

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

VŨ XUÂN OANH

**NGHIÊN CỨU CÁC ĐẶC TÍNH
CỦA PHÂN MẢNH ĐỌC TRONG CSDL PHÂN TÁN
VÀ ỨNG DỤNG**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN, NĂM 2020

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

VŨ XUÂN OANH

**NGHIÊN CỨU CÁC ĐẶC TÍNH
CỦA PHÂN MẢNH ĐỌC TRONG CSDL PHÂN TÁN
VÀ ỨNG DỤNG**

Chuyên ngành: Khoa học máy tính

Mã số: 84 8 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. LÊ VĂN PHÙNG

THÁI NGUYÊN, NĂM 2020

LỜI CAM ĐOAN

Tôi xin cam đoan toàn bộ nội dung luận văn này là do tôi tự sưu tầm, tra cứu thông tin trên mạng Internet, trong một số sách tham khảo để sắp xếp, hoàn thiện cho phù hợp với nội dung yêu cầu của đề tài.

Đến nay, nội dung luận văn của tôi chưa từng được công bố hay xuất bản dưới bất kỳ hình thức nào. Nếu sai tôi xin chịu hoàn toàn trách nhiệm.

Ngày 7 tháng 11 năm 2020

Tác giả

Vũ Xuân Oanh

LỜI CẢM ƠN

Trong suốt quá trình học tập và thực hiện đề tài, em đã nhận được sự giúp đỡ tận tình và những chỉ bảo ân cần của các Thầy cô trong viện Công nghệ thông tin – Viện khoa học và công nghệ Việt nam, các Thầy cô trong trường đại học Công nghệ Thông tin và Truyền thông, cùng các bạn bè đồng nghiệp. Đặc biệt là sự giúp đỡ của ***TS Lê Văn Phùng***, người thầy trực tiếp hướng dẫn, đưa ra ý tưởng, định hướng, chỉnh sửa các kiến thức chuyên môn và tận tình giúp đỡ em trong suốt quá trình nghiên cứu và thực hiện luận văn.

Qua đây cho phép em được bày tỏ lời cảm ơn tới tất cả các thầy cô giáo ở Viện Công nghệ thông tin và Trường Đại học Công nghệ Thông tin và Truyền thông, đã giảng dạy và tạo mọi điều kiện thuận lợi giúp đỡ chúng em trong quá trình học tập, nghiên cứu.

Cuối cùng, tôi xin cảm ơn đến gia đình, các bạn bè đồng nghiệp đã chia sẻ động viên giúp đỡ tôi về chuyên môn cũng như về mọi mặt trong cuộc sống, đó là nguồn động viên khích lệ giúp tôi có nghị lực hơn để hoàn thành khoá học.

Học viên

Vũ Xuân Oanh

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC	iii
DANH MỤC CÁC BẢNG VÀ CÁC HÌNH	v
TRONG LUẬN VĂN	v
LỜI MỞ ĐẦU	1
Chương 1. TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU PHÂN TÁN VÀ PHÂN MẢNH	1
1.1. Những nét chung nhất về cơ sở dữ liệu phân tán.....	3
1.2. Vấn đề phân mảnh trong cơ sở dữ liệu phân tán	8
1.2.1. Lý do phân mảnh	8
1.2.2. Giải pháp phân mảnh	9
1.2.3. Mức độ phân mảnh	9
1.2.4. Các quy tắc phân mảnh	10
1.2.5. Các chiến lược phân mảnh.....	11
Chương 2. CÁC ĐẶC TÍNH CỦA PHÂN MẢNH DỌC	13
2.1. Định hướng heuristic để phân mảnh dọc	13
2.2. Đặc tính có nối không mất thông tin.....	14
2.2.1. Khái niệm có nối không mất thông tin	14
2.2.2. Thuật toán kiểm tra tính nối không mất thông tin	15
2.2.3. Thuật toán phân mảnh dọc có nối không mất thông tin	18
2.3. Đặc tính bảo toàn phụ thuộc	20
2.3.1. Định nghĩa phân mảnh dọc bảo toàn phụ thuộc	20
2.3.4. Thuật toán kiểm tra phân mảnh dọc có nối không mất thông tin và bảo toàn phụ thuộc	23
2.4. Phân mảnh dọc thành các BCNF, bảo toàn phụ thuộc, nối không mất thông tin	26
2.4.1. Một số mệnh đề bổ trợ	27

