ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

TÔ NGỌC ANH

CÁC KỸ THUẬT PHÂN MẢNH, GỘP NHÓM TRONG CSDL PHÂN TÁN

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2013

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

TÔ NGỌC ANH

CÁC KỸ THUẬT PHÂN MẢNH, GỘP NHÓM TRONG CSDL PHÂN TÁN

Chuyên ngành: Khoa học máy tính
Mã số: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: PGS.TS LÊ HUY THẬP

Thái Nguyên - 2013

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này là do bản thân tự nghiên cứu và thực hiện theo sự hướng dẫn khoa học của ***PGS. TS. Lê Huy Thập***

Tôi hoàn toàn chịu trách nhiệm về tính pháp lý quá trình nghiên cứu khoa học của luận văn này.

Người Cam Đoan

TÔ NGỌC ANH

LỜI CẢM ƠN

Lời đầu tiên tôi xin gửi lời cảm ơn đến thầy giáo ***PGS. TS. Lê Huy Thập*** đã định hướng, hướng dẫn và giúp đỡ tôi rất nhiều về mặt chuyên môn trong quá trình tìm hiểu và thực hiện luận văn.

Tôi xin gửi lời biết ơn sâu sắc đến các thầy, các cô đã dạy dỗ và truyền đạt những kinh nghiệm quý báu cho chúng tôi trong suốt hai năm cao học ở trường Đại học Công nghệ thông tin và truyền thông Thái Nguyên.

Cuối cùng, xin chân thành cảm ơn gia đình và bạn bè đã động viên, quan tâm, giúp đỡ tôi hoàn thành khóa học và luận văn.

Thái nguyên, tháng 12 năm 2013

Tác giả

Tô Ngọc Anh

MỤC LỤC

MỞ ĐẦU

1. Đặt vấn đề

Nhằm giải quyết vấn đề chậm trễ thường gặp trong các hệ *CSDL* song song, ngoài việc áp dụng một kiến trúc phần cứng thích hợp, người ta tiến hành phân mảnh dữ liệu một cách hợp lý cho các bộ xử lý. Một chiến lược phân mảnh dữ liệu tốt sẽ tăng mức độ thực hiện song song đồng thời khai thác tốt hơn các hàm gộp nhóm từ các mảnh. Chúng ta sẽ đề cập đến một số kỹ thuật phân mảnh dữ liệu theo chiều ngang phổ biến như phân mảnh theo vòng tròn Robin, phân mảnh theo hàm băm, phân mảnh theo khoảng, phân mảnh theo chiều dọc, ... và một số hàm gộp nhóm trong CSDL phân tán như: SUM, COUNT, AVERAGE...

2. Đối tượng và phạm vi nghiên cứu

Các hàm gộp nhóm trong cơ sở dữ liệu quan hệ

Các phương pháp phân mảnh

Các hàm gộp nhóm trong trường hợp CSDL phân tán

3. Hướng nghiên cứu của đề tài

Nghiên cứu các phương pháp phân mảnh.

Nghiên cứu các hàm gộp nhóm.

Nghiên cứu cách đưa các hàm gộp nhóm vào các mảnh và ứng dụng

4. Những nội dung nghiên cứu chính

Luận văn được trình bày trong 3 chương, có phần mở đầu, phần kết luận, phần mục lục, phần tài liệu tham khảo. Các nội dung cơ bản của luận văn được trình bày theo cấu trúc như sau:

Mở đầu

Chương 1: Cơ sở lý thuyết

Chương 2: Phân mảnh và gộp nhóm trong CSDL phân tán

Chương 3: Ứng dụng

Kết luận và hướng phát triển của luận văn

Chương 1: CƠ SỞ LÝ THUYẾT

1.1. GIỚI THIỆU VỀ LOGIC

1. Mệnh đề là một phát biểu để diễn tả một ý tưởng trọn vẹn và chúng ta có thể khẳng định một cách khách quan là đúng hoặc sai, nó không thể vừa đúng lại vừa sai, hay mang tính chất mập mờ.
2. Giá trị trị đúng hay sai của mệnh đề được gọi là chân trị của mệnh đề. Chân trị đúng của mệnh đề thường được kí hiệu là 1 hoặc T hoặc True, còn chân trị sai được kí hiệu là 0 hoặc F hoặc False
3. Mệnh đề logic tuy đơn giản nhưng rất quan trọng trong khoa học máy tính. Là cơ sở lập luận hàng ngày và trong lập trình.

