Chương 4

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

- Hầu hết các truy vấn *SQL* đều có thể được biến đổi thành các biểu thức đại số quan hệ tương đương và ngược lại.
- Một biểu thức đại số quan hệ (*expression of relational algebra*) không chỉ là sự đặc tả ngữ nghĩa của một truy vấn, mà còn là sự đặc tả của một *chuỗi các phép toán* (*sequence of operations*).
- Hai biểu thức có cùng ngữ nghĩa có thể mô tả hai chuỗi phép toán khác nhau.

$\Pi_{\text{họtên, mãbm}}\ \sigma_{\text{mãbm = 15}}$ (giảngviên)

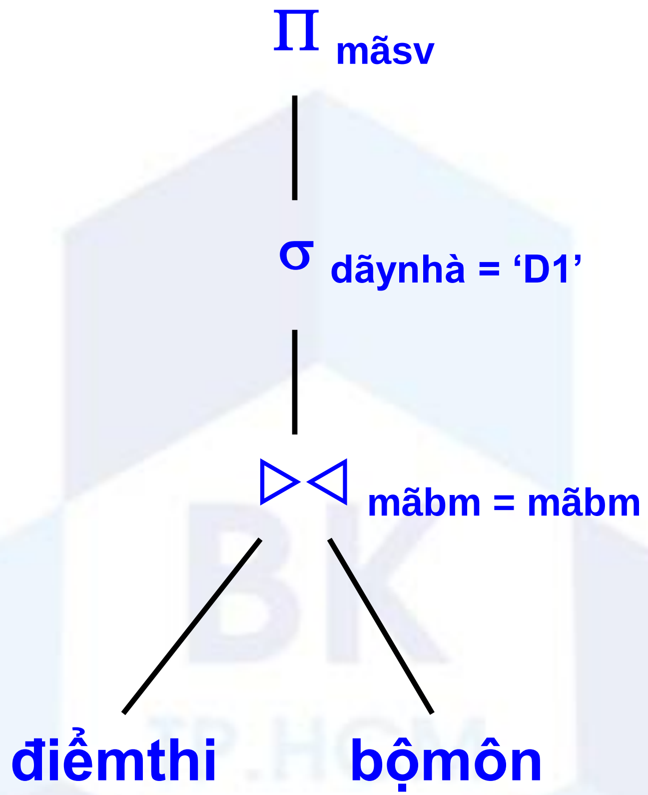
$\sigma_{\text{mãbm = 15}}\ \Pi_{\text{họtên, mãbm}}$ (giảngviên)

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 sinh viên có học môn học thuộc các bộ môn ở dãy nhà D1.

$$Q_1: \Pi_{\text{mãsv}}\ \sigma_{\text{dãynhà = 'D1'}}\ (\text{điểmthi} \bowtie_{\text{mãbm = mãbm}} \text{bộmôn})$$

Cây toán tử của truy vấn



Hình 4.1. Cây toán tử của truy vấn Q_1

Cây toán tử của truy vấn

❖ Biểu diễn cây toán tử

- Các nút lá là các quan hệ toàn cục.
- Mỗi nút trung gian là phép toán một ngôi (*unary operation*) hoặc phép toán hai ngôi (*binary operation*).
- Nút gốc là phép toán tạo ra kết quả cuối cùng của truy vấn.
- Thứ tự thực hiện của cây toán tử là phép duyệt cây *LRN*.

Các phép biến đổi tương đương

- **Định nghĩa của quan hệ không phụ thuộc vào thứ tự của các thuộc tính trong lược đồ quan hệ.**
- ***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

- Gọi *U* là phép toán đại số quan hệ một ngôi và *B* là phép toán đại số quan hệ hai ngôi.
- 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:**

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

Các phép biến đổi tương đương

❖ Một số phép biến đổi tương đương

- **Hai phép chọn kề nhau**

$$\sigma_{F1}(\sigma_{F2}(R)) \equiv \sigma_{F2}(\sigma_{F1}(R))$$

$$\sigma_{F1}(\sigma_{F2}(R)) \equiv \sigma_{F1 \wedge F2}(R)$$

- **Hai phép chiếu kề nhau**

$$\Pi_{X1}(\Pi_{X2}(R)) \equiv \Pi_{X1}(R) \quad \text{với } X_1 \subseteq X_2$$

- **Phép chiếu kề với phép chọn**

$$\Pi_X(\sigma_F(R)) \leftrightarrow \sigma_F(\Pi_X(R)) \quad (\rightarrow \text{ nếu } Attr(F) \subseteq X)$$

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

- **Phép hợp kề với phép chọn**

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

Các phép biến đổi tương đương

❖ Một số phép biến đổi tương đương

▸ Phép kết kề với phép chọn

$\sigma_F (R \rhd\lhd_{F1} S) \leftrightarrow \sigma_F (R) \rhd\lhd_{F1} S$ $(\rightarrow$ nếu $Attr(F) \subseteq Attr(R))$

$\sigma_{F1 \wedge F2} (R \rhd\lhd_{F3} S) \leftrightarrow \sigma_{F1} (R) \rhd\lhd_{F3} \sigma_{F2} (S)$

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

$\sigma_F (R \rhd\lhd_{F3} S) \leftrightarrow \sigma_{F2} (\sigma_{F1} (R) \rhd\lhd_{F3} S)$

$(\rightarrow$ nếu $F = F_1 \wedge F_2$ và

$Attr(F_1) \subseteq Attr(R)$ và $Attr(F_2) \subseteq Attr(R) \cup Attr(S))$

▸ Phép nửa kết kề với phép chọn

$\sigma_F (R \rhd<_{F1} S) \leftrightarrow \sigma_F (R) \rhd<_{F1} S$ $(\rightarrow$ nếu $Attr(F) \subseteq Attr(R))$

Các phép biến đổi tương đương

- Một số phép biến đổi tương đương
  - Phép tích *Descartes* kề với phép chọn

$\sigma_F(R \times S) \leftrightarrow \sigma_F(R) \times S$ ($\rightarrow$ nếu Attr(F) $\subseteq$ Attr(R))

$\sigma_{F1 \wedge F2}(R \times S) \leftrightarrow \sigma_{F1}(R) \times \sigma_{F2}(S)$

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

$\sigma_F\ (R \times S) \leftrightarrow \sigma_{F2}(\sigma_{F1}(R) \times S)$

($\rightarrow$ nếu $F = F_1 \wedge F_2$ và

$Attr(F_1) \subseteq Attr(R)$ và $Attr(F_2) \subseteq Attr(R) \cup Attr(S)$)

  - Phép giao kề với phép chọn

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

$\sigma_{F1 \wedge F2}(R \cap S) \leftrightarrow \sigma_{F1}(R) \cap \sigma_{F2}(S)$

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

Các phép biến đổi tương đương

- Một số phép biến đổi tương đương
  - Phép hiệu kề với phép chọn

    $\sigma_F(R - S) \equiv \sigma_F(R) - \sigma_F(S)$
  - Phép chia kề với phép chọn

    $\sigma_F(R \div S) \leftrightarrow \sigma_F(R) \div S$ ($\rightarrow$ nếu $Attr(F) \subseteq Attr(R)$)
  - Phép hợp kề với phép chiếu

    $\Pi_X(R \cup S) \equiv \Pi_X(R) \cup \Pi_X(S)$
  - Phép kết kề với phép chiếu

    $\Pi_X(R \rhd\lhd_F S) \leftrightarrow \Pi_X(R) \rhd\lhd_F S$

($\rightarrow$ nếu thỏa hai điều kiện: 1. $Attr(F_R) \subseteq X$; 2. $X \subseteq Attr(R)$)

$\Pi_{X1 \cup X2}(R \rhd\lhd_F S) \leftrightarrow \Pi_{X1}(R) \rhd\lhd_F \Pi_{X2}(S)$

($\rightarrow$ nếu thỏa ba điều kiện:

1. $Attr(F) \subseteq X1 \cup X2$; 2. $X1 \subseteq Attr(R)$; 3. $X2 \subseteq Attr(S)$)

Các phép biến đổi tương đương

- Một số phép biến đổi tương đương
  - Phép nửa kết kề với phép chiếu

$$\Pi_{X1 \cup X2} (R \triangleright\!<_F S) \leftrightarrow \Pi_{X1} (R) \triangleright\!<_F \Pi_{X2} (S)$$

(→ nếu thỏa ba điều kiện: 1. *Attr*(*F*) ⊆ *X1* ∪ *X2*; 2. *X1* ⊆ *Attr*(*R*); 3. *X2* ⊆ *Attr*(*S*) ∩ *Attr*(*F*))

  - Phép tích *Descartes* kề với phép chiếu

$$\Pi_{X1 \cup X2} (R \times S) \leftrightarrow \Pi_{X1} (R) \times \Pi_{X2} (S)$$

(→ nếu thỏa hai điều kiện: 1. *X1* ⊆ *Attr*(*R*); 2. *X2* ⊆ *Attr*(*S*))

Các phép biến đổi tương đương

- Một số phép biến đổi tương đương
  - **Các phép hợp, phép giao, phép tích *Descartes* và phép kết**

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

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

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

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

$$(\rightarrow \text{ nếu } Attr(F_2) \subseteq Attr(R) \cup Attr(T))$$

$$(\leftarrow \text{ nếu } Attr(F_1) \subseteq Attr(R) \cup Attr(S))$$

  - **Các phép hiệu và phép chia**

$$R - S - T \equiv R - T - S$$

$$R \div S \div T \equiv R \div T \div S$$

Các phép biến đổi tương đương

- Một số phép biến đổi tương đương
  - Phép kết đi sau phép hợp

    $$(R \cup S) \bowtie_F T \equiv (R \bowtie_F T) \cup (S \bowtie_F T)$$

  - Phép giao đi sau phép hợp

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

  - Phép hiệu đi sau phép hợp

    $$(R \cup S) - T \equiv (R - T) \cup (S - T)$$

  - Phép hiệu đi trước phép hợp

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

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

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

$$(R \bowtie_F S) - T \equiv (R - T) \bowtie_F S \equiv (R \bowtie_F S) - (T \bowtie_F S)$$

