

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

NGÔ HUY THẮNG

**BẢNG CÂU VẤN TIN TRÊN CÁC QUAN HỆ
VÀ XỬ LÝ CÂU VẤN TIN TRÊN BẢNG**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên, năm 2012

LỜI CẢM ƠN

Những kiến thức căn bản trong luận văn này là kết quả của quá trình tự nghiên cứu trong quá trình công tác và hai năm học Thạc sỹ (2010 - 2012) tại Trường Đại học Công nghệ thông tin và Truyền thông Thái Nguyên. Dưới sự giảng dạy, đào tạo và dìu dắt trực tiếp của các thầy cô trong trường và Viện Công nghệ thông tin Việt Nam.

Tôi xin bày tỏ lời cảm ơn chân thành tới các thầy cô trong Khoa Công nghệ thông tin, Phòng Đào tạo, Phòng Công tác học sinh sinh viên, Phòng Đào tạo sau đại học Trường Đại học Công nghệ thông tin và Truyền thông Thái Nguyên, đã tạo điều kiện thuận lợi cho tôi trong thời gian học tập tại trường.

Tôi xin bày tỏ lòng biết ơn chân thành, lời cảm ơn sâu sắc nhất đối với thầy giáo **PGS.TS Lê Huy Thập** đã trực tiếp hướng dẫn, định hướng cho tôi giải quyết các vấn đề trong luận văn.

Tôi cũng xin cảm ơn đến các anh chị đồng nghiệp ở Sở Thông tin và Truyền thông tỉnh Bắc Kạn, người thân, bạn bè và các bạn đồng môn lớp cao học CH 9A, đã ủng hộ và giúp đỡ tôi trong quá trình làm luận văn tốt nghiệp.

Thái Nguyên, ngày 15 tháng 8 năm 2012

Học viên

Ngô Huy Thắng

LỜI CAM ĐOAN

Với mục đích nghiên cứu, tìm hiểu để nâng cao kiến thức và trình độ chuyên môn để áp dụng trong các bài toán cụ thể trong tương lai nên tôi đã làm luận văn này một cách nghiêm túc và hoàn toàn trung thực. Nội dung luận văn do tự tôi tìm hiểu và hoàn thành.

Trong luận văn, tôi có sử dụng tài liệu tham khảo của một số tác giả trong và ngoài nước để hoàn thành luận văn được nêu ở phần tài liệu tham khảo.

Tôi xin cam đoan và chịu trách nhiệm về nội dung, sự trung thực trong luận văn tốt nghiệp Thạc sỹ của mình.

Thái Nguyên, Ngày 15 tháng 8 năm 2012

Học viên

Ngô Huy Thắng

MỤC LỤC

LỜI CẢM ƠN	I
LỜI CAM ĐOAN	II
MỤC LỤC	III
BẢNG CÁC KÝ HIỆU	V
BẢNG CÁC CHỮ VIẾT TẮT	VI
DANH MỤC HÌNH VẼ	VII
DANH MỤC BẢNG BIỂU	VIII
MỞ ĐẦU	10
CHƯƠNG I: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU QUAN HỆ	11
1.1. Khái quát về cơ sở dữ liệu	11
1.1.1. Những vấn đề mà CSDL cần phải giải quyết	12
1.1.2. Định nghĩa Hệ thống cơ sở dữ liệu (Database Systems)	13
1.1.3. Cơ sở dữ liệu quan hệ và lược đồ cơ sở dữ liệu quan hệ	14
1.2. Các loại câu vấn tin SQL	17
1.3. Phương pháp chuyển câu vấn tin SQL sang câu vấn tin đại số quan hệ AQL	19
1.3.1. Ngôn ngữ truy vấn đại số quan hệ (AQL)	20
1.3.2. Các phép biến đổi tương đương trong đại số quan hệ	22
1.3.3. Thuật toán chuyển câu vấn tin SQL sang câu vấn tin đại số quan hệ AQL	25
1.4. Kết luận chương 1	32
CHƯƠNG II: PHƯƠNG PHÁP TÁCH GỘP CÁC HÀNG – CỘT	33
2.1. Phương pháp chuyển câu vấn tin đại số quan hệ sang bảng	34
2.1.1. Mục tiêu của xử lý vấn tin	35
2.1.2. Mô tả đặc trưng của xử lý vấn tin	35
2.2. Định nghĩa và cách thể hiện câu truy vấn bằng bảng	37

IV

2.2.1. Cách thể hiện bảng vận tin đại số quan hệ	38
2.2.2. Độ phức tạp của phép toán đại số quan hệ	40
2.3. Phương pháp tách gộp hàng trong bảng vận tin	40
2.3.1. Kỹ thuật Gộp các hàng – cột	41
2.3.2. Kỹ thuật Tách các hàng – cột	44
2.3. Kết luận chương 2	51
CHƯƠNG III: ỨNG DỤNG GIẢI BÀI TOÁN CỤ THỂ	52
3.1. Cách tính tải trên hàng và các phân hoạch của bảng vận tin	52
3.1.1. Phân hoạch bảng vận tin.....	53
3.1.2. Cách tính tải trên hàng	53
3.2. Phương pháp tìm Cell có chi phí truyền thông lớn	56
3.3. Thuật toán tạo ra bảng vận tin tiền xử lý	58
3.4. Ví dụ minh họa	61
3.4.1. Cây tối ưu cho truy vấn dạng ống có cân bằng tải	61
3.4.2. Thuật toán chia công việc	61
3.4.3. Mở rộng thuật toán nhất cắt cục bộ cho bài toán POM	62
3.5. Kết luận chương 3	68
KẾT LUẬN	69
HƯỚNG PHÁT TRIỂN	70
TÀI LIỆU THAM KHẢO	71

