

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

**NGUYỄN VĂN CHUNG**

**MÔ HÌNH TỐI ƯU HÓA TRUY VẤN HAI PHA  
TRONG CƠ SỞ DỮ LIỆU VÀ ỨNG DỤNG**

**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 thầy 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**

**Nguyễn Văn Chung**

## 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 09 năm 2013*

**Tác giả**

**Nguyễn Văn Chung**

## MỤC LỤC

LỜI CAM ĐOAN .....	i
LỜI CẢM ƠN .....	ii
MỤC LỤC.....	iii
DANH MỤC CÁC KÝ HIỆU, VIẾT TẮT .....	v
DANH MỤC CÁC BẢNG.....	vi
DANH MỤC CÁC HÌNH VẼ .....	vii
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. Những nội dung nghiên cứu chính.....	1
Chương 1: CƠ SỞ LÝ THUYẾT.....	3
1.1. Giới thiệu về logic.....	3
1.2. Tổng quan về CSDL phân tán.....	9
1.2.1. Không gian tìm kiếm .....	10
1.2.2. Các chiến lược tìm kiếm.....	13
1.2.3. Mô hình chi phí phân tán.....	15
1.2.4. Các dạng chi phí song song và mô hình chi phí song song trên bộ tối ưu hóa truy vấn.....	22
1.3. Kết luận chương 1 .....	25
Chương 2: MÔ HÌNH TỐI ƯU HÓA TRUY VẤN HAI PHA .....	26
2.1. Mô hình tối ưu hóa truy vấn hai pha JOQR.....	26
2.1.1. Cây truy vấn tiền xử lý .....	26
2.1.2. Cây toán tử.....	29
2.2. Tối ưu hóa giai đoạn JOQR .....	31

2.2.1. Cực tiểu hóa chi phí phân mảnh lại .....	32
2.2.2. Khả phân mảnh và toán tử cảm thuộc tính.....	34
2.2.3. Bài toán tối ưu hóa .....	37
2.3. Kết luận chương 2 .....	48
Chương 3: CHƯƠNG TRÌNH THỬ NGHIỆM.....	49
3.1. Ứng dụng tại trường Cao đẳng kinh tế - kỹ thuật Vĩnh Phúc (Dạng demo) ....	49
3.1.1. Giới thiệu CSDL của trường Cao đẳng kinh tế - kỹ thuật Vĩnh Phúc.....	49
3.1.2. Cực tiểu hóa chi phí phân mảnh lại CSDL tại mục 3.1.1 .....	62
3.2. Kết luận chương 3 .....	66
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA LUẬN VĂN.....	67
TÀI LIỆU THAM KHẢO.....	68

## DANH MỤC CÁC KÝ HIỆU, VIẾT TẮT

- DBMS (Database management system)
- ESPS (Executor Sever Process)
- JOQR (Join Ordering and Query Rewriting)
- LAN (Local Area Network)
- QEP (Query Execution Plan)
- SPJ (Selection Projection Joint)
- SQL (Structured Query Language)
- WAN (Wide area network)
- TW (Total Work)
- RT (Response Time)
- MC (Memory Consumption)

## DANH MỤC CÁC BẢNG

<i>Bảng 1-1. Bảng chân trị các phép toán mệnh đề.....</i>	<i>4</i>
<i>Bảng 1-2. Thứ tự ưu tiên của các phép toán .....</i>	<i>4</i>

