

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

OUTSA PHOMPHIPHAK

**MỘT SỐ KỸ THUẬT XỬ LÝ TRONG
CƠ SỞ DỮ LIỆU PHÂN TÁN VÀ ỨNG DỤNG**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái nguyên – 2015

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

OUTSA PHOMPHIPHAK

**MỘT SỐ KỸ THUẬT XỬ LÝ TRONG
CƠ SỞ DỮ LIỆU PHÂN TÁN VÀ ỨNG DỤNG**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60 48 01 01

Người hướng dẫn khoa học: PGS.TS PHẠM VIỆT BÌNH

Thái nguyên – 2015

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

**MỘT SỐ KỸ THUẬT XỬ LÝ TRONG
CƠ SỞ DỮ LIỆU PHÂN TÁN VÀ ỨNG DỤNG**

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60 48 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học

Học viên thực hiện

PGS.TS PHẠM VIỆT BÌNH

Outsa PHOMPhiphak

Thái nguyên – 2015

MỤC LỤC

DANH MỤC CÁC HÌNH ẢNH TRONG LUẬN VĂN.....	vi
DANH SÁCH CÁC BẢNG TRONG LUẬN VĂN	vii
BẢNG DANH MỤC CÁC CHỮ VIẾT TẮT.....	viii
MỞ ĐẦU.....	1
CHƯƠNG 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU VÀ CƠ SỞ DỮ LIỆU PHÂN TÁN.....	4
1.1. Khái niệm cơ sở dữ liệu	4
1.2. Nguyên tắc của hệ thống xử lý phân tán.....	5
1.2.1. Khái niệm xử lý phân tán.....	5
1.2.2. Hệ thống phân tán	5
1.3. Cơ sở dữ liệu phân tán và cơ sở dữ liệu tập trung	6
1.4. Kiến trúc Mô hình cơ sở dữ liệu phân tán.....	8
1.4.1. Điểm mạnh của cơ sở dữ liệu phân tán.....	9
1.4.2. Hệ quản trị cơ sở dữ liệu phân tán	10
1.4.3. Các mức của cơ sở dữ liệu phân tán	13
1.5. Yêu cầu về quản trị cơ sở dữ liệu phân tán	17
1.5.1. Quản lý bảng danh mục trong cơ sở dữ liệu phân tán	18
1.5.2. Nội dung của bảng danh mục.....	18
1.5.3. Phân tán bảng danh mục	19
1.5.4. Quản trị và bảo vệ.....	21
1.5.5. Áp đặt luật phân quyền	22
1.5.6. Phân lớp người sử dụng:	23
CHƯƠNG 2: MỘT SỐ KỸ THUẬT XỬ LÝ TRONG CƠ SỞ DỮ LIỆU PHÂN TÁN.....	25
2.1. Kỹ thuật phân mảnh ngang	25
2.1.1. Thuật toán phân mảnh ngang nguyên thủy	27
2.1.2. Thuật toán xác định tập vị từ đầy đủ và cực tiểu	29

2.1.3 Phân mảnh ngang dẫn xuất	31
2.2. Kỹ thuật phân Mảnh dọc	33
2.2.1. Thuật toán tụ nhóm	34
2.2.2. Thuật toán phân mảnh	38
2.3. Kỹ thuật phân mảnh hỗn hợp:	42
2.4. Kỹ thuật trộn	43
2.4.1. Thuật toán trộn tập trung	43
2.4.2. Thuật toán trộn phân tán	45
2.5. Đánh giá một số kỹ thuật xử lý trong CSDL phân tán	46
2.5.1. Đánh giá kỹ thuật phân mảnh dữ liệu:	46
2.5.2. Đánh giá kỹ thuật trộn	49
CHƯƠNG 3: XÂY DỰNG HỆ THỐNG QUẢN LÝ HÓA ĐƠN ĐIỆN TỬ NGÀNH ĐIỆN CHO CỘNG HÒA DÂN CHỦ NHÂN DÂN LÀO	50
3.1. Phát biểu bài toán	50
3.2 Các yếu tố tác động đến sự cần thiết xây dựng hệ thống	51
3.3. Phân tích nghiệp vụ các yêu cầu hệ thống	52
3.3.1. Các yêu cầu phi chức năng	52
3.3.2 Phân tích các yêu cầu chức năng của hệ thống	55
3.4. Thiết kế mô hình dữ liệu	64
3.5. Xây dựng ứng dụng và đánh giá hiệu năng	66
3.5.1. Lựa chọn các công nghệ	66
3.5.2. So sánh hiệu năng và phân tích kết quả đầu ra	67
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	69
TÀI LIỆU THAM KHẢO	70