BẢNG CÁC KÝ HIỆU

\cap	<i>Phép giao</i>
\cup	<i>Phép hợp</i>
\in	<i>Ký hiệu thuộc</i>
$-$	<i>Phép trừ</i>
\times	<i>Tích đề các</i>
\boxtimes	<i>Phép nối</i>
π	<i>Phép chiếu</i>
θ	<i>Tê ta</i>
$*$	<i>Kết nối tự nhiên</i>
\div	<i>Phép chia</i>
\wedge	<i>Phép và</i>
\vee	<i>Phép hoặc</i>
σ	<i>Phép chọn</i>
\subseteq	<i>Tập con</i>

BẢNG CÁC CHỮ VIẾT TẮT

SQL	Structured Query Language – <i>Ngôn ngữ truy vấn dữ liệu</i>
CSDL	<i>Cơ sở dữ liệu</i>
HQTCSDL	Database management system - <i>Hệ quản trị Cơ Sở Dữ Liệu</i>
AQL	Algebraic Query Language (<i>Ngôn ngữ vấn tin đại số</i>)
POT	Pipeline Operator Tree – <i>Cây toán tử dạng ống</i>
POM	Pipeline Operator Matrix – <i>Ma trận toán tử dạng ống</i>
QH	<i>Quan hệ</i>
QH _{<i>i</i>}	<i>Quan hệ <i>i</i>, <i>i</i> = 1, 2, 3, ...</i>
Ip	Isomorphous – <i>Ma trận đặc trưng</i>
Cell	<i>Ô</i>

DANH MỤC HÌNH VẼ

Hình 1.1. Mô hình xử lí thông tin	11
Hình 1.2. Mô hình hệ thống cơ sở dữ liệu	13
Hình 1.3. Cây đại số quan hệ ví dụ	31
Hình 2.1. Gộp hai đỉnh i và j thành đỉnh m	43
Hình 2.2. Tách hai đỉnh i và j	46
Hình 2.3. Cắt các cạnh của cây toán tử	50
Hình 2.4. Gộp các đỉnh của cây toán tử	50
Hình 3.1. Cây toán tử tổng quát	59
Hình 3.2. Cây toán tử đã được tiền xử lí	60
Hình 3.3. Thuật toán nhất cắt cục bộ	64
Hình 3.4. Cây toán tử gốc	65

VIII

DANH MỤC BẢNG BIỂU

<i>Bảng 1.1. Bảng quan hệ NHANVIEN</i>	15
<i>Bảng 1.2. Bảng quan hệ DONVI</i>	16
<i>Bảng 1.3. Bảng quan hệ DONVI_DIADIEM</i>	16
<i>Bảng 1.4. Bảng quan hệ DUAN</i>	16
<i>Bảng 1.5. Bảng quan hệ NHANVIEN_DUAN</i>	17
<i>Bảng 2.1. Bảng vận tin dạng đại số quan hệ</i>	39
<i>Bảng 2.2. Toán tử Collapse (i,j) gốc</i>	42
<i>Bảng 2.3. Toán tử Collapse (i,j) sau khi gộp i,j</i>	42
<i>Bảng 2.4. POM dữ liệu</i>	43
<i>Bảng 2.5. POM sau khi gộp</i>	44
<i>Bảng 2.6. Toán tử cut(i,j) gốc</i>	45
<i>Bảng 2.7. Toán tử cut(i,j) sau khi gộp i,j</i>	45
<i>Bảng 2.8. Dữ liệu POM</i>	46
<i>Bảng 2.9 POM₁</i>	47
<i>Bảng 2.9a. POM_{1,1}</i>	47
<i>Bảng 2.9b. POM_{1,2}</i>	47
<i>Bảng 2.10. POM₄</i>	47
<i>Bảng 2.11. Ma trận Ip truy vấn</i>	48

<i>Bảng 2.12. Ma trận Ip truy vấn mảnh F1</i>	48
<i>Bảng 2.13. Ma trận Ip truy vấn mảnh F2</i>	48
<i>Bảng 2.14. Ma trận Ip truy vấn mảnh F3</i>	49
<i>Bảng 2.15. Ma trận Ip truy vấn lớp mảnh F1</i>	49
<i>Bảng 2.16. Ma trận Ip truy vấn lớp mảnh F2</i>	49
<i>Bảng 2.17. Ma trận Ip truy vấn lớp mảnh F3</i>	49
<i>Bảng 3.1. Bảng truy vấn Ip với các phép toán đại số quan hệ</i>	54
<i>Bảng 3.2. Các t_i và $cell_{ij}$ của Ip</i>	55
<i>Bảng 3.3. Ma trận Ip tương ứng với cây toán tử gốc</i>	58
<i>Bảng 3.4. Gộp các cạnh có trọng số lớn đã qua tiền xử lí</i>	60
<i>Bảng 3.5. Ip truy vấn tương ứng với cây toán tử gốc</i>	65
<i>Bảng 3.6. Phân hoạch F1</i>	66
<i>Bảng 3.7. Phân hoạch F2</i>	66
<i>Bảng 3.8. Phân hoạch F3</i>	67
<i>Bảng 3.9. Phân hoạch F4</i>	67
<i>Bảng 3.10. Phân hoạch F5</i>	67
<i>Bảng 3.11. Phân hoạch F6</i>	68