

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

VŨ ANH TÀI

**XỬ LÝ CÂU TRUY VẤN SQL
QUA PHÁN ĐOÁN VÀ CƯỜNG CHẾ**

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, kết quả của luận văn hoàn toàn là kết quả của tự bản thân tôi tìm hiểu, nghiên cứu. Các tài liệu tham khảo được trích dẫn và chú thích đầy đủ.

Tác giả
VŨ ANH TÀI

LỜI CẢM ƠN

Lời đầu tiên tôi xin gửi lời cảm ơn chân thành nhất đến thầy PGS.TS LÊ HUY THẬP đã định hướng và nhiệt tình hướng dẫn, giúp đỡ tôi rất nhiều về mặt chuyên môn trong quá trình làm 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, Đại học Thái Nguyên.

Cuối cùng tôi xin dành tình cảm thân thiết nhất cho bạn bè, đồng nghiệp, cha mẹ và gia đình, những người luôn gần gũi để động viên, chia sẻ cùng tôi trong suốt thời gian qua.

Thái Nguyên, tháng 9 năm 2013

Vũ Anh Tài

MỤC LỤC

LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN	ii
DANH MỤC CÁC BẢNG.....	v
DANH MỤC CÁC HÌNH.....	vi
MỞ ĐẦU	1
1. Đặt vấn đề.....	1
2. Đối tượng và phạm vi nghiên cứu	1
3. Hướng nghiên cứu của đề tài.....	1
4. Phương pháp nghiên cứu	2
5. Ý nghĩa khoa học của đề tài	2
6. Các kết quả dự kiến đạt được	2
Chương 1: CƠ SỞ LÝ THUYẾT	3
1.1. Giới thiệu về logic	3
1.1.1 Các loại mệnh đề	3
1.1.2. Khái niệm vị từ.....	4
1.2. Tổng quan về cơ sở dữ liệu phân tán.....	4
1.2.1. Các phương pháp phân mảnh	4
1.2.1.1 Phương pháp phân mảnh ngang	4
1.2.1.2 Phương pháp phân mảnh dọc	15
1.2.1.3 Phương pháp phân mảnh hỗn hợp.....	16
1.2.1. Kiểm soát toàn vẹn ngữ nghĩa phân tán	16
1.2.3 Khái niệm về cường chế và phán đoán	17
1.3 Kết luận	20
Chương 2: XỬ LÝ CÂU TRUY VẤN SQL	21
2.1. Kiểm soát toàn vẹn ngữ nghĩa phân tán	21
2.1.1 Lớp phán đoán riêng.....	21
2.1.2. Phán đoán hướng tập hợp	22
2.2. Cường chế thi hành các phán đoán toàn vẹn phân tán	24

2.2.1. Phán đoán riêng	24
2.2.2. Phán đoán hướng tập hợp	24
2.2.3. Phán đoán có hàm gộp.....	27
2.3. Các phương pháp rút gọn dựa vào phán đoán và cường chế	27
2.3.1 Rút gọn cho phân mảnh ngang nguyên thủy	27
2.3.1.1 Rút gọn với phép chọn	28
2.3.1.2. Rút gọn với phép nối	29
2.3.2. Rút gọn cho phân mảnh ngang dẫn xuất	31
2.3.3. Rút gọn cho phân mảnh dọc	35
2.3.4 Rút gọn cho phân mảnh hỗn hợp.....	37
2.4. Kết luận	41
Chương 3: ỨNG DỤNG.....	43
3.1. Ứng dụng tạo CSDL quản lý sinh viên tại Trường Cao đẳng Công nghệ Viettronics.....	43
3.2. Ứng dụng các phán đoán và cường chế thi hành câu SQL về các thông tin cần kết xuất trong CSDL trên	45
3.3. Kết luận chương 3	52
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA LUẬN VĂN.....	53
TÀI LIỆU THAM KHẢO.....	54