2.4.2. Thuật toán phân mảnh lược đồ quan hệ thành các BCNF, có nối không mất thông tin	27
2.4.3. Thuật toán phân mảnh thành các BCNF, có bảo toàn phụ thuộc .	32
2.4.4. Thuật toán phân mảnh dọc thành các BCNF, có nối không mất thông tin và bảo toàn phụ thuộc	33
Chương 3. ỨNG DỤNG THIẾT KẾ CSDL VỀ THÔNG TIN CÁC CUNG ĐƯỜNG BỘ TRÊN ĐỊA BÀN LẠNG SƠN	36
3.1. Bài toán quản lý thông tin các cung đường bộ trên địa bàn TP. Lạng Sơn...	36
3.1.1. Giới thiệu Thành phố Lạng Sơn	36
3.1.2. Hiện trạng quản lý thông tin các cung đường bộ trên địa bàn TP. Lạng Sơn	37
3.2. Thuật toán sử dụng và xác định dữ liệu đầu vào	38
3.2.1. Thuật toán sử dụng.....	38
3.2.2. Dữ liệu đầu vào	38
3.3. Môi trường thử nghiệm.....	40
3.4. Nội dung và kết quả thử nghiệm.....	41
3.4.1. Nội dung thiết kế cơ sở dữ liệu các cung đường TP. Lạng Sơn..	41
3.4.2. Phương án đề xuất phân mảnh dữ liệu (nếu có yêu cầu xây dựng CSDL phân tán)	46
3.4.3. Một số giao diện chính.....	51
3.4.4. Hướng dẫn sử dụng chương trình thử nghiệm.....	53
3.5. Đánh giá chương trình thử nghiệm	58
PHẦN KẾT LUẬN.....	62
TÀI LIỆU THAM KHẢO	63

DANH MỤC CÁC BẢNG VÀ CÁC HÌNH TRONG LUẬN VĂN

Bảng 3.1. Bảng so sánh nội dung các bước trong quy trình thiết kế CSDL mức logic	59
Hình 1.1. Minh họa về một DDBS.....	7
Hình 1.2. CSDL tập trung, không phải là DDBS.....	8
Hình 2.1. Một bảng gồm hai hàng tổng quát	17
Hình 3.1. Mô hình Thực thể- Mối quan hệ (Mô hình E_R):	45
Hình 3.2. Sơ đồ định vị của các mảnh tại các vị trí	48
Hình 3.3. Các mảnh và hình ảnh vật lý của một quan hệ tổng thể	49
Hình 3.4. Mô hình mạng của hệ thống quản lí các cung đường	50
Hình 3.5. Giao diện trang chủ	51
Hình 3.6. Giao diện nhập liệu	51
Hình 3.7. Giao diện tìm khóa	52
Hình 3.8. Giao diện phân mảnh thành hệ lược đồ đạt 3NF	52
Hình 3.9. Các bước thiết kế CSDL mức logic trong mô hình CSDL tập trung	59

LỜI MỞ ĐẦU

Cơ sở dữ liệu (CSDL) phân tán là một lĩnh vực được nghiên cứu từ lâu, nhưng gần đây do sự phát triển nhanh chóng của công nghệ truyền tin và mạng internet, cùng với xu thế toàn cầu hoá trong mọi lĩnh vực, đặc biệt trong lĩnh vực phân tán dữ liệu và thiết kế, CSDL phân tán đã trở thành một lĩnh vực thu hút nhiều sự quan tâm của các nhà nghiên cứu trong lĩnh vực CNTT.

Khái niệm phân tán (hoặc phân mảnh) ở đây được hiểu là phân tán thông tin và các thông tin đó được chứa trên các máy tính ở các vị trí khác nhau của một hệ thống máy tính có liên hệ với nhau được gọi là mạng máy tính. Việc phân mảnh sẽ làm tăng mức độ hoạt động đồng thời (song song) và như thế làm tăng lưu lượng hoạt động của hệ thống.

