheùp choïn vaø pheùp chieáu xuoáng khoûi pheùp hôïp ta ñöôic:

$$\begin{aligned} & \Pi_{\text{malop,tenlop}}(\sigma_{\text{tenkhoa}='CNTT'}(\text{lop1} \cup \text{lop2})) \\ &= \Pi_{\text{malop,tenlop}}(\sigma_{\text{tenkhoa}='CNTT'}(\text{lop1})) \cup \Pi_{\text{malop,tenlop}}(\sigma_{\text{tenkhoa}='CNTT'}(\text{lop2})) \end{aligned}$$

Ta nhaän thaáy keát quaû cuûa pheùp choïn  $\sigma_{\text{tenkhoa}='CNTT'}(\text{lop2})$  laø roãng vaø pheùp choïn  $\sigma_{\text{tenkhoa}='CNTT'}(\text{lop1})$  laø khoâng caàn thieát vì ñieàu kieän choïn cuûa lop1 laø tenkhoa= 'CNTT'.

Do ñoù:

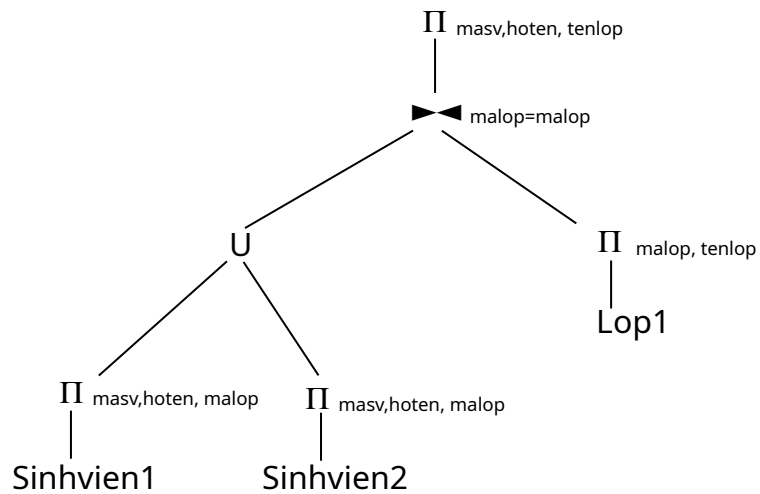
$$\Pi_{\text{malop,tenlop}}(\sigma_{\text{tenkhoa}='CNTT'}(\text{lop1} \cup \text{lop2})) = \Pi_{\text{malop,tenlop}}(\text{lop1})$$

Nhảy phép chiếu xuống khỏi phép join trong biểu thức:

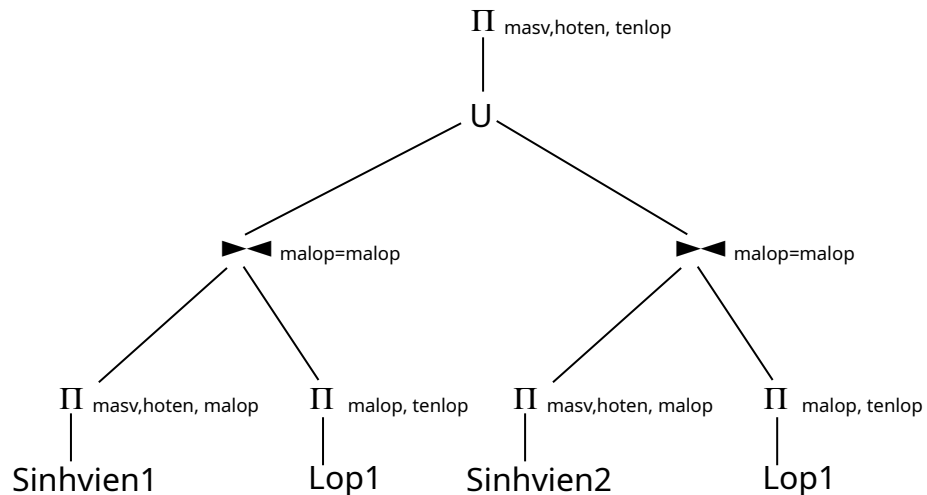
$$\Pi_{\text{masv,hoten,malop}}(\text{sinhvien1} \cup \text{sinhvien2}) =$$

$$\Pi_{\text{masv,hoten,malop}}(\text{sinhvien1}) \cup \Pi_{\text{masv,hoten,malop}}(\text{sinhvien2})$$