## DANH MỤC CÁC HÌNH VẼ

<i>Hình 1-1. Quá trình tối ưu hoá vấn tin</i> .....	9
<i>Hình 1-2. Sơ đồ kết nối các quan hệ</i> .....	11
<i>Hình 1-3. Các cây nối tương đương</i> .....	12
<i>Hình 1-4. Các loại cây</i> .....	13
<i>Hình 1-5. Xây dựng tối ưu hoá một cách đơn định theo kiểu quy hoạch động</i> .....	14
<i>Hình 1-6. Hành động của thể tối ưu hoá trong một chiến lược ngẫu nhiên hoá</i> .....	15
<i>Hình 1-7. Truyền dữ liệu trong câu vấn tin</i> .....	17
<i>Hình 2-1. Cây truy vấn tiền xử lý</i> .....	27
<i>Hình 2-2. Cây toán tử tương ứng với cây trong hình 2-1</i> .....	31
<i>Hình 2-3. Sơ đồ phân mảnh ngang dữ liệu tại các nút</i> .....	33
<i>Hình 2-4. Các cây truy vấn khác nhau về phân hoạch dữ liệu, đường nét đứt cho thấy phải phân bố lại quan hệ</i> .....	33
<i>Hình 2-5. Cây toán tử tương ứng với câu truy vấn</i> .....	37
<i>Hình 2-6. Cây gốc và các phương án tô màu</i> .....	39
<i>Hình 2-7. Đồ thị vấn tin</i> .....	42
<i>Hình 2-8. Cây nối của đồ thị vấn tin trên hình 2-7</i> .....	43
<i>Hình 2-9. Ảnh hưởng của thứ tự phép nối đến chi phí phân mảnh ngang</i> .....	43
<i>Hình 3-1. Sơ đồ kết nối các quan hệ</i> .....	53
<i>Hình 3-2. Màn hình chính của chương trình</i> .....	54
<i>Hình 3-3. Cây truy vấn ban đầu của ví dụ 1</i> .....	55
<i>Hình 3-4. Cây sau khi sắp lại phép nối ví dụ 1</i> .....	55
<i>Hình 3-5. Màn hình nhập câu truy vấn</i> .....	56
<i>Hình 3-6. Câu truy vấn ban đầu và sau biểu diễn lại ví dụ 1</i> .....	56
<i>Hình 3-7. Kết quả của câu truy vấn ví dụ 1</i> .....	57
<i>Hình 3-8. Cây truy vấn ban đầu của ví dụ 2</i> .....	58



<i>Hình 3-9. Cây sau khi sắp lại phép nối ví dụ 2</i> .....	58
<i>Hình 3-10. Giao diện câu truy vấn ban đầu và sau biểu diễn lại ví dụ 2</i> .....	59
<i>Hình 3-11. Kết quả của câu truy vấn ví dụ 2</i> .....	59
<i>Hình 3-12. Cây truy vấn ban đầu của ví dụ 3</i> .....	60
<i>Hình 3-13. Cây sau khi xếp lại phép nối của ví dụ 3</i> .....	61
<i>Hình 3-14. Giao diện câu truy vấn ban đầu và sau biểu diễn lại ví dụ 3</i> .....	61
<i>Hình 3-15. Kết quả của câu truy vấn ví dụ 3</i> .....	62
<i>Hình 3-16. Sơ đồ phân mảnh ngang dữ liệu tại các nút của ví dụ 1</i> .....	62
<i>Hình 3-17. Cây gốc và các phương án tô màu của ví dụ 1</i> .....	63
<i>Hình 3-18. Giao diện pha 2 của ví dụ 1</i> .....	63
<i>Hình 3-19. Giao diện kết quả pha 2 của ví dụ 1</i> .....	64
<i>Hình 3-20. Sơ đồ phân mảnh ngang dữ liệu tại các nút của ví dụ 2</i> .....	64
<i>Hình 3-21. Cây gốc và các phương án tô màu của ví dụ 2</i> .....	65
<i>Hình 3-22. Giao diện pha 2 của ví dụ 2</i> .....	65
<i>Hình 3-23. Giao diện kết quả pha 2 của ví dụ 1</i> .....	66

## MỞ ĐẦU

### 1. Đặt vấn đề

Tối ưu hóa vấn tin là quá trình tìm một phương án thực hiện câu vấn tin QEP (Query Execution Plan) tối ưu (theo nghĩa hạ thấp tối đa hàm chi phí, hoặc cực đại hàm lợi ích ở một dạng nào đó). Tối ưu câu truy vấn trong cơ sở dữ liệu song song bằng mô hình tối ưu hóa truy vấn hai pha bao gồm:

- i. Sắp xếp lại thứ tự các phép nối
- ii. Biểu diễn lại cây truy vấn.

Bộ tối ưu hóa thực hiện hai bước này để tạo ra một cây truy vấn tiền xử lý, xác định những yếu tố như thứ tự thực hiện các phép toán và chiến lược thực hiện mỗi phép toán. Bộ tối ưu sẽ triển khai các mô hình và giải thuật song song để tìm kiếm một phương án tốt nhất cho việc thi hành song song.

### 2. Đối tượng và phạm vi nghiên cứu

Các biểu thức logic

Cơ sở dữ liệu phân tán

Xử lý song song và phân tán

### 3. Hướng nghiên cứu của đề tài

Các dạng chi phí song song

Nghiên cứu mô hình tối ưu hóa hai pha.

### 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.

Chương 1: Cơ sở lý thuyết

Chương 2: Mô hình tối ưu hóa truy vấn hai pha

Chương 3: Chương trình thử nghiệm

Kết luận và hướng phát triển của luận văn