Ví dụ 1.1.1.

1. "12 là số chẵn" là mệnh đề đúng
2. "12 là số nguyên tố" là mệnh đề sai
3. "x + ay = z" không phải mệnh đề

Các kí hiệu dùng trong mệnh đề logic

( ) dùng để gom nhóm biểu thức logic

$\neg$ Phủ định (NOT)

$\wedge$ Hội (Conjunction AND)

$\vee$ Tuyển (Disjunction OR)

$\rightarrow$ Ký hiệu điều kiện (If…Then…)

$\leftrightarrow$ Kéo theo hai chiều (If AND Only If)

Chúng ta giả thiết rằng tập các ký tự trong biểu thức logic là hữu hạn hoặc đếm được, nhưng hầu hết các kết luận vẫn đúng cho trường hợp không đếm được.

Mệnh đề được chia làm hai loại cơ bản: mệnh đề sơ cấp (elementary), nó là các nguyên tử (atom)-không thể chia nhỏ được; mệnh đề phức hợp (compound), mệnh đề được tạo ra từ một hoặc nhiều mệnh đề khác bằng cách sử dụng các phép toán mệnh

Để máy tính hiểu được, chúng ta dùng các kí hiệu cho các mệnh đề, các biến mệnh đề thường được dùng là các chữ thường.

Ví dụ 1.1.2.

p = "15 MOD 3 = 0", là mệnh đề sơ cấp.

r = "15 MOD 3 = 0" AND "3 là số nguyên tố", là mạnh đề phức hợp

Các phép toán mệnh đề: $\neg$ (phủ định) ; $\wedge$ (hội) ; $\vee$ (tuyển) ; $\veebar$ (hoặc hay tổng trực giao) ; $\rightarrow$ (kéo theo) ; $\leftrightarrow$ (kéo theo hai chiều)

Biểu thức logic

Biểu thức logic có thể nói chính là mệnh đề phức hợp, biểu thức logic thường được ký hiệu bởi các chữ in to và nó là sự kết hợp của:

- *Các mệnh đề hay các giá trị hằng*
- *Các biến mệnh đề hoặc các biểu thức logic*
- *Các phép toán logic và các dấu ( )*
- ***Bảng chân trị của các phép toán mệnh đề***

p	q	$\neg$p	p $\wedge$ q	p $\vee$ q	p $\veebar$ q	p $\rightarrow$ q	p $\leftrightarrow$ q
0	0	1	0	0	0	1	1
0	1	1	0	1	1	1	0
1	0	0	0	1	1	0	0
1	1	0	1	1	0	1	1

Bảng chân trị của các phép toán mệnh đề

Mức ưu tiên của các phép toán logic

Thứ tự ưu tiên của các phép toán logic được liệt kê theo mức yếu dần từ trên xuống dưới, từ trái qua phải theo bảng sau :

Ký hiệu phép toán	Nghĩa của phép toán
$\neg$, $-$,	Phủ định
$\wedge$, $\vee$	Hội, tuyển
$\rightarrow$, $\leftrightarrow$	Kéo theo, tương đương

Bảng ưu tiên các phép toán mệnh đề

Tương đương của hai biểu thức logic

Hai biểu thức logic E và F được gọi là tương đương với nhau và viết E $\Leftrightarrow$ F khi E và F có cùng chân trị.

Các quy tắc thay thế

Quy tắc 1: (Quy tắc thay thế tương đương).

