ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN&TRUYỀN THÔNG

LƯƠNG KIM CƯƠNG

TỐI ƯU HÓA TRUY VẤN TRONG CÁC CƠ SỞ DỮ LIỆU PHÂN TÁN

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2019

ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN&TRUYỀN THÔNG

LƯƠNG KIM CƯƠNG

TỐI ƯU HÓA TRUY VẤN TRONG CÁC CƠ SỞ DỮ LIỆU PHÂN TÁN

Chuyên ngành: Khoa học máy tính

Mã số: 8480101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS.TS. ĐOÀN VĂN BAN

Thái Nguyên - 2019

LỜI CAM ĐOAN

Tôi xin cam đoan, kết quả của luận văn hoàn toàn là kết quả của tự bản thân tôi tìm hiểu và nghiên cứu thông qua tham khảo các tài liệu và được thực hiện dưới sự hướng dẫn của PGS.TS Đoàn Văn Ban . Các tài liệu tham khảo được trích dẫn và chú thích đầy đủ.

Tác giả

Lương Kim Cương

LỜI CẢM ƠN

Lời đầu tiên, tôi xin chân thành cảm ơn PGS.TS Đoàn Văn Ban, người đã trực tiếp giảng dạy tôi trong thời gian học tập và cũng là người đã trực tiếp hướng dẫn, giúp đỡ và tạo mọi điều kiện thuận lợi cho tôi từ lúc nhận đề tài đến khi hoàn thành luận văn.

Tôi xin gửi lời cảm ơn sâu sắc đến tất cả các Thầy cô đã tham gia giảng dạy và truyền đạt kiến thức, kinh nghiệm quý báu cho chúng tôi trong hai năm học cao học tại trường Đại học Công Nghệ Thông Tin và Truyền Thông – Đại học Thái Nguyên. Những kiến thức này đã giúp tôi rất nhiều trong quá trình làm luận văn của mình.

Cuối cùng, tôi xin cảm ơn tất cả người thân, bạn bè và đồng nghiệp đã khích lệ, động viên, đóng góp ý kiến và giúp đỡ tôi hoàn thành luận văn này.

Thái Nguyên, ngày…..tháng….năm 2019

Lương Kim Cương

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

STT	Ký hiệu	Diễn giải
1	CSDL	Cơ sở dữ liệu
2	CPU	Bộ xử lý trung tâm
3	I/O	Cổng vào/ ra
4	DP	Quy hoạch động
5	ACO	Tối ưu đàn kiến

DANH MỤC CÁC HÌNH VẼ

MỤC LỤC

GIỚI THIỆU

1. Lý do chọn đề tài

Cơ sở dữ liệu phân tán đã đáp ứng một phần lớn các nhu cầu trong thực tế về dữ liệu phục vụ công tác quản lý ngày càng lớn và đa dạng. Đặc biệt, các hệ quản trị cơ sở dữ liệu phân tán đã giải quyết được vấn đề lưu trữ dữ liệu và phục vụ cho nhiều người dùng ở phân tán khắp mọi nơi.

Khi khối lượng thông tin phải xử lý ngày càng lớn, đa dạng và phong phú, dữ liệu được phân bố nhiều nơi thì vấn đề đặt ra là xử lý thông tin như thế nào để giảm chi phí đến mức tối thiểu. Một trong các giải pháp có tính khả thi là phải tối ưu hóa các câu lệnh khi truy vấn dữ liệu. Nghiên cứu về tối ưu hóa truy vấn trong cơ sở dữ liệu phân tán là cần thiết để khai thác có hiệu quả dữ liệu phân tán. Do đó, tôi chọn nghiên cứu đề tài **"Tối ưu hóa truy vấn trong các cơ sở dữ liệu phân tán"** làm luận văn tốt nghiệp của mình.

2. Mục đích nghiên cứu

Đề tài phân tích, trình bày một cách có hệ thống các nghiên cứu về cơ sở dữ liệu quan hệ, nghiên cứu các phương pháp thiết kế cơ sở dữ liệu phân tán, các kỹ thuật tối ưu hóa câu truy vấn trong cơ sở dữ liệu phân tán, cài đặt thử nghiệm một số thuật toán tối ưu hóa câu truy vấn trong cơ sở dữ liệu phân tán, từ đó đưa ra lựa chọn phù hợp với từng bài toán trên thực tế.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng và phạm vi nghiên cứu của luận văn là cơ sở dữ liệu phân tán, các câu truy vấn phân tán, một số thuật toán tối ưu hóa truy vấn phân tán và cài đặt một thuật toán tối ưu hóa truy vấn.

4. Phương pháp nghiên cứu

Nghiên cứu lý thuyết: Tìm hiểu các nghiên cứu từ các tài liệu, tạp chí và các bài viết trên mạng internet… sau đó tổng hợp so sánh để viết thành luận văn.

Nghiên cứu thực nghiệm: Cài đặt thử nghiệm thuật toán R*

5. Bố cục luận văn

Với những yêu cầu trên, nội dung của bản luận văn này trình bày khái quát chung về cơ sở dữ liệu phân tán, các kỹ thuật xây dựng cơ sở dữ liệu phân tán, tối ưu hóa truy vấn trong quá trình xử lý truy vấn, trình bày một số thuật toán tối ưu hóa truy vấn phân tán, cài đặt thuật toán R* để tối ưu hóa truy vấn. Luận văn được chia làm 3 chương:

Chương 1: Khái quát về cơ sở dữ liệu phân tán. Trong chương này trình bày khái quát về cơ sở dữ liệu phân tán: Khái niệm về cơ sở dữ liệu phân tán, ưu nhược điểm của cơ sở dữ liệu phân tán, các mức trong suốt phân tán, kiến trúc tham chiếu của cơ sở dữ liệu phân tán, các kỹ thuật xây dựng cơ sở dữ liệu phân tán.

Chương 2: Trình bày tối ưu hóa truy vấn trong cơ sở dữ liệu phân tán: Quá trình xử lý truy vấn, tối ưu hóa truy vấn dựa vào phương pháp tối ưu đàn kiến, trình bày một số thuật toán tối ưu hóa truy vấn

Chương 3: Cài đặt thuật toán tối ưu hóa truy vấn phân tán: Xác định bài toán, mô hình phân tán cơ sở dữ liệu, sử dụng thuật toán R* để áp dụng, trình bày kết quả thử nghiệm và kết luận thực nghiệm.