Việc phân mảnh được tiến thành theo 2 cách ngang và dọc. Trong đó việc chia dọc một quan hệ thành các quan hệ con chứa một tập con các thuộc tính của quan hệ gốc được gọi là phân mảnh dọc. Việc phân mảnh dọc có nhiều đặc trưng (đặc tính) quan trọng đã và đang được tập trung nghiên cứu ở trong cũng như ngoài nước cả về mặt lý luận cũng như ứng dụng thực tiễn. Đặc biệt trong lĩnh vực thiết kế, cho phép các vấn đề ảnh hưởng đến các quan hệ nhỏ hơn, vì thế giảm bớt truy xuất và tiết kiệm bộ nhớ.

Phân mảnh dọc được nghiên cứu trong ngữ cảnh của các hệ CSDL cả tập trung và phân tán. Tuy CSDL phân tán được phát triển từ CSDL tập trung nhưng nó vẫn có những ứng dụng giá trị để thiết kế CSDL tập trung.

Với mong muốn nắm vững hơn các tri thức phục vụ công tác nghiên cứu chuyên môn cũng như áp dụng vào đời sống kinh tế - xã hội cùng với sự gợi ý của Thầy hướng dẫn, tôi đã lựa chọn đề tài "***Nghiên cứu các đặc tính của phân***

mảnh dọc trong CSDL phân tán và ứng dụng" để làm luận văn tốt nghiệp thạc sĩ. Luận văn này tập trung vào việc nghiên cứu các đặc trưng của phân mảnh dọc và áp dụng để thiết kế CSDL quản lý thông tin cung đường tỉnh Lạng Sơn, nơi tôi đang sống và công tác.

Chương 1. TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU PHÂN TÁN VÀ PHÂN MẢNH

1.1. Những nét chung nhất về cơ sở dữ liệu phân tán

Cơ sở dữ liệu (CSDL) phân tán là một tập hợp các dữ liệu phục thuộc logic lẫn nhau của cùng một hệ thống và được lưu trữ trên các trạm của một mạng máy tính. CSDL phân tán làm tăng khả năng truy nhập tới CSDL lớn trên mạng. Trong hệ thống đó mỗi máy tính quản lý một CSDL thành phần được gọi là 1 node hoặc site.

Hệ quản trị CSDL phân tán (DBMS) là phần mềm quản trị CSDL, đảm bảo trong suốt đối với người sử dụng và cho phép tính tự trị nghĩa là mỗi cơ sở dữ liệu thành phần vẫn được quản trị độc lập và riêng biệt [5].

Định nghĩa này nhấn mạnh 2 khía cạnh quan trọng của CSDL phân tán:

1- *Tính phân tán:* thực tế dữ liệu không cư trú ở cùng một trạm, vì vậy chúng ta có thể phân biệt một CSDL phân tán với CSDL tập trung.

2- *Sự tương quan logic:* Các dữ liệu có một số tính chất ràng buộc lẫn nhau và như vậy có thể phân biệt CSDL phân tán với tập các CSDL địa phương hoặc với các tệp ở các trạm khác nhau trên mạng.

3 - *Các đặc trưng trong suốt của CSDL phân tán thể hiện ở chỗ:*

+ *Trong suốt phân tán:* Cho phép xử lý dữ liệu trên hệ CSDL phân tán giống như đối với CSDL tập trung. Người sử dụng (NSD) không cần biết dữ liệu đã được phân mảnh như thế nào, các bản sao dữ liệu để ở đâu, vị trí vật lý lưu trữ dữ liệu ở đâu. Trong suốt phân tán thể hiện:

+ *Trong suốt địa điểm:* NSD không cần biết lưu trữ vật lý của dữ liệu ở đâu, mà có quyền truy cập đến CSDL tại bất cứ nút nào trên mạng. Trong truy vấn chỉ cần chỉ ra tên dữ liệu mà không cần chỉ ra vị trí. Các thao tác để lấy hoặc cập nhật một dữ liệu từ xa được tự động thực hiện bởi hệ thống tại địa điểm đưa ra yêu cầu. Tính trong suốt về vị trí rất hữu ích, nó cho phép người