Các phép biến đổi tương đương

- Một số phép biến đổi tương đương
  - Phép giao

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

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

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

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

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

  - Phép hiệu

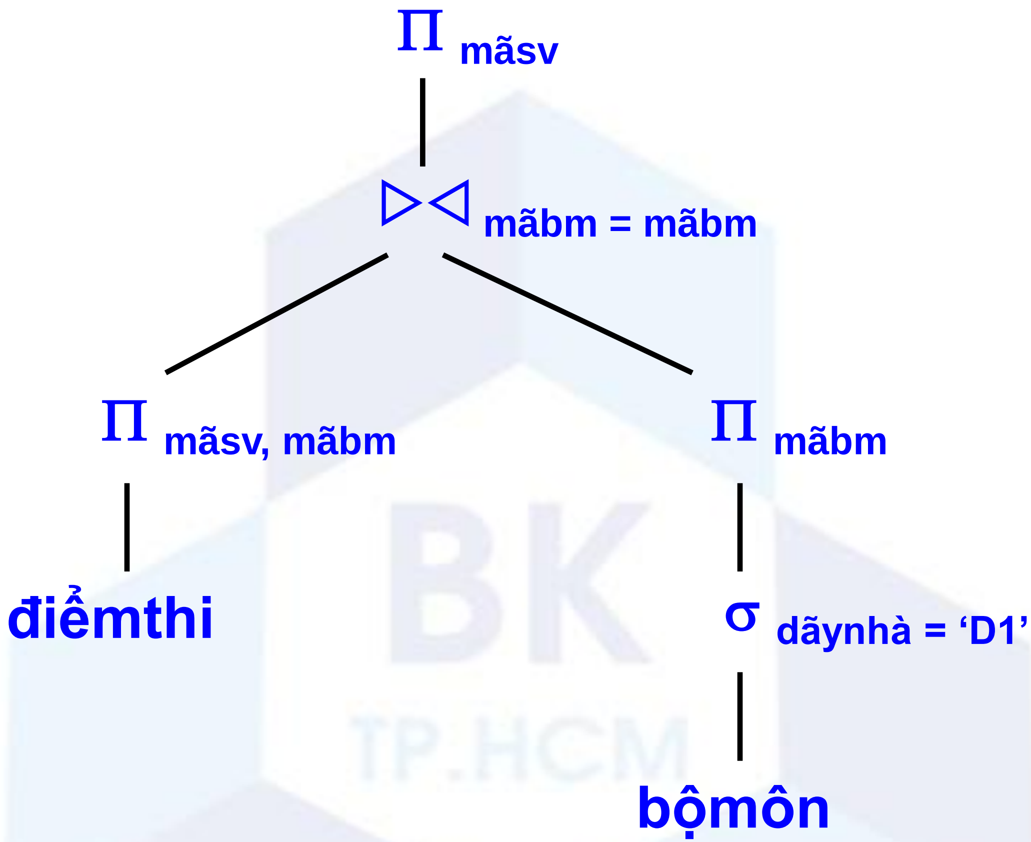
$$(R - S) \bowtie_F T \equiv (R \bowtie_F T) - (S \bowtie_F T)$$

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

Tiêu chuẩn 1 và 2

- Mục đích là đơn giản hóa việc thực hiện các truy vấn, giảm kích thước của các toán hạng của các phép toán hai ngôi trước khi thực hiện chúng, đặc biệt là các phép kết.
- **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 2



Hình 4.2. Cây toán tử của truy vấn Q_1 đã bị thay đổi.

Tiêu chuẩn 1 và 2

- Khi một phép chiếu di chuyển xuống dưới qua một phép kết thì các thuộc tính trong điều kiện kết phải có trong tập thuộc tính chiếu của phép chiếu này.

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.
- Tìm ra các biểu thức con chung sẽ tiết kiệm thời gian thực hiện của truy vấn nếu các biểu thức con chung chỉ được định trị một lần.

Đồ thị toán tử và biểu thức con chung

- Một phương pháp để nhận biết các biểu thức con chung là biến đổi cây toán tử tương ứng thành một đồ thị toán tử.
  - Trước tiên, gộp các nút lá giống nhau của cây (nghĩa là các quan hệ toán hạng giống nhau).
  - Sau đó, gộp các nút trung gian khác của cây tương ứng với cùng các phép toán và có cùng các toán hạng.

Đồ thị toán tử và biểu thức con chung

- **Ví dụ**
  - Truy vấn Q_2 – Hãy cho biết họ tên của các giảng viên thuộc bộ môn có mã sếp là 373 nhưng tiền lương của họ không lớn hơn \$35.000.

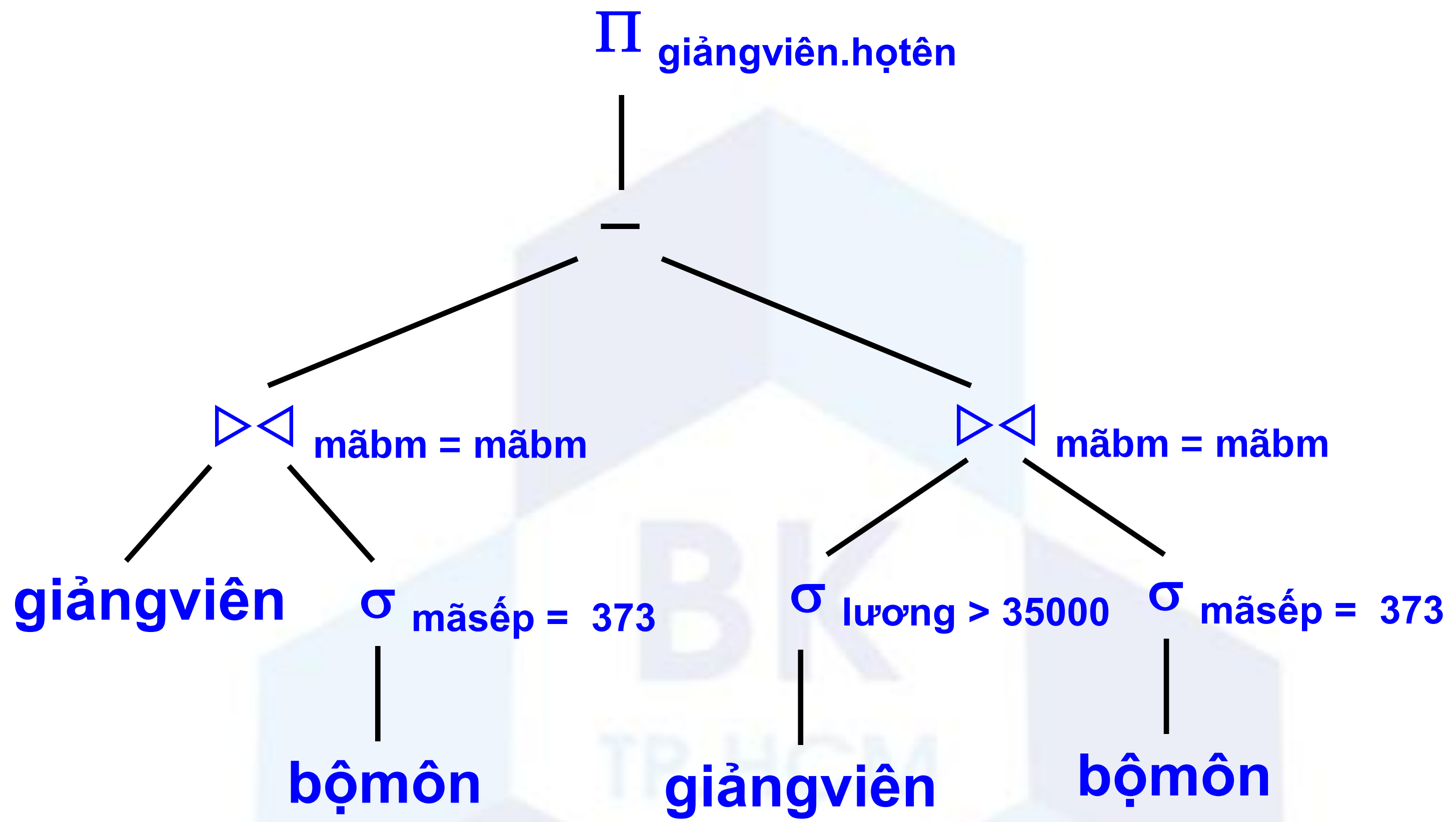
Q_2:

$$\Pi_{\text{giảngviên.họtên}} ((\text{giảngviên} \bowtie_{\text{mãbm = mãbm}} \sigma_{\text{mãsếp = 373}} \text{bộmôn}) - (\sigma_{\text{lương > 35000}} \text{giảngviên} \bowtie_{\text{mãbm = mãbm}} \sigma_{\text{mãsếp = 373}} \text{bộmôn}))$$

  - Biểu thức con chung

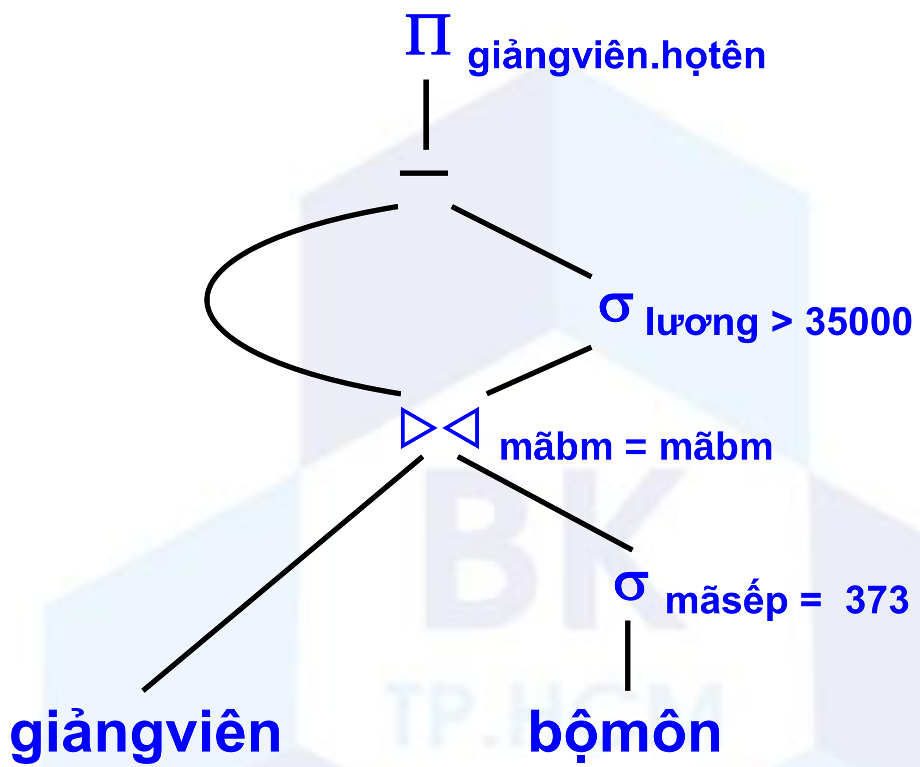
$$\text{giảngviên} \bowtie_{\text{mãbm = mãbm}} \sigma_{\text{mãsếp = 373}} \text{bộmôn}$$