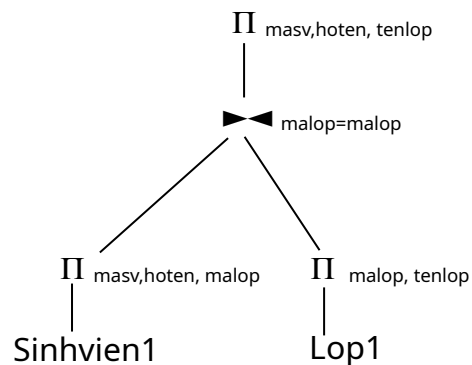
Ta có cây toàn bộ:



Sau đó phân phối phép kết nối với phép hợp ta có:



Tuy nhiên phép kết nối giữa **sinhvien2** và **lop1** là rỗng do nhiều kiến thức của phân môn **lop1** và **sinhvien2** mâu thuẫn nhau. Cuối cùng ta có cây toàn bộ trên lược đồ phân môn như sau:



Nên giao hoán biểu thức nào đó quan hệ trong logic nào phân minh còn dữ liệu vào một hệ suy diễn nào đó gọi là Bộ chứng minh hình lý (Theorem Prover).

Ví dụ: Giả sử chúng ta cần có hai khoa là 'CNTT', 'DIEN' và có toán tử 20 lớp, các lớp có mã lớp từ 1 đến 10 thuộc khoa 'CNTT' và các lớp có mã từ 11 đến 20 thuộc khoa 'DIEN'. Từ đó, chúng ta có các luật suy diễn sau:

Malop > 10 →	tenkhoa = 'DIEN'
Malop ≤ 10 →	NOT (Malop > 10)
Malop > 10 →	NOT (Malop ≤ 10)
tenkhoa = 'CNTT' →	Malop ≤ 10
tenkhoa = 'DIEN' →	Malop > 10
tenkhoa = 'CNTT' →	not(tenkhoa = 'DIEN')
tenkhoa = 'DIEN' →	not(tenkhoa = 'CNTT')

Xét truy vấn Q14 cho biết tên lớp của lớp có mã lớp bằng 1:

Q14: Select tenlop  
From lop  
Where malop = 1

Trước khi thực hiện truy vấn này, chúng ta cần các suy diễn sau đây:

Malop = 1 →	malop ≤ 10
Malop ≤ 10 →	tenkhoa = 'CNTT'

Do đó truy vấn này cần liên quan đến lớp 1 vì nhiều kiến thức của lớp 1 là tenkhoa = 'CNTT'. Vì thế biểu thức nào đó quan hệ của truy vấn này là:

$\Pi_{tenlop}(\sigma_{malop=1}(lop1))$

### 4.3.3 Bổ đề 3 Toán tử toán truy vấn toán tử

Bổ đề toán tử toán truy vấn toán tử nhằm để tìm ra một chiến lược thực hiện truy vấn sao cho chiến lược này gần tối ưu (theo nghĩa giảm thời gian thực hiện truy vấn trên dữ liệu nào đó phân tán, giảm vãng nhữ trung gian).

Một chiến lược nào đó cần trông bề ngoài thì tối thực hiện các phép toán nào đó quan hệ và các tài nguyên thông tin của



(gôûi/nhaän) duøng ñeã truyeàn döõ lieäu giöõa caùc vò trí. Baèng caùc hoàùn ñoãi thöù töï cuûa caùc pheùp toàùn trong bieäu thöùc truy vaán phaân maûnh, ta coù theã coù ñöôïc nhieàu truy vaán töông ñöông.