DANH MỤC CÁC BẢNG

Bảng 1.1. Bảng chân trị các phép toán mệnh đề.....	3
Bảng 1.2. Mức ưu tiên của các phép toán logic.....	4
Bảng 1.3. Mảnh ngang $DuAn_{H1}$	7
Bảng 1.4. Mảnh ngang $DuAn_{H2}$	7
Bảng 1.5. Mảnh ngang $DuAn_{H3}$	7
Bảng 1.6. $TraLuong_1$	10
Bảng 1.7 $TraLuong_2$	10
Bảng 1.8. Phân hoạch ngang cho quan hệ $DuAn$: $DuAn_{H1}, DuAn_{H3}, DuAn_{H4}, DuAn_{H6}$	12
Bảng 1.9. Khoa.....	13
Bảng 1.10. Sinhvien.....	13
Bảng 1.11. Monhoc.....	13
Bảng 1.12. Diem.....	14
Bảng 1.13: $Sinhvien_1$	15
Bảng 1.14: $Sinhvien_2$	15
Bảng 1.15: $Monhoc_1$	16
Bảng 1.16: $Monhoc_a$	16
Bảng 2.1. Quan hệ EMP.....	38
Bảng 2.2. Mảnh hỗn hợp EMP_{HH1}	38
Bảng 2.3. Mảnh hỗn hợp EMP_{HH2}	38
Bảng 2.4. Mảnh hỗn hợp EMP_{HH3}	39
Bảng 3.1. Khoa.....	48
Bảng 3.2. Sinhvien.....	48
Bảng 3.3. monnganh.....	48
Bảng 3.4. Lophoc.....	48
Bảng 3.5. $Sinhvien_1$	49
Bảng 3.6. $Sinhvien_2$	49
Bảng 3.7: $Monnganh_1$	49

DANH MỤC CÁC HÌNH

Hình 1.1 Biểu diễn mối liên hệ giữa các quan hệ nhờ các đường nối	5
Hình 1.2: Mối liên hệ giữa các quan hệ	14
Hình 2.1. Rút gọn cho phân mảnh ngang (với phép chọn)	29
Hình 2.2. (a) vấn tin gốc.....	31
Hình 2.2. (b) vấn tin đã rút gọn.....	31
Hình 2.3. Vấn tin gốc của ví dụ 2.10	33
Hình 2.4. Vấn tin gốc đã được giao hoán.....	34
Hình 2.5. Vấn tin sau khi dùng vị từ mâu thuẫn	34
Hình 2.6. Cây vấn tin sau khi giao hoán phép hợp và phép nối	35
Hình 2.7. Rút gọn cho phân mảnh dẫn xuất.....	35
Hình 2.8a. Phép chiếu vô dụng	36
Hình 2.8 b,c,d. Rút gọn cho phân mảnh dọc	37
Hình 2.9. Cây vấn tin gốc.....	40
Hình 2.10. Cây vấn tin đã loại EMP_{HHI}	40
Hình 2.11. Cây vấn tin đã đẩy phép chiếu xuống, phép nối lên	41
Hình 2.12. Cây vấn tin đã rút gọn	41
Hình 3.1. Sơ đồ kết nối các quan hệ.....	45
Hình 3.2. Cây truy vấn tiền xử lý của ví dụ 1	46
Hình 3.3. Các cây nối tương đương của ví dụ 1	46
Hình 3.4. Các cây nối tương đương của ví dụ 2	47
Hình 3.5. Các cây nối tương đương của ví dụ 2	47

MỞ ĐẦU

1. Đặt vấn đề

Ngày nay cùng với sự phát triển về khoa học kỹ thuật, bùng nổ về thông tin. cung cấp đa dạng các loại hình thương mại, quản lý và các dịch vụ đa phương tiện cho người sử dụng. Kết nối máy tính thành mạng với mục tiêu chia sẻ tài nguyên, khai thác có hiệu quả các tài nguyên thông tin, nâng cao khả năng tích hợp và trao đổi các loại dữ liệu giữa các thành phần trên mạng, nhu cầu thu thập, lưu trữ, xử lý và trao đổi thông tin ngày càng tăng .