Đồ thị toán tử và biểu thức con chung



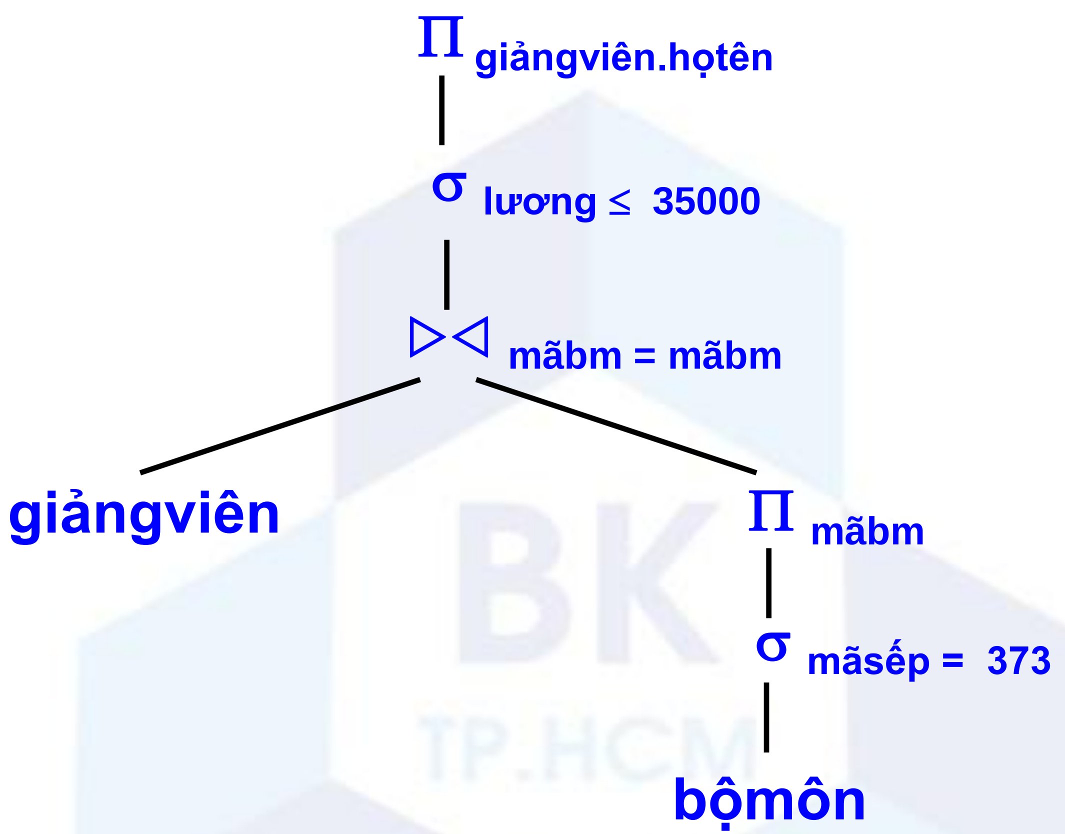
Hình 4.3. (a) Các bước đơn giản hóa của truy vấn Q_2

Đồ thị toán tử và biểu thức con chung



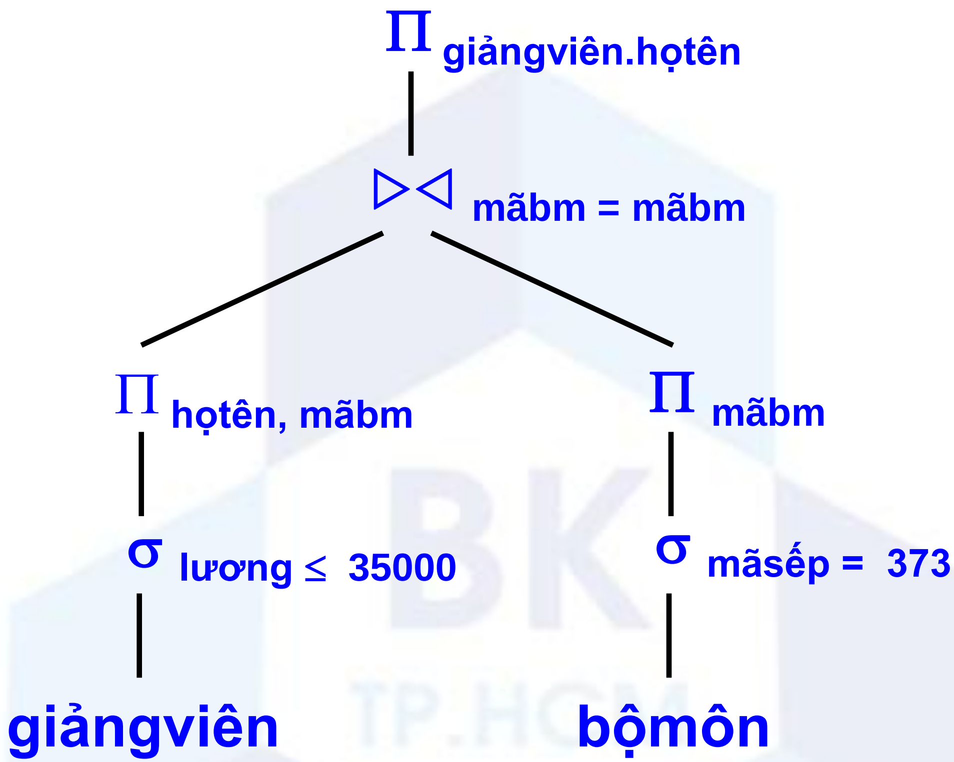
Hình 4.3. (b) Các bước đơn giản hóa của truy vấn Q_2

Đồ thị toán tử và biểu thức con chung



Hình 4.3. (c) Các bước đơn giản hóa của truy vấn Q_2

Đồ thị toán tử và biểu thức con chung



Hình 4.3. (d) Các bước đơn giản hóa của truy vấn Q_2

Đồ thị toán tử và biểu thức con chung

- Sử dụng các phép biến đổi tương đương để đơn giản hóa cây toán tử.

$$R \rhd\lhd R \leftrightarrow R$$

$$R \cup R \leftrightarrow R$$

$$R - R \leftrightarrow \varnothing$$

$$R \rhd\lhd \sigma_F R \leftrightarrow \sigma_F R$$

$$R \cup \sigma_F R \leftrightarrow R$$

$$R - \sigma_F R \leftrightarrow \sigma_{\neg F} R$$

$$(\sigma_{F1} R) \rhd\lhd (\sigma_{F2} R) \leftrightarrow \sigma_{F1 \wedge F2} R$$

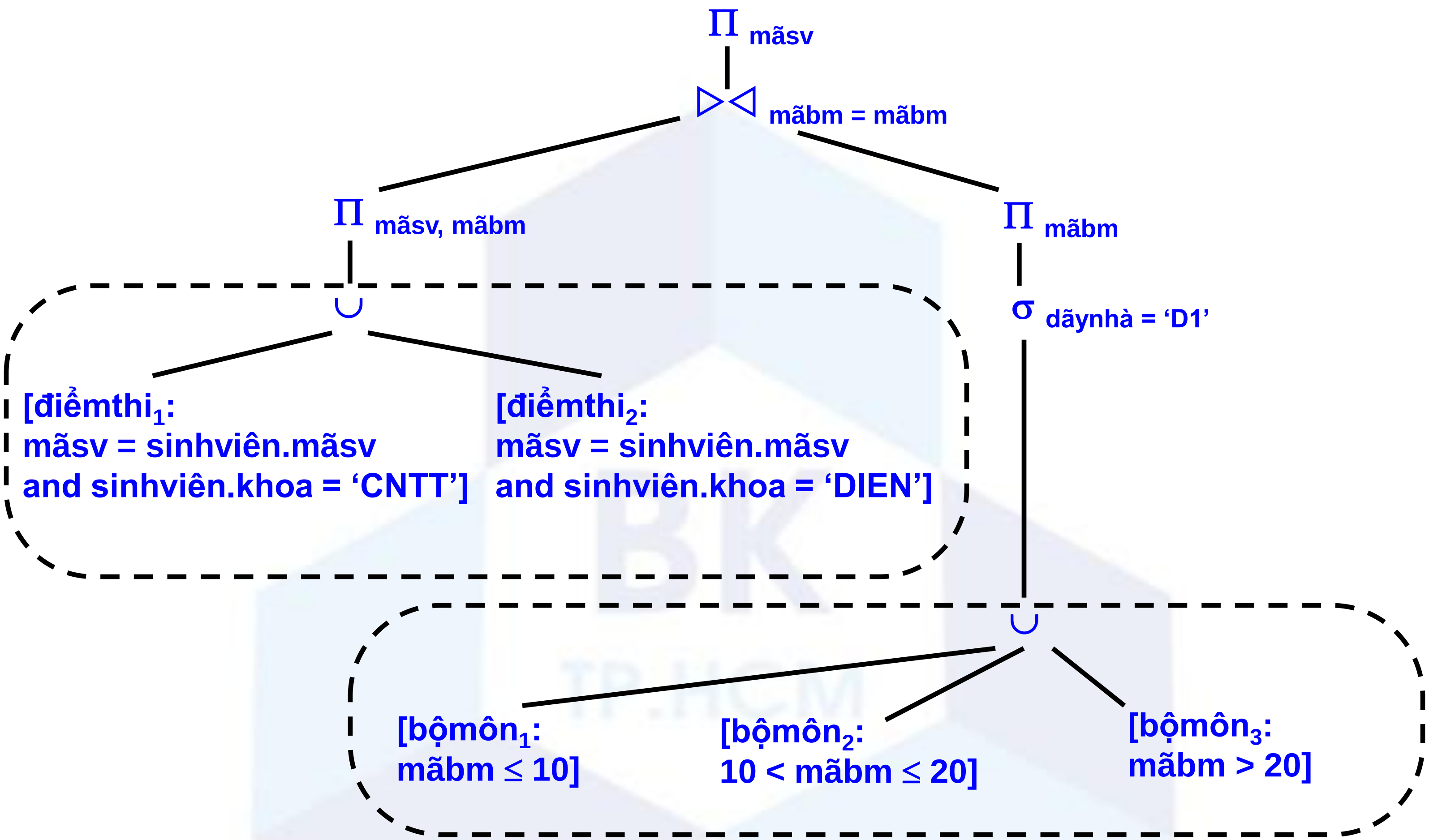
$$(\sigma_{F1} R) \cup (\sigma_{F2} R) \leftrightarrow \sigma_{F1 \vee F2} R$$

$$(\sigma_{F1} R) - (\sigma_{F2} R) \leftrightarrow \sigma_{F1 \wedge \neg F2} R$$