Toái ous hoùa truy vaán toaøn cuïc laø tìm ra moät thöù töï thöïc hieän caùc pheùp toaùn trong bieäu thöcù truy vaán sao cho ít toán thôøi gian nhaát. Ñeëc bieät khaâu toáùn keøm thôøi gian trong cô sôu döõ lieäu phaân taùn laø khaâu truyeàn döõ lieäu do toác ñoã vaø baêng thoâng giöùì haïn.

Trong tröôøng hôïp nhaân baûn thì coøn phaûi tính xem nhaân baûn naøo ñöôïc sôu duïng nhaèm giaûm chi phí truyeàn thoâng.

Moät khía caïnh quan troïng cuûa toái ous hoùa truy vaán laø *thöù töï thöïc hieän caùc pheùp keát phaân taùn*. Nhöø tính giao hoàùn cuûa caùc pheùp keát, chuùng ta coù theã laøm giaûm chi phí thöïc hieän caùc pheùp keát naøy. Moät kyõ thuaät cô baûn ñeã toái ous hoùa moät chuoãi caùc pheùp keát phaân taùn laø sôu duïng pheùp nöûa keát nhaèm laøm giaûm chi phí truyeàn thoâng giöõa caùc vò trí vaø taêng tính xöû lý cuïc boã taïi caùc vò trí.

$$R \rhd_{A=B} S = S \rhd_{A=B} (R \rhd_{A=B} \Pi_B S)$$

Ví duï: Giaû sö coù söï phaân taùn döõ lieäu sau:

- maûnh *sinhvien1* ñaët taïi vò trí 1 vaø
- maûnh *lop1* ñaët taïi vò trí 2

Chuùng ta caàn thöïc hieän pheùp keát phaân taùn sau:

$$\text{Sinhvien1} \rhd \text{lop1}$$

Baèng caùch aùp duïng pheùp nöûa keát bieäu thöùc treân töông ñöông vöùì:

$$\text{Lop1} \rhd (\text{sinhvien1} \rhd \Pi_{\text{malop}}(\text{lop1}))$$

Do ñoù ta coù moät chieán löôïc thöïc hieän cho pheùp keát phaân taùn naøy vöùì caùc taùc vụ truyeàn thoâng sau:

- 1) Thöïc hieän  $T_1 = \Pi_{\text{malop}}(\text{lop1})$  cuïc boã taïi vò trí 2.

- 2) Truyeàn  $T_1$  töø vò trí 2 qua vò trí 1.
- 3) Thöïc hieän  $T_2 = \text{sinhvien1} \triangleright < T_1$  cuïc boä taïi vò trí 1.
- 4) Truyeàn  $T_2$  töø vò trí 1 qua vò trí 2.
- 5) Thöïc hieän  $T_3 = \text{lop1} \triangleright < T_2$  cuïc boä taïi vò trí 2.
- 6) Truyeàn  $T_3$  töø vò trí 2 qua vò trí cuûa öùng düïng caàn thöïc hieän cuûa pheùp keát naøy.

#### 4.3.4 Böôùc 4 Toái öu hoà trùy vaán cuïc boä

Toái öu hoà trùy vaán cuïc boä nhaèm ñeå thöïc hieän caùc trùy vaán con ñöôïc phaân taùn taïi moãi vò trí, goïi laø trùy vaán cuïc boä coù chöùa caùc maûnh, sau ñoù ñöôïc toái öu hoà treân löôïc ñòa cuïc boä taïi moãi vò trí. Toái öu hoà trùy vaán cuïc boä sôu düïng caùc thuaät toàùn toái öu hoà trùy vaán cuûa cô sôu döõ lieäu taäp trung.