

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

**PHANNOUVONG NISASON**

**RÚT GỌN CÂU TRUY VẤN VÀ ỨNG DỤNG TRONG  
CSDL PHÂN TÁN**

**LUẬN VĂN THẠC SĨ: KHOA HỌC MÁY TÍNH**

**Thái Nguyên, 2017**

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

**PHANNOUVONG NISASON**

**RÚT GỌN CÂU TRUY VẤN VÀ ỨNG DỤNG TRONG  
CSDL PHÂN TÁN**

**LUẬN VĂN THẠC SĨ: KHOA HỌC MÁY TÍNH**

*Mã số: 60 48 01 01*

**Người hướng dẫn khoa học:**

**TS. Vũ Đức Thái**

**Thái Nguyên, 2017**

## **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 xử lý. Các tài liệu tham khảo được trích dẫn và chú thích đầy đủ.

Tác giả

**PHANNOUVONG Nisason**

## LỜI CẢM ƠN

Lời đầu tiên tôi xin gửi lời cảm ơn chân thành nhất đến thầy **TS. VŨ ĐỨC THÁI** đã định hướng và nhiệt tình hướng dẫn, giúp đỡ tôi rất nhiều về mặt chuyên môn, kiến thức 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 kiến thức, 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 cha mẹ và gia đình, bạn bè, đồng nghiệp những người luôn gần gũi để động viên, chia sẻ cùng tôi trong suốt thời gian qua.

*Thái Nguyên, Tháng 4 năm 2017*

**PHANNOUVONG Nisason**

## MỤC LỤC

<b>LỜI CAM ĐOAN</b> .....	I
<b>LỜI CẢM ƠN</b> .....	IV
<b>MỤC LỤC</b> .....	V
<b>DANH MỤC CHỮ VIẾT TẮT</b> .....	VII
<b>PHẦN MỞ ĐẦU</b> .....	1
1. Đặt vấn đề.....	1
2. Đối tượng và phạm vi nghiên cứu.....	1
3. Những nội dung nghiên cứu chính.....	2
4. Phương pháp nghiên cứu.....	2
5. Ý nghĩa khoa học của đề tài .....	2
<b>CHƯƠNG 1: LÝ THUYẾT VỀ THAO TÁC TRÊN CƠ SỞ DỮ LIỆU QUAN HỆ</b> .....	3
1.1. Các phép toán logic .....	3
1.1.1. Logic hình thức .....	3
1.1.2. Logic mệnh đề và chân lý .....	3
1.2 Tổng quan về CSDL phân tán.....	7
1.2.1. Các phương pháp phân mảnh và tái thiết.....	10
1.2.2. Cách tạo cây toán tử từ câu SQL. ....	27
1.3. Kết luận .....	30
<b>CHƯƠNG 2: PHÂN RÃ VẤN TIN VÀ CỤC BỘ HÓA DỮ LIỆU</b> .....	32
2.1. Phân rã vấn tin câu vấn tin SQL . ....	32
2.1.1. Phân tích câu truy vấn trên cơ sở “kiểu dữ liệu” và “ngữ nghĩa”.....	33
2.1.2. Loại bỏ dư thừa và tối ưu câu truy vấn .....	36
2.2. Cục bộ hóa dữ liệu phân tán và các phương pháp rút gọn câu truy vấn. .	39
2.2.1. Cục bộ hóa dữ liệu phân tán.....	39
2.2.2. Rút gọn câu truy vấn SQL phân tán.....	40
2.3. Kết luận .....	64

<b>CHƯƠNG 3: TỐI ƯU TRUY VẤN PHÂN TÁN CHO CSDL TẠI TRƯỜNG CĐ KỸ THUẬT VIỆT CHĂN.....</b>	<b>65</b>
3.1. Ứng dụng trong xử lý dữ liệu học tập tại trường cao đẳng kỹ thuật Việt Chăn.....	65
3.1.1. Lý thuyết và kỹ thuật trong các câu lệnh .....	65
3.1.2. Giới thiệu CSDL của trường cao đẳng kỹ thuật Việt Chăn năm học 2015-2016.....	72
3.1.2. Một số dạng rút gọn câu vấn tin trên CSDL (tại mục 3.1.1) đã được phân mảnh. ....	74
3.2. Kết luận .....	85
<b>KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA LUẬN VĂN.....</b>	<b>86</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>87</b>

## DANH MỤC CHỮ VIẾT TẮT

CSDL: Cơ sở dữ liệu.

CĐ: Cao đẳng.

DDBMS: Distributed Database Management System.

DDBS: Distributed Database System.

LAN: Local Area Network.

SQL: Structured Query Language.