Biểu thức chuẩn tắc

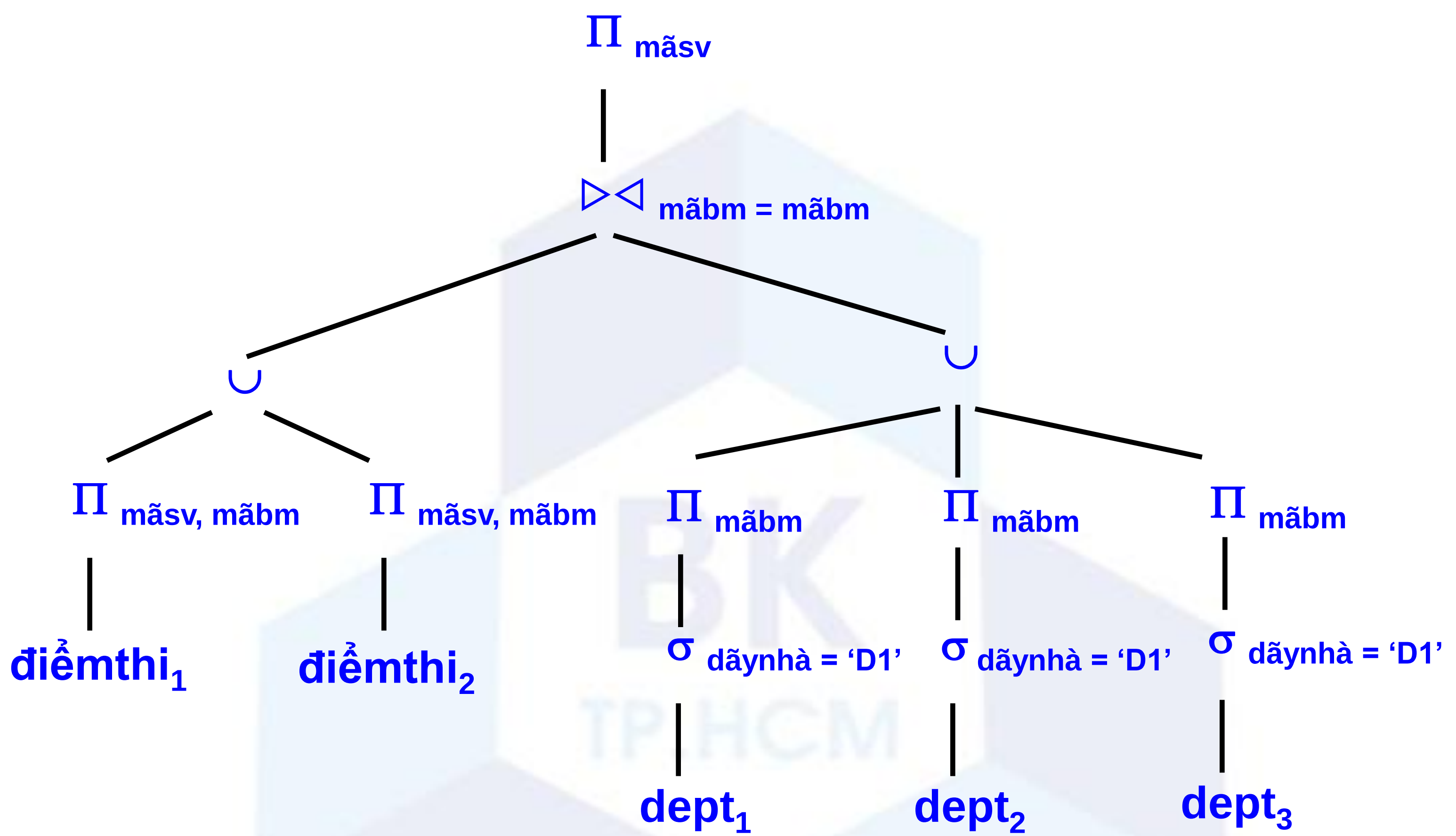
- ***Biểu thức chuẩn tắc*** **(*canonical expression*) của một biểu thức đại số quan hệ trên lược đồ toàn cục có được bằng cách thay thế mỗi tên quan hệ toàn cục xuất hiện trong nó bởi biểu thức đại số quan hệ tái tạo các quan hệ toàn cục từ các mảnh.**
- **Biểu thức chuẩn tắc cũng là một biểu thức đại số quan hệ.**
- **Có thể sử dụng các phép biến đổi tương đương, đặc biệt 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.**

Biểu thức chuẩn tắc



Hình 4.4. (a) Dạng chuẩn tắc của truy vấn Q_1

Biểu thức chuẩn tắc



Hình 4.4. (b) Đẩy các phép chọn và phép chiếu xuống dưới cây.

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.
- Vị từ định tính (*qualification*) có thể được xem là một đặc tính ngữ nghĩa của tất cả các bộ của quan hệ.
- 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

- Các mảnh ngang là các ví dụ tiêu biểu của các quan hệ định tính, trong đó vị từ định tính tương ứng với vị từ phân mảnh.
- Đại số quan hệ định tính là sự mở rộng của đại số quan hệ, trong đó các toán hạng là các quan hệ định tính.

Đại số quan hệ định tính

- Các quy tắc sau đây xác định kết quả của các phép toán đại số quan hệ cho các 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**

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

Đại số quan hệ định tính

- **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**

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

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.**

Đại số quan hệ định tính

- Các quan hệ định tính với các vị từ định tính mâu thuẫn về thực chất là quan hệ rỗng.
- Các phép biến đổi tương đương (*R* có thể là một quan hệ thông thường hoặc là một quan hệ định tính):

$$\sigma_F(\emptyset) \leftrightarrow \emptyset \qquad R \bowtie_F \emptyset \leftrightarrow \emptyset$$

$$\Pi_A(\emptyset) \leftrightarrow \emptyset \qquad R \rhd\!<_F \emptyset \leftrightarrow \emptyset$$

$$R \times \emptyset \leftrightarrow R \qquad \emptyset \bowtie_F R \leftrightarrow \emptyset$$

$$R \cup \emptyset \leftrightarrow R$$

$$R - \emptyset \leftrightarrow R$$

$$\emptyset - R \leftrightarrow \emptyset$$

Tiêu chuẩn 3 và 4

- Mục đích là 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 *bộmôn* được phân mảnh ngang:**

$$Q_3: \sigma_{\text{mãbm} = 1}\ \text{bộmôn}$$

Đơn giản hóa các quan hệ được phân mảnh ngang

$\sigma_{\text{mãbm} = 1}$

$\cup$

[bộmôn$_1$: mãbm $\leq$ 10] [bộmôn$_2$: 10 < mãbm $\leq$ 20] [bộmôn$_3$: mãbm > 20]

Hình 5.5. (a) Dạng chuẩn tắc của truy vấn Q_3

$\sigma_{\text{mãbm} = 1}$

[bộmôn$_1$: mãbm $\leq$ 10]

Hình 4.5. (b) Đơn giản hóa truy vấn Q_3

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

- Để đơn giản, xét phép kết giữa hai quan hệ được phân mảnh ngang là *R* và *S*. Có hai giải pháp để kết chúng:
  - **Giải pháp 1:** Tập hợp tất cả các mảnh của *R* và của *S* trước khi thực hiện phép kết.

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

  - **Giải pháp 2:** Thực hiện phép kết giữa các mảnh, sau đó tập hợp tất cả các kết quả vào trong cùng một quan hệ kết quả. Phép kết này được gọi là ***phép kết phân tán*** (*distributed join*).

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

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

- Trong hai giải pháp trên thì không có giải pháp nào hay hơn.
  - 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.
- Đối với một phép kết cho trước, đồ thị kết cho thấy các mảnh nào cần được kết với nhau.