Khi khối lượng thông tin phải xử lý ngày càng lớn, phong phú và đa dạng thì vấn đề đặt ra là xử lý thông tin như thế nào để tìm cách giảm thiểu thời gian và các chi phí hoặc tăng cao hiệu năng xử lý

Một trong những giải pháp có tính khả thi là phải xử lý các câu lệnh SQL khi truy vấn dữ liệu dựa vào các quy tắc phán đoán – cưỡng chế và các phương pháp rút gọn ngay trên cây thứ tự hoặc gộp nhóm khi phân mảnh mà chúng ta sẽ gọi là các giải pháp tiền xử lý.

Đó cũng chính là mục đích tôi chọn nghiên cứu “ *Xử lý câu truy vấn SQL qua phán đoán và cưỡng chế* ” làm đề tài luận văn tốt nghiệp của mình.

2. Đối tượng và phạm vi nghiên cứu

- Đối tượng nghiên cứu là cơ sở dữ liệu phân tán, xử lý song song và phân tán.
- Phạm vi nghiên cứu là một số các phương pháp xử lý câu truy vấn SQL trong cơ sở dữ liệu phân tán.

3. Hướng nghiên cứu của đề tài

- Nghiên cứu tối ưu hóa truy vấn trong cơ sở dữ liệu phân tán, các phương pháp, kỹ thuật, các thuật toán liên quan đến xử lý câu truy vấn SQL đặc biệt là phương pháp phán đoán, cưỡng chế.
- Nghiên cứu các mô hình chi phí song song và mô hình chi phí song song trên bộ tối ưu hóa truy vấn. Nghiên cứu mô hình tối ưu hóa hai pha.

4. Phương pháp nghiên cứu

- Nghiên cứu kỹ các kiến thức, chủ đề có liên quan đến đề tài.
- Nghiên cứu các quy tắc và phương pháp phán đoán và cường chế.
- Nắm vững các kiến thức về rút gọn câu văn tin dựa vào phán đoán.

5. Ý nghĩa khoa học của đề tài

Luận văn giúp cho việc tiền xử lý câu lệnh SQL bằng phương pháp phán đoán cường chế và rút gọn các câu lệnh loại này.

6. Các kết quả dự kiến đạt được

- Giới thiệu tổng quan về CSDL phân tán.
- Trình bày các phương pháp, thuật toán xử lý, tối ưu hóa truy vấn phân tán đặc biệt bằng phương pháp phán đoán và cường chế.
- Cài đặt thử nghiệm một thuật toán xử lý câu truy vấn phân tán.

Chương 1

CƠ SỞ LÝ THUYẾT

1.1. Giới thiệu về logic

1.1.1 Các loại mệnh đề

Mệnh đề sơ cấp: (elementary proposition)

Là mệnh đề không thể phân nhỏ hơn được nữa – có thể nói đó là phát biểu đơn giản nhất.

Ví dụ: “Một cộng một bằng 2”

“ 7 là số nguyên tố”

Các mệnh đề sơ cấp thường được gán với các ký hiệu như các chữ p, q, r, \dots mà ta gọi là các biến mệnh đề logic.

Mệnh đề phức hợp: (compound proposition)

Mệnh đề phức hợp là mệnh đề được tạo ra từ các mệnh đề bằng cách dùng các từ liên kết như “và” (AND), “hoặc” (OR),....

Ví dụ:

- “Số 2 là số chẵn và là số nguyên tố” gồm 2 mệnh đề: “Số 2 là số chẵn” từ nối “và” và “ Số 2 là số nguyên tố”;

Các phép toán mệnh đề

Bao gồm: Phép phủ định(\neg), phép hội(\wedge), phép tuyển(\vee), phép kéo theo($\rightarrow, \leftrightarrow$)

Bảng chân trị của các phép toán mệnh đề

Bảng 1.1. Bảng chân trị các phép toán mệnh đề

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

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 ở bảng sau:

Số hóa bởi trung tâm học liệu

<http://lrc.tnu.edu.vn/>