

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

BÙI MẠNH HÙNG

**RÚT GỌN CÂU TRUY VẤN PHÂN TÁN QUA PHÁN
ĐOÁN VÀ CƯỜNG CHẾ**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2015

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

BÙI MẠNH HÙNG

**RÚT GỌN CÂU TRUY VẤN PHÂN TÁN QUA PHÁN
ĐOÁN VÀ CƯỜNG CHẾ**

Thuộc chuyên ngành : Khoa học máy tính

Mã số chuyên ngành : 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

PGS. TS. LÊ HUY THẬP

THÁI NGUYÊN - 2015

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Qua đây em xin chân thành cảm ơn toàn thể các thầy cô trong khoa đào tạo sau đại học trường Đại học Công nghệ Thông tin và Truyền thông và đặc biệt là thầy PGS.TS. Lê Huy Thập, đã tạo điều kiện thuận lợi và hướng dẫn em để hoàn thành luận văn này.

MỤC LỤC

LỜI CAM ĐOAN	i
DANH MỤC BẢNG BIỂU	iv
DANH MỤC HÌNH	v
BẢNG DANH MỤC CÁC KÍ HIỆU	vii
LỜI 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.	1
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. Mệnh đề logic và các phép toán mệnh đề.....	3
1.1.1 Khái niệm về mệnh đề	3
1.1.2 Biến mệnh đề và biểu thức mệnh đề.....	3
1.1.3 Các phép toán mệnh đề.....	4
1.1.4. Các biểu thức logic cơ bản.....	5
1.2. Các khái niệm cơ bản của CSDL phân tán	6
1.2.1. Khái niệm về phán đoán	8
1.2.2. Khái niệm về cưỡng chế	13
1.3. Kết luận chương 1	19
Chương 2: RÚT GỌN CÂU TRUY VẤN PHÂN TÁN QUA PHÁN ĐOÁN VÀ CƯỖNG CHẾ	20
2.1. Các thuật toán rút gọn câu truy vấn	20
2.1.1. Rút gọn cho phân mảnh ngang.	21
2.1.2. Rút gọn phân mảnh dọc	32
2.1.3. Rút gọn phân mảnh hỗn hợp.....	35
2.2. Rút gọn câu truy vấn phân tán qua phán đoán và cưỡng chế	39
2.2.1. Rút gọn phân mảnh ngang qua phán đoán và cưỡng chế	40
2.2.2. Rút gọn phân mảnh dọc qua phán đoán và cưỡng chế.....	47

2.2.3. Rút gọn phân mảnh hỗn hợp qua phán đoán và cưỡng chế	49
2.3. Kết luận chương 2.....	53
Chương 3: CHƯƠNG TRÌNH ỨNG DỤNG.....	54
3.1. Chương trình ứng dụng hỗ trợ khách hàng sử dụng dịch vụ viễn thông tại Bưu điện Cầu Giấy	54
3.2. Kết luận chương 3	54
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA LUẬN VĂN.....	68

DANH MỤC BẢNG BIỂU

Bảng 1.1. Bảng chân trị các phép toán mệnh đề	5
Bảng 1.2. Mức ưu tiên của các phép toán logic	5
Bảng 2.1-1. Quan hệ NhanVien	20
Bảng 2.1-2. Quan hệ DuAn.....	20
Bảng 2.1-3. Quan hệ TraLuong.....	21
Bảng 2.1-4. Quan hệ PhanNhiem.....	21
Bảng 2.1-5. Mảnh ngang DuAn _{H1}	23
Bảng 2.1-6. Mảnh ngang DuAn _{H2}	24
Bảng 2.1-7. Mảnh ngang DuAn _{H3}	24
Bảng 2.1-8. TraLuong ₁	26
Bảng 2.1-9. TraLuong ₂	26
Bảng 2.1-10. Phân hoạch ngang cho quan hệ DuAn: DuAn _{H1} , DuAn _{H3} , DuAn _{H4} , DuAn _{H6} ...	28
Bảng 2.1-11. Khoa	28
Bảng 2.1-12. Sinhvien.....	29
Bảng 2.1-13. Monhoc.....	29
Bảng 2.1-14. Diem	29
Bảng 2.1-15: Sinhvien ₁	30
Bảng 2.1-16: Sinhvien ₂	30
Bảng 2.1-17	33
Bảng 2.1-18	33
Bảng 2.2-1. Quan hệ EMP	50
Bảng 2.2-2. Mảnh hỗn hợp EMPHH1	50
Bảng 2.2-3. Mảnh hỗn hợp EMPHH2	50
Bảng 2.2-4. Mảnh hỗn hợp EMPHH3	50

DANH MỤC HÌNH

Hình 1.1. CSDL tập trung, không phải là DDBS	7
Hình 1.2. CSDL được phân tán trên mạng, DDBS.....	7
Hình 2.1: Biểu diễn mối liên hệ giữa các quan hệ nhờ các đường nối.....	22
Hình 2.2: Mối liên hệ giữa các quan hệ.....	29
Hình 2.3. Rút gọn phân mảnh ngang với phép nối.....	34
Hình 2.4. Phép chiếu vô dụng.....	35
Hình 2.7. (b) vấn tin đã rút gọn.....	44
Hình 2.8. Vấn tin gốc của ví dụ 2.16	46
Hình 2.9. Vấn tin gốc đã được giao hoán	46
Hình 2.10. Vấn tin sau khi dùng mệnh đề mâu thuẫn.....	46
Hình 2.11. Cây vấn tin sau khi giao hoán phép hợp và phép nối	47
Hình 2.12. Rút gọn cho phân mảnh dẫn xuất.....	47
Hình 2.13 a,b,c. Rút gọn cho phân mảnh dọc	49
Hình 2.14. Cây vấn tin gốc	51
Hình 2.15. Cây vấn tin đã loại EMPHH1	52
Hình 2.16: Cây vấn tin đã đẩy phép chiếu xuống, phép nối lên	52
Hình 2.17. Câu vấn tin đã rút gọn.....	53

DANH MỤC CÁC CHỮ VIẾT TẮT

CSDL	Cơ sở dữ liệu
CPU	Central Processing Unit (<i>Bộ xử lý trung tâm</i>)
SQL	Structured Query Language (<i>Ngôn ngữ truy vấn có cấu trúc</i>)
DDBS	Distributed Database System (<i>Hệ cơ sở dữ liệu phân tán</i>)

BẢNG DANH MỤC CÁC KÍ HIỆU

\cap	Phép giao
\cup	Phép hợp
\in	Ký hiệu thuộc
\notin	Ký hiệu không thuộc
-	Phép trừ
\times	Tích đề các
\boxtimes	Phép nối
π	Phép chiếu
θ	Tê ta
*	Kết nối tự nhiên
$>$	Phép so sánh lớn hơn
$<$	Phép so sánh bé hơn
\div	Phép chia
\wedge	Phép và
\vee	Phép hoặc
\emptyset	Tập rỗng
\neg	Phủ định
$=$	Phép bằng
\geq	Lớn hơn hoặc bằng
\leq	Nhỏ hơn hoặc bằng
σ	Phép chọn
Π	Pi

LỜI 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 “ *Rút gọn 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 rút gọn 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 rút gọn 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 cưỡng chế.
- Nắm vững các kiến thức về rút gọn câu truy vấn tin dựa vào phán đoán.