Tiêu chuẩn 5

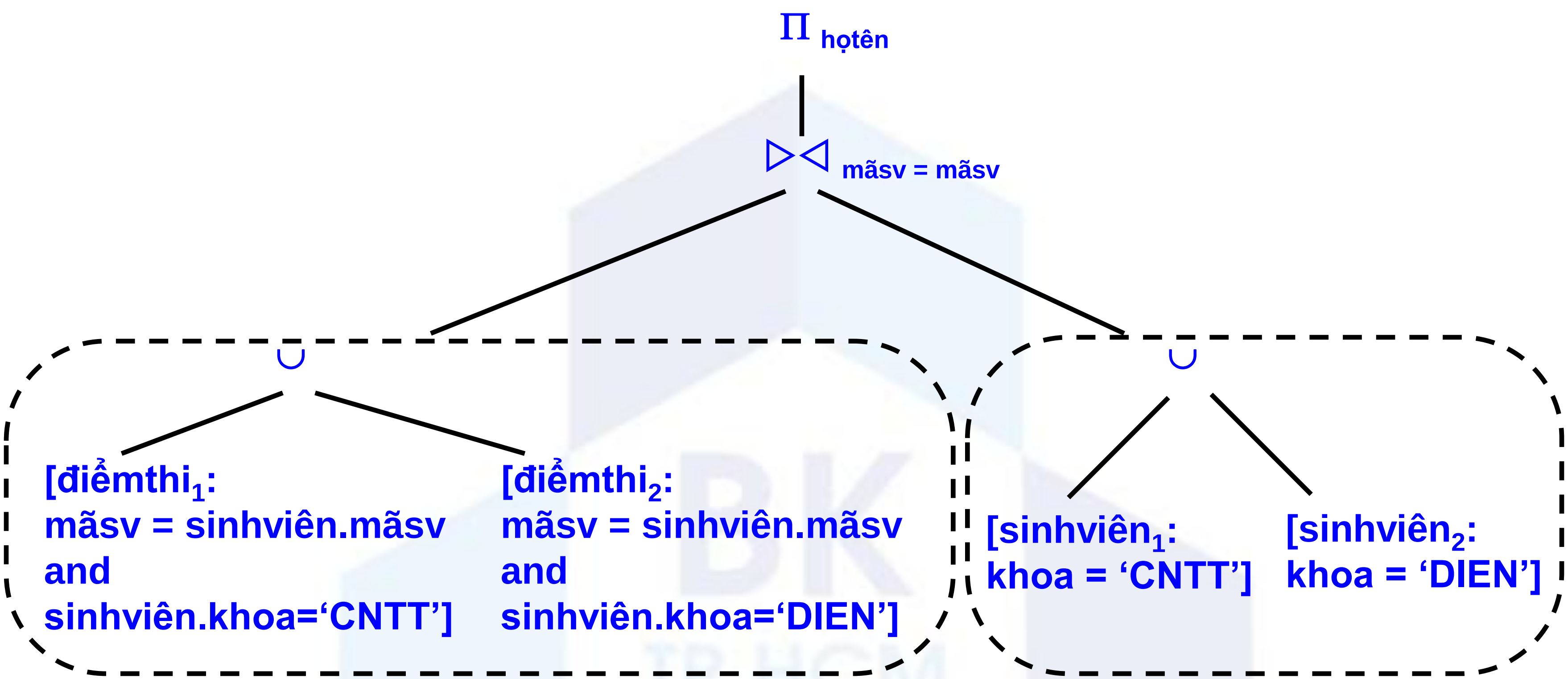
- Mục đích là biến đổi một truy vấn không có các phép kết phân tán thành một truy vấn có phép kết phân tán.
- Tiêu chuẩn 5 - Để phân phối các phép kết xuất hiện trong một truy vấn toàn cục, các phép hợp (biểu diễn việc tập hợp các mảnh) phải được đẩy lên phía trên các phép kết muốn phân phối.

Tiêu chuẩn 5

- Việc xây dựng một đồ thị kết cần phải áp dụng tiêu chuẩn 5 (để phân phối phép kết), kế tiếp là tiêu chuẩn 4 (để loại bỏ các phép kết giữa các mảnh mà kết quả là rỗng).
- Ví dụ
  - Truy vấn Q_4 - Hãy cho biết họ tên của tất cả sinh viên có điểm thi:

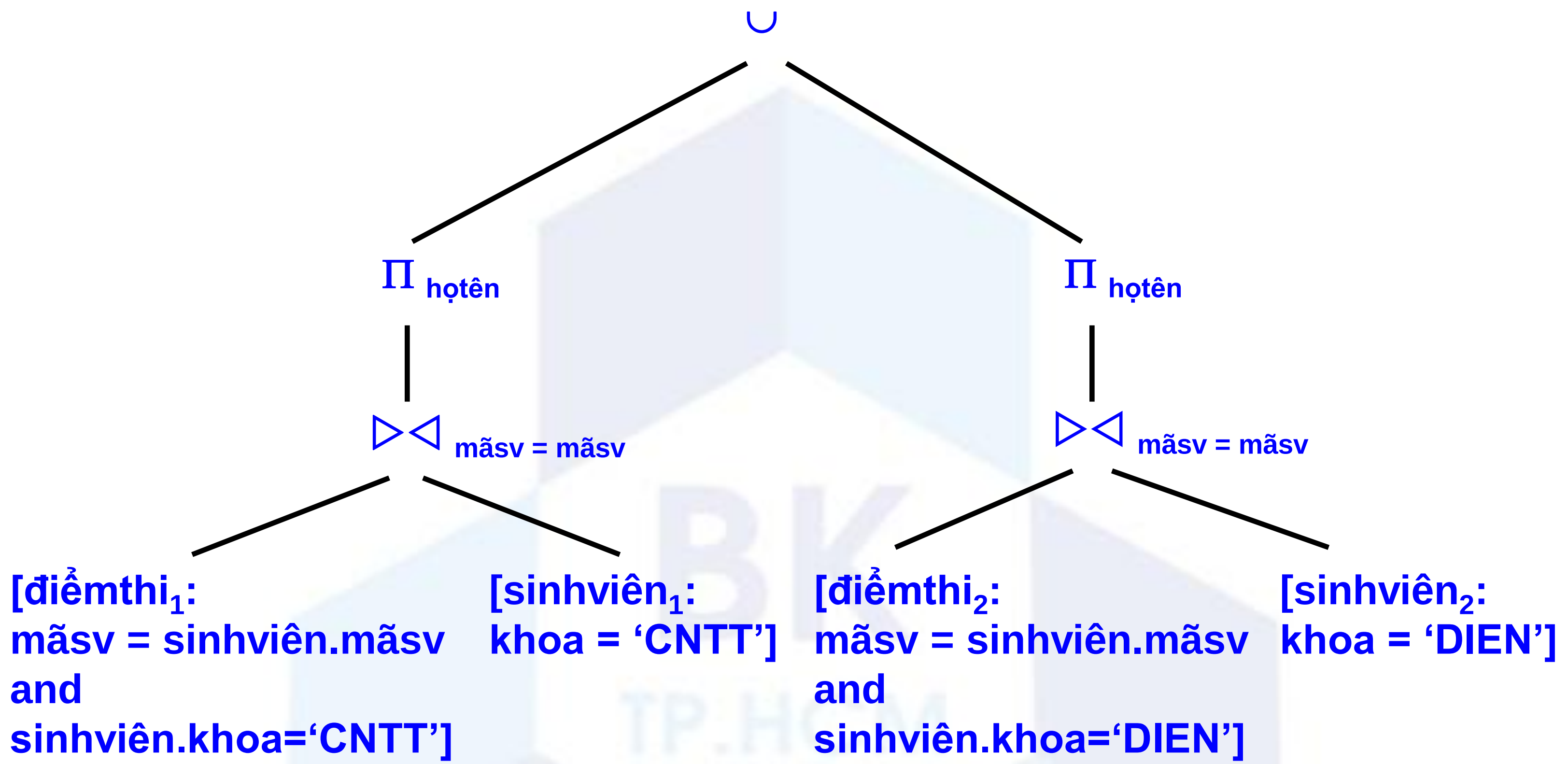
$$Q_4: \Pi_{\text{họtên}} (\text{điểmthi} \bowtie \text{sinhviên})$$

Tiêu chuẩn 5



Hình 4.6. (a) Dạng chuẩn tắc của truy vấn Q_4

Tiêu chuẩn 5



Hình 4.6. (b) Phép kết phân tán dùng cho truy vấn Q_4

Sử dụng phép suy diễn cho các phép đơn giản hóa

- Xác định các điều mâu thuẫn giữa các điều kiện chọn của các truy vấn và các vị từ định tính của các mảnh.
- Việc xác định một công thức là mâu thuẫn có thể được thực hiện bằng cách sử dụng các thông tin ngữ nghĩa phức tạp. Sử dụng ***bộ chứng minh định lý*** (*theorem prover*).
- Ví dụ
  - Xét truy vấn Q_1 – Cho biết mã của các sinh viên có học các môn học thuộc các bộ môn ở dãy nhà D1.
  - Cây toán tử của Q_1 trong Hình 4.4.

Sử dụng phép suy diễn cho các phép đơn giản hóa

- **Giả sử:**

(1) Dãy nhà D1 chỉ bao gồm các bộ môn có mã từ 1 đến 10.

(2) Tất cả môn học thuộc các bộ môn có mã từ 1 đến 10 dành riêng cho các sinh viên thuộc khoa CNTT.

- **Sử dụng các điều hiểu biết này để suy ra các mâu thuẫn cho phép loại bỏ các biểu thức con.**
- **Từ (1), có thể viết các điều suy diễn sau đây:**

dãynhà = 'D1' $\Rightarrow$ not (10 < mãbm $\leq$ 20)

dãynhà = 'D1' $\Rightarrow$ not (mãbm > 20)