## PHẦN MỞ ĐẦU

### **1. Đặt vấn đề.**

Ngày nay các lĩnh vực như giáo dục, thương mại ... ngày càng mở rộng và phát triển. Để hoạt động được thuận lợi và có hiệu quả cần phải nắm bắt thông tin nhanh nhất khi các địa điểm ở xa nhau. Theo mô hình đám mây, các máy chủ CSDL lắp đặt tập trung tại một địa điểm an toàn, dữ liệu được phân bố trong các thiết bị lưu trữ cũng coi như là kiến trúc phân tán.

Như vậy, mô hình cơ sở dữ liệu phân tán đã, đang được nghiên cứu và ứng dụng vào thực tế. Việc cần làm là nghiên cứu và làm sáng tỏ thêm khả năng ứng dụng thực tiễn của cơ sở dữ liệu phân tán.

Lợi ích cơ bản nhất của cơ sở dữ liệu phân tán là dữ liệu của các cơ sở dữ liệu vật lý riêng biệt được tích hợp logic với nhau làm cho nhiều người sử dụng trên mạng có thể truy nhập được.

Cơ sở dữ liệu phân tán với cấp độ tự trị cao ở mỗi điểm. Trong hệ phân tán các giải pháp rút gọn câu vấn tin phân tán làm giảm thời gian tính toán, giảm khối lượng truyền thông tin giữa các trạm và giảm không gian bộ nhớ, tránh được các công việc dư thừa, loại bỏ những phần vô ích và để phục vụ cho việc xử lý song song. Được sự đồng ý của giáo viên hướng dẫn thực hiện nghiên cứu "*Rút gọn câu truy vấn và ứng dụng trong CSDL phân tán*" làm đề tài luận văn tốt nghiệp của mình.

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

- Cơ sở dữ liệu có cấu trúc và cấu trúc phân tán
- Vấn đề truy vấn trong cơ sở dữ liệu phân tán và các giải pháp tối ưu các câu truy vấn.
- Xây dựng một cơ sở dữ liệu trong thực tế minh họa cho các giải pháp
- Cài đặt mô phỏng các thao tác xử lý.



### ***3. Những nội dung nghiên cứu chính***

- Kiến trúc cơ sở dữ liệu phân tán: Các mô hình, cách tổ chức, phân mảnh CSDL theo kiến trúc phân tán.
- Xây dựng một CSDL phân tán nhỏ để thử nghiệm thuật toán
- Xây dựng một số câu truy vấn và tối ưu
- Đánh giá hiệu quả tối ưu

### ***4. Phương pháp nghiên cứu***

- Phương pháp phân tích thiết kế CSDL
- Phương pháp phân mảnh CSDL phân tán
- Phương pháp xây dựng câu truy vấn và rút gọn câu truy vấn SQL

### ***5. Ý nghĩa khoa học của đề tài***

Luận văn ứng dụng các lý thuyết về CSDL phân tán, câu truy vấn SQL vào thực tế làm rõ hơn ưu điểm của mô hình phân tán. Xây dựng bài toán ứng dụng như một CASE STUDY để nghiên cứu nâng cao kỹ năng kinh nghiệm triển khai cho người đọc.

## CHƯƠNG 1: LÝ THUYẾT VỀ THAO TÁC TRÊN CƠ SỞ DỮ LIỆU QUAN HỆ

### 1.1. Các phép toán logic

#### 1.1.1. Logic hình thức

Logic hình thức là loại logic được xác định bởi cú pháp và ngữ nghĩa như sau:

##### *Cú pháp*

Cú pháp của logic hình thức là việc sử dụng các đối tượng và các phép kết nối logic các đối tượng lại để tạo ra một biểu thức logic chuẩn (theo một nghĩa nào đó). Như vậy biểu thức logic bao gồm:

- Tập kí hiệu là các kí tự:  $a, b, c, \dots, A, B, C, \dots$
- Tập các phép toán logic: NOT, AND, OR, XOR (tương ứng với các phép toán đại số quan hệ  $\neg, \wedge, \vee, \oplus$ )
- Dùng một dãy hữu hạn các ký hiệu và các phép toán trên để tạo ra biểu thức logic.
- Cho trước một tập các quy tắc, dựa vào tập quy tắc này để tạo ra các biểu thức logic chuẩn.

##### *Ngữ nghĩa*

Ngữ nghĩa của biểu thức logic là ý nghĩa của biểu thức logic đó. Có thể dùng các phương pháp: quy nạp, đệ quy, suy luận, các quy tắc, các lược đồ chứng minh,... để chỉ ra một biểu thức logic cho trước là chuẩn hay không chuẩn và cho biết chân lý của biểu thức đó.

#### 1.1.2. Logic mệnh đề và chân lý

- Mệnh đề là một phát biểu để diễn tả một khẳng định là đúng hoặc sai, nó không thể vừa đúng lại vừa sai, hay mang tính chất mập mờ.