Cho E là một biểu thức logic, nếu thay thế một biểu thức con của nó bởi một biểu thức tương đương với biểu thức con đó, biểu thức logic E' mới nhận được sẽ tương đương với E.

Quy tắc 2: (Tính bất biến đối với biểu thức logic hằng đúng)

Cho E là biểu thức hằng đúng, nếu thay thế một biến mệnh đề p nào đó trong E bởi một biểu thức logic bất kỳ ta sẽ nhận được biểu thức logic E' mới cũng là hằng đúng.

Tương tự cho biểu thức hằng sai.

Các dạng chính tắc

Biểu thức hội cơ bản.

Biểu thức logic F = F (p1, p2, ...pn), trong đó pi ($i = \overline{1,n}$) là các biến mệnh đề sơ cấp, được gọi là biểu thức hội cơ bản, nếu: F = q1 $\wedge$ q2 $\wedge$ ...$\wedge$ qn ; với qi = pi hoặc qi = $\overline{p}$i ($i = \overline{1,n}$)

Biểu thức tuyển cơ bản.

Biểu thức logic E = E (p1, p2, ...pn), trong đó pi ($i=\overline{1,n}$) là các biến mệnh đề sơ cấp, được gọi là biểu thức tuyển cơ bản, nếu: E = q1 ∨ q2 ∨ ...∨ qn ; với qi = pi hoặc qi = $\bar{p}$i ($i=\overline{1,n}$)

Biểu thức logic E = E (p1, p2, ...pn), trong đó pi ($i=\overline{1,n}$) là các biến mệnh đề sơ cấp, được gọi là dạng tuyển chính tắc, nếu: E = E1 ∨ E2 ∨ ...∨ En ; trong đó mỗi Ei ($i=\overline{1,n}$) là những biểu thức hội cơ bản của các pi ($i=\overline{1,n}$).

Định lý 1:

Mọi biểu thức logic E ($p_1, p_2, ...p_n$) đều tương đương với một biểu thức tuyển chính tắc duy nhất. Tức là E ($p_1, p_2, ...p_n$) $\Leftrightarrow E_1 \vee E_2 \vee ... \vee E_m$ (duy nhất) với E_i ($i=\overline{1,m}$) là các biểu thức hội cơ bản.

Biểu thức logic hội chính tắc

Biểu thức logic F = F (p1, p2, ...pn), trong đó pi ($i=\overline{1,n}$) là các biến mệnh đề sơ cấp, được gọi là dạng hội chính tắc, nếu: F = F1 ∧ F2 ∧ ...∧ Fn , trong đó mỗi Fi ($i=\overline{1,n}$) là một biểu thức tuyển cơ bản của các p_i ($i=\overline{1,n}$)

Định lý 2:

Mọi biểu thức logic F ($p_1, p_2, ...p_n$) đều tương đương với một biểu thức hội chính tắc duy nhất. Tức là F ($p_1, p_2, ...p_n$) $\Leftrightarrow F = F_1 \wedge F_2 \wedge ... \wedge F_m$ (duy nhất) với F_i ($i=\overline{1,m}$) là các biểu thức tuyển cơ bản.

Mộ số luật hay được dùng nhất

1/ Luật phủ định của phủ định: $\neg \neg p \Leftrightarrow p$

2/ Luật giao hoán: $p \vee q \Leftrightarrow q \vee p$

$p \wedge q \Leftrightarrow q \wedge p$

3/ Luật kết hợp: $p \vee (q \vee r) \Leftrightarrow (p \vee q) \vee r$

$p \wedge (q \wedge r) \Leftrightarrow (p \wedge q) \wedge r$

4/ Luật phân phối: $p \vee (q \wedge r) \Leftrightarrow (p \vee q) \wedge (p \vee r)$

$p \wedge (q \vee r) \Leftrightarrow (p \wedge q) \vee (p \wedge r)$

5/ Luật Demorgan: $\neg (p \wedge q) \Leftrightarrow \neg p \vee \neg q$