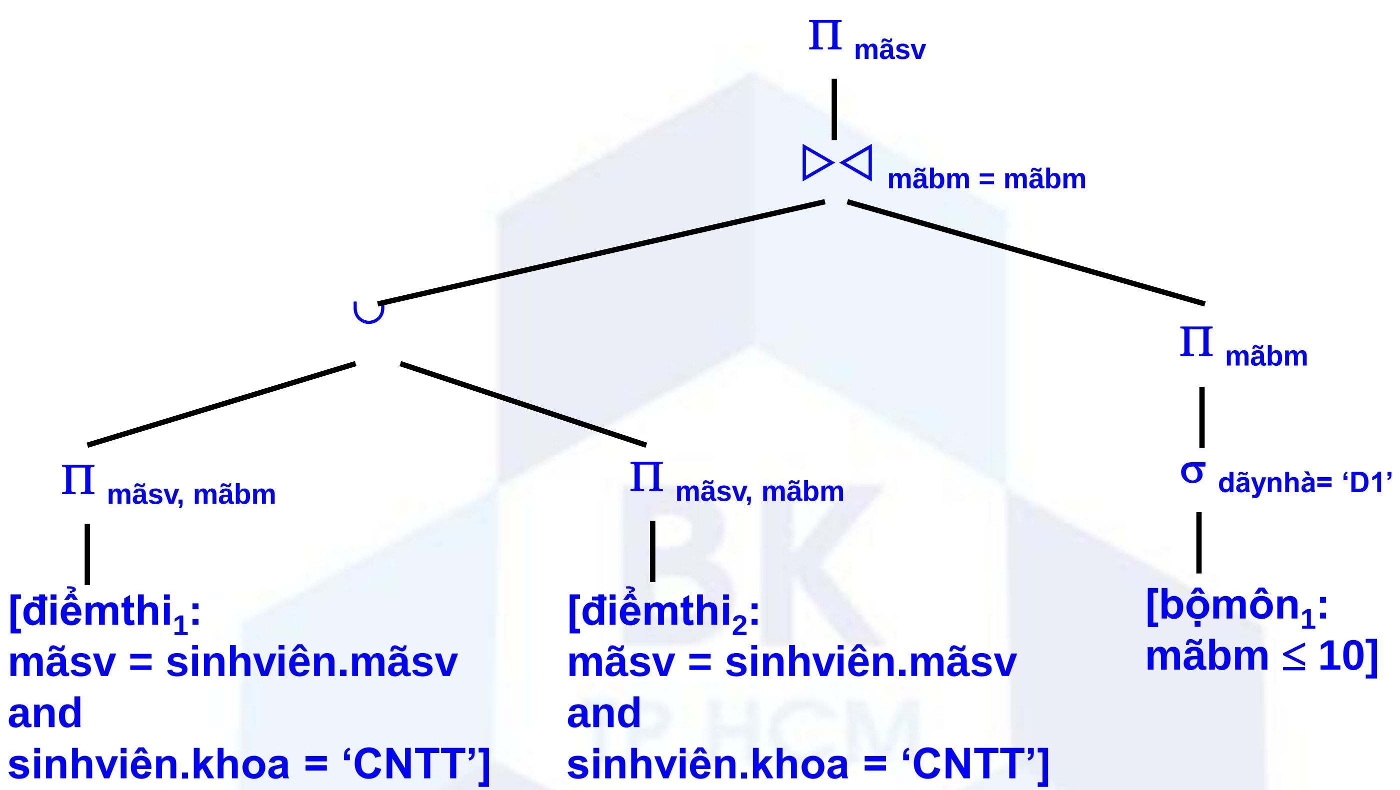
dãynhà = 'D1' $\Rightarrow$ mãbm $\leq$ 10

- **Từ (2) :**

mãbm $\leq$ 10 $\Rightarrow$

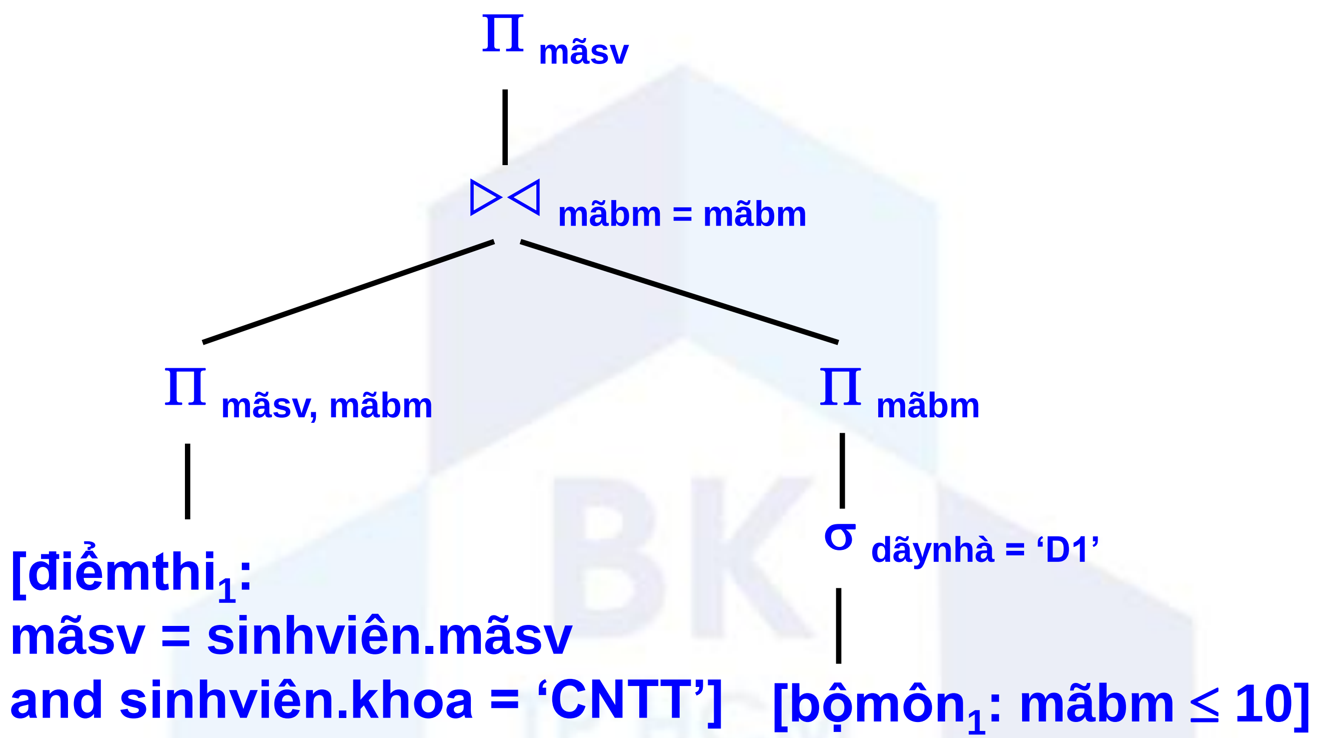
not (mãsv = sinhviên. mãsv and sinhviên.khoa = 'DIEN')

Sử dụng phép suy diễn cho các phép đơn giản hóa



Hình 4.7. (a) Đơn giản hóa cây toán tử bằng sự suy diễn.

Sử dụng phép suy diễn cho các phép đơn giản hóa



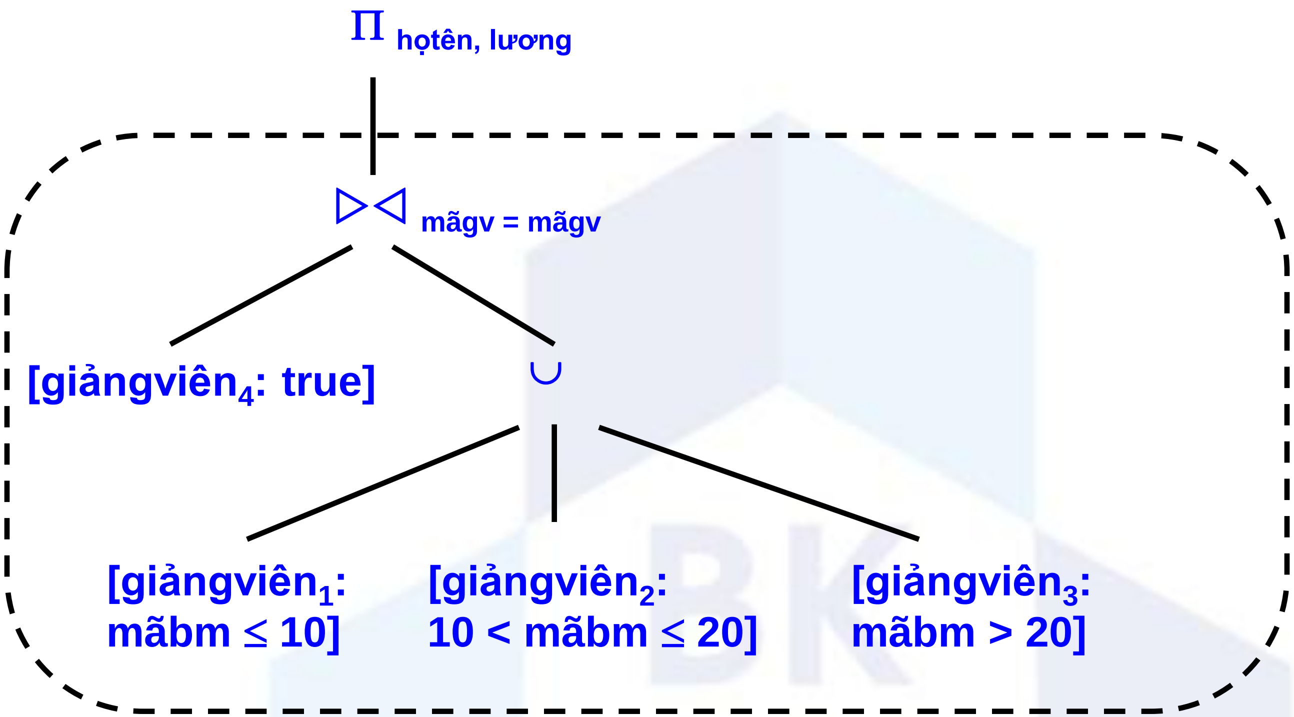
Hình 4.7. (b) Đơn giản hóa cây toán tử bằng sự suy diễn.

Đơn giản hoá các quan hệ được phân mảnh dọc

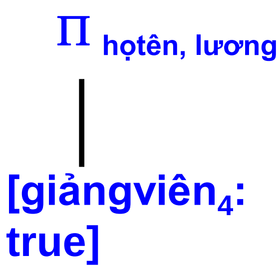
- Mục đích là xác định một tập con bao gồm các mảnh đủ để trả lời truy vấn, sau đó loại bỏ tất cả các mảnh khác từ biểu thức truy vấn và các phép kết được dùng trong phép đổi ngược của lược đồ phân mảnh để tái tạo các quan hệ toàn cục.
- **Ví dụ**
  - Truy vấn Q_5 – Hãy cho biết tên và tiền lương của các nhân viên:

$$Q_5: \Pi_{\text{họtên, lương}}\ \text{giảngviên}$$

Đơn giản hoá các quan hệ được phân mảnh dọc



Hình 4.8. (a) Dạng chuẩn tắc của truy vấn Q_5



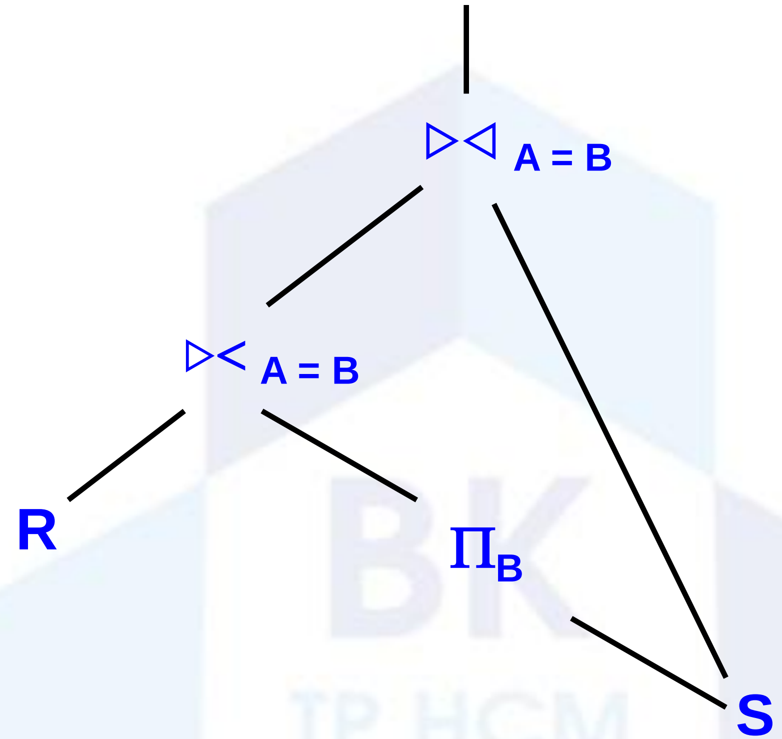
Hình 4.8. (b) Truy vấn Q_5 đã được đơn giản hóa.

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*) $R \rhd\lhd_{A = B} S$, 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:**

$$S \rhd\lhd_{A = B} (R \rhd<_{A = B} \Pi_{S.B})$$

Chương trình nửa kết



Hình 4.9. Đồ thị toán tử của chương trình nửa kết $R \rhd\!\!<_{A = B} S$.

Chương trình nửa kết

Chương trình nửa kết

(1) Thực hiện $T_1 = \Pi_B$ *S* tại nơi của *S*.

(2) Gửi T_1 đến nơi của *R*.

(3) Thực hiện $T_2 = R \triangleright\!<_{A=B} T_1$ tại nơi của *R*.

(4) Gửi T_2 đến nơi của *S*.

(5) Thực hiện $T_3 = T_2 \triangleright\!\triangleleft_{A=B} S$ tại nơi của *S*.

(6) Gửi kết quả cuối cùng T_3 đến nơi của ứng dụng.

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(*điểm*)
from *điểmthi*
where *mãmh* = 'CNPM';

$\Psi_{\text{AVG(điểmthi)}}\ \sigma_{\text{mãmh = 'CNPM'}}$ điểmthi

Q_7: select *mãsv*, *mãmh*, SUM(*điểm*)
from *điểmthi*
group by *mãsv*, *mãmh*;

$\Psi_{\text{mãsv, mãmh, SUM(điểm)}}$ điểmthi

Q_8: select *mãsv*, *mãmh*, SUM(*điểm*)
from *điểmthi*
group by *mãsv*, *mãmh*
having SUM(*điểm*) > 300;

$\sigma_{\text{SUM(quan) > 300}}\ \Psi_{\text{mãsv, mãmh, SUM(điểm)}}$ điểmthi

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.

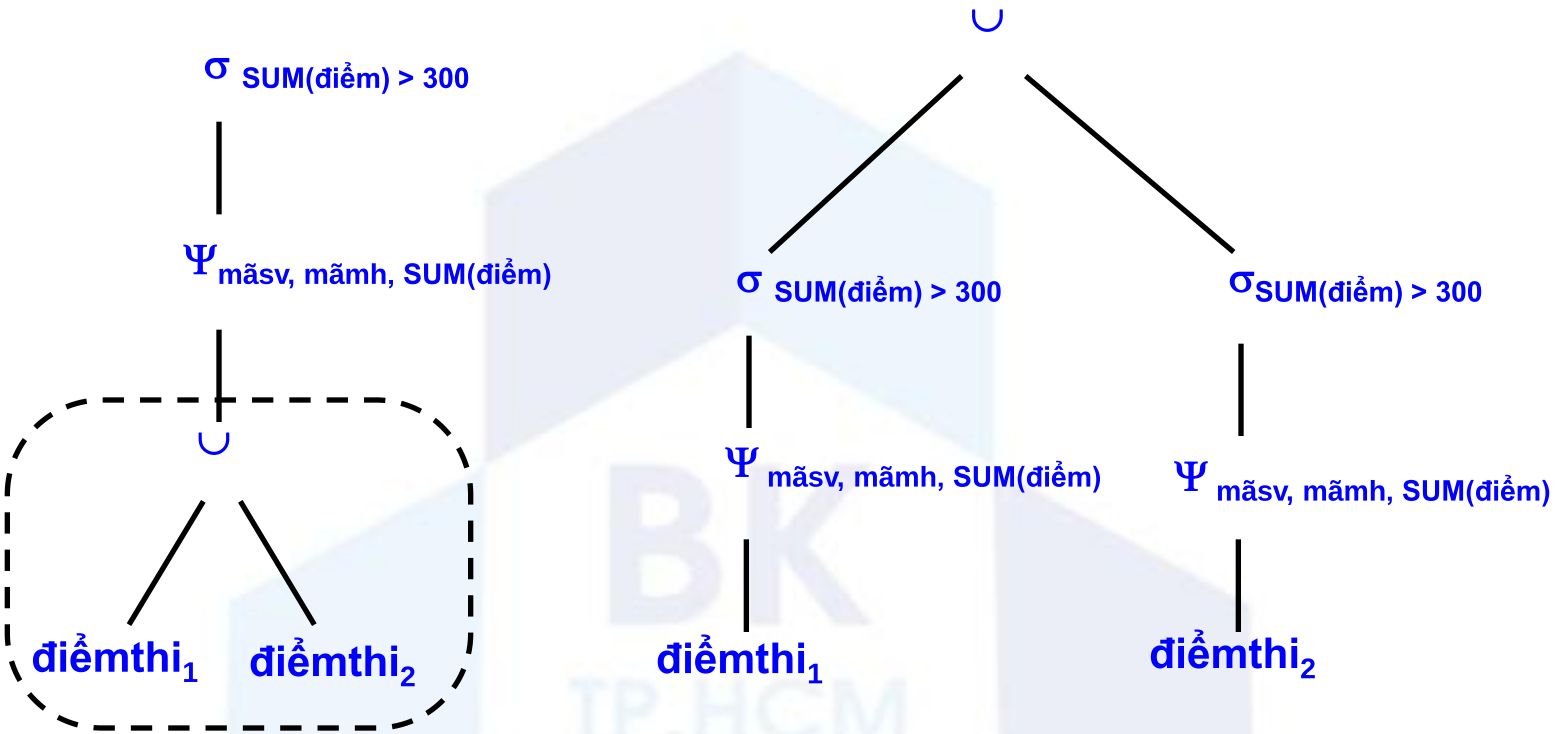
$$\forall i, j : (G_i \subseteq R_j) \text{ hoặc } (G_i \cap R_j = \varnothing)$$

  - Mỗi nhóm phải được chứa hoàn toàn trong một mảnh, nghĩa là việc gom nhóm phải nhỏ hơn việc phân mảnh.
  - Việc thực hiện phép gom nhóm trên các toán hạng của phép hợp và sau đó hợp các kết quả này thì tương đương với việc thực hiện phép gom nhóm trực tiếp trên kết quả của phép hợp.

Tiêu chuẩn 6

- **Mục đích là 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 Q_8 (b) Bản phân tán của truy vấn Q8

Hình 4.10. Một truy vấn với việc gom nhóm và các hàm kết hợp.

Tính chất của hàm kết hợp

- Hàm kết hợp *F* có tính toán phân tán nếu đối với một đa tập (*multiset*) *S* bất kỳ và một phân rã bất kỳ của *S* thành các đa tập S_1, S_2,…, S_n, thì có thể xác định một tập hợp các hàm kết hợp F_1,…, F_m và một biểu thức $E(F_1, \ldots, F_m)$ sao cho:

$$F(S) = E(F_1(S_1),\ldots,F_1(S_n),F_2(S_1),\ldots,F_2(S_n),\ldots,F_m(S_1),\ldots,F_m(S_n))$$

Tính chất của hàm kết hợp

- **Hàm tìm giá trị nhỏ nhất**

$$MIN(S) = MIN(MIN(S_1), MIN(S_2), \ldots, MIN(S_n))$$

- **Hàm tìm giá trị lớn nhất**

$$MAX(S) = MAX(MAX(S_1), MAX(S_2), \ldots, MAX(S_n))$$

- **Hàm đếm**

$$COUNT(S) = SUM(COUNT(S_1), COUNT(S_2), \ldots, COUNT(S_n))$$

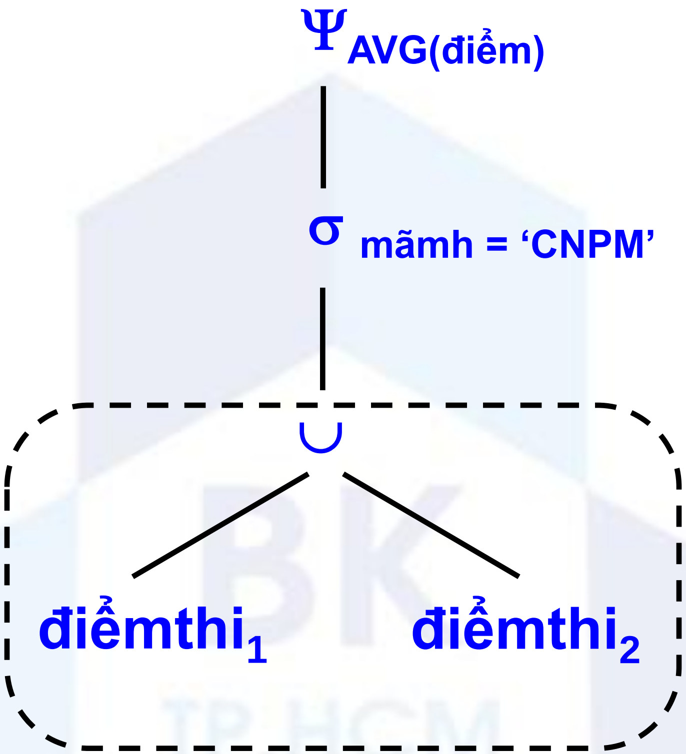
- **Hàm tính giá trị tổng cộng**

$$SUM(S) = SUM(SUM(S_1), SUM(S_2), \ldots, SUM(S_n))$$

- **Hàm tính giá trị trung bình**

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

Tính chất của hàm kết hợp



Hình 4.11. (a) Định trị phân tán của các hàm kết hợp.

Tính chất của hàm kết hợp

$$E{:}AVG(\text{điểm}) = \frac{SUM(S_1, S_2)}{SUM(C_1, C_2)}$$

$S_1,C_1{:}\Psi_{SUM(\text{điểm}),\ COUNT}$ $S_2,C_2{:}\Psi_{SUM(\text{điểm}),\ COUNT}$

$\sigma_{\text{mãmh} = \text{‘CNPM’}}$ $\sigma_{\text{mãmh} = \text{‘CNPM’}}$

điểmthi$_1$ điểmthi$_2$

Hình 4.11. (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.

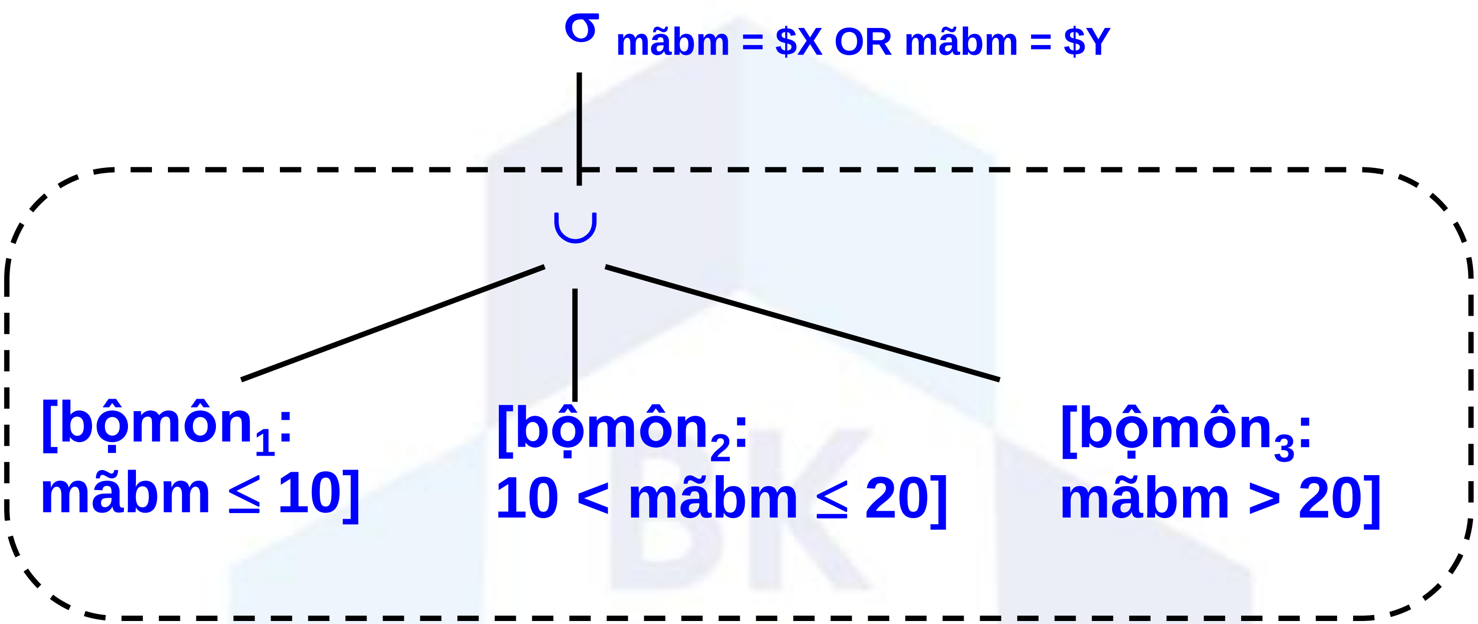
Đơn giản hóa truy vấn có tham số

- **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:

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

  - **Ở thời gian biên dịch:** không biết các mảnh nào của quan hệ toàn cục *dept* sẽ được sử dụng. Tuy nhiên, nhiều nhất là hai mảnh của nó sẽ có liên quan đến truy vấn.
  - **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 4.12. (a) Dạng chuẩn tắc của truy vấn Q_9

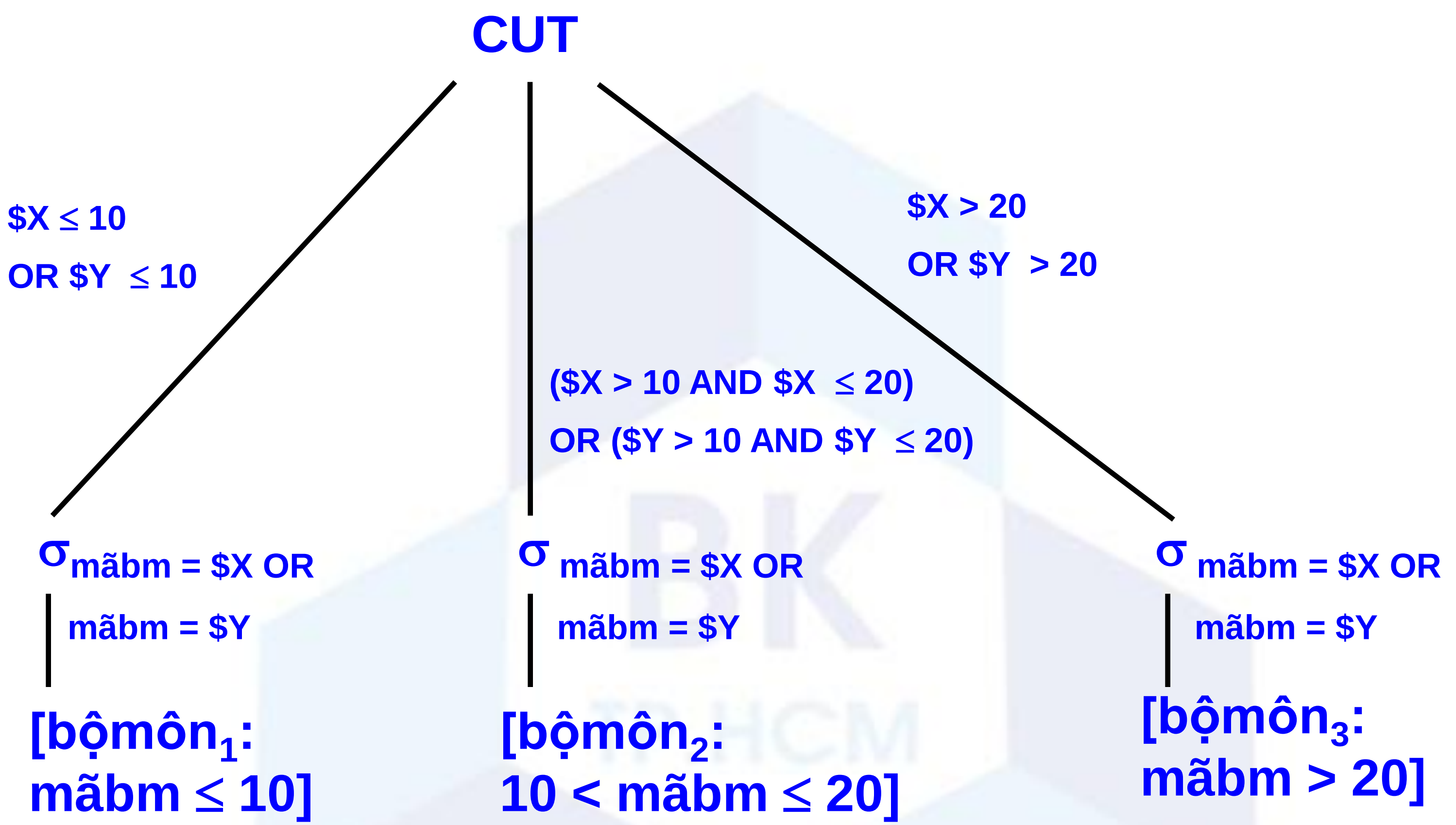
Đơn giản hóa truy vấn có tham số

- Phép đơn giản hóa truy vấn có tham số liên quan đến việc 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.
- Để biểu diễn sự hiện diện của các kiểm tra trên cây toán tử đối với 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

- Đối với mỗi toán hạng, phép toán *CUT* có một mô tả gồm:
  (1) Một công thức sử dụng các tham số của truy vấn .
  (2) Vị từ định tính của các mảnh.
  - Mỗi công thức được chuẩn bị ở thời gian biên dịch và được định trị ở thời gian chạy.
  - Nếu công thức trả về *true*, thì toán hạng tương ứng được đưa vào phép hợp.
  - Nếu công thức trả về *false*, thì toán hạng tương ứng bị loại bỏ ra khỏi cây.

Đơn giản hóa truy vấn có tham số



Hình 4.12. (b) Cây truy vấn với phép CUT.

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

- Truy vấn có tham số được sử dụng nhiều lần, với các giá trị khác nhau của các tham số.
- Để làm giảm chi phí thực hiện, có thể sử dụng các quan hệ tạm thời ở nơi gốc của truy vấn, mà nơi này lưu trữ một tập cha chứa dữ liệu cần thiết cho mỗi lần lặp.
- 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):

$$Q_{10}: \Pi_{\text{hotên}}\ \sigma_{\text{mãsếp} = \$X \text{ AND } \text{mãbm} = 12}\ \text{giảngviên}$$

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

$\Pi_{\text{họtên}}$

$\sigma_{\text{mãsếp = \$X}}$

$\Pi_{\text{họtên, mãsếp}}$

$\sigma_{\text{mãbm = 12}}$

giảngviên$_2$: $10 < \text{mãbm} \leq 20$

$\Pi_{\text{họtên}}$

$\sigma_{\text{mãsếp = \$X}}$

T

$T \equiv \Pi_{\text{họtên, mãsếp}}$

$\sigma_{\text{mãbm = 12}}$

giảngviên$_2$: $10 < \text{mãbm} \leq 20$

Hình 4.13. Sử dụng các quan hệ tạm thời cho các truy vấn có tham số.