DANH MỤC CÁC HÌNH ẢNH TRONG LUẬN VĂN

Hình 1.1: Mô hình các thành phần của hệ quản trị cơ sở dữ liệu phát triển theo kiểu thương mại (Truy cập từ xa trực tiếp).....	11
Hình 1.2: Cách thức truy cập csdl của ứng dụng	12
Hình 1.3 Cách thức truy cập gián tiếp của chương trình ứng dụng vào csdl.....	13
Hình 1.4: Sơ đồ các mức của cơ sở dữ liệu phân tán.....	14
Hình 2.1: Đồ thị kết nối.....	32
Hình 2.2 Định vị một điểm tách.....	39
Hình 2.3: Phân đoạn hỗn hợp.....	42
Hình 2.4. Tính tái thiết được của phân mảnh hỗn hợp	43
Hình 2.5: Danh sách nhân viên công ty	44
Hình 2.6 Kết quả trộn tập trung	44
Hình 2.7: Danh sách nhân viên của công ty tại các chi nhánh.....	45
Hình 2.8: Gộp nhóm theo phương pháp trộn phân tán	46
Hình 3.1: Hóa đơn điện tử.....	50
Hình 3.2: Yêu cầu về siêu liên kết	54
Hình 3.3: Sơ đồ thiết kế tổng thể	55
Hình 3.4: Cấu trúc XML của hóa đơn điện tử	57
Hình 3.5: Biểu đồ Actor.....	58
Hình 3.6: Biểu đồ Use Case	59
Hình 3.7: Biểu đồ trình tự ca sử dụng download hóa đơn điện tử.....	60
Hình 3.8: Biểu đồ trình tự ca sử dụng xác thực hóa đơn	61
Hình 3.9: Biểu đồ hoạt động ca sử dụng kí số hóa đơn	63
Hình 3.10: Mô hình logic phân tán	65
Hình 3.11: Mô hình thực thể dữ liệu.....	65
Hình 3.12: Mô hình trường dữ liệu bảng Hóa đơn	66
Hình 3.13: Bảng chỉ mục hóa đơn.....	66
Hình 3.14: Indexer database process.....	67
Hình 3.15: Biểu đồ so sánh hiệu năng	68

DANH SÁCH CÁC BẢNG TRONG LUẬN VĂN

Bảng 3.1: Danh sách các tác nhân.....	57
Bảng 3.2: Danh sách các ca sử dụng.....	59
Bảng 3.3 Mô tả biểu đồ hoạt động UC kí lên hóa đơn điện tử	62
Bảng 3.4:Mô tả biểu đồ hoạt động UC Import dữ liệu	64
Bảng 3.5: Hiệu năng tìm kiếm bản ghi được đánh chỉ mục primary cluster (khóa chính).....	67
Bảng 3.6: Tốc độ xuất báo cáo.....	68

BẢNG DANH MỤC CÁC CHỮ VIẾT TẮT

Thuật ngữ	Chi tiết	Ý nghĩa
CSDL	Cơ sở dữ liệu	
DBMS	Database management system	Hệ quản trị cơ sở dữ liệu
DB	Data base	Cơ sở dữ liệu
UC	Use case	Cả sử dụng
DC	Data Communication	Phản truyền thông dữ liệu
DD	Data Dictionary	Từ điển dữ liệu
DDB	Distributed Database	Phần cơ sở dữ liệu phân tán
DN	Doanh nghiệp	
HĐĐT	Hóa đơn điện tử	
CNTT	Công nghệ thông tin	
SSL	Secure socket layer	Giao thức mạng an toàn
PKI	Public key infrastructure	Hạ tầng khóa công khai

MỞ ĐẦU

1. Đặt vấn đề

Ngày nay, xu thế công nghệ thông tin toàn cầu cũng như sự phát triển mạng internet, cả thế giới đã được kết nối không khoảng cách. Như một tất yếu, chính phủ và các doanh nghiệp cần có những hệ thống, ứng dụng và cơ sở dữ liệu khổng lồ, chứa hàng tỉ bản ghi để phục vụ các nhu cầu đặc thù. Các bài toán mang tầm vóc quốc gia hay các hệ thống hoạt động trong thời gian dài đều có số bản ghi lên đến hàng triệu, hàng tỉ.

Khi hệ thống dữ liệu bùng nổ, các cơ chế lưu trữ và quản lý dữ liệu tập trung đã bộc lộ nhiều điểm yếu, như khả năng tìm kiếm, cập nhật dữ liệu chậm, hệ thống hay xuất hiện lỗi tràn bộ nhớ, các lỗi về tương tranh câu truy vấn khi vận hành hệ thống. Ngoài ra, sự quản lý dữ liệu tập trung rất dễ gặp rủi ro sau các sự cố và khả năng vận hành liên tục không hiệu quả.

Để khắc phục các điểm yếu của cơ chế dữ liệu tập trung, các chuyên gia đã đưa ra mô hình dữ liệu phân tán. Theo đó, mỗi một đặc thù hệ thống cần cân nhắc đến khả năng phân tán, cách thức phân tán và cường độ phân tán dữ liệu. Trong luận văn này, tôi tập trung nghiên cứu các kỹ thuật phân tán dữ liệu, tính hiệu quả của kỹ thuật đó và khả năng áp dụng bài toán đặc thù thực tế. Bài toán cụ thể ở trong luận văn là xây dựng hệ thống quản lý hóa đơn điện tử của CHDCND Lào, phân tích đặc thù hệ thống để đưa ra kỹ thuật phân tán phù hợp, đánh giá mức độ hiệu quả của kỹ thuật phân tán áp dụng trong đó

Do đó, ***“Một số kỹ thuật xử lý trong cơ sở dữ liệu phân tán và ứng dụng”*** được tôi chọn làm đề tài.

2. Đối tượng và phạm vi nghiên cứu

- Đối tượng nghiên cứu:

- ✓ Một số kỹ thuật xử lý trong cơ sở dữ liệu phân tán
- ✓ Ứng dụng quản lý hóa đơn đặc thù ngành điện

- Phạm vi nghiên cứu:

- ✓ Nguyên tắc của hệ thống phân tán

- ✓ Kiến trúc mô hình cơ sở dữ liệu phân tán
- ✓ Các kỹ thuật xử lý trong cơ sở dữ liệu phân tán
- ✓ Nghiệp vụ quản lý tiền điện, nghiệp vụ của hóa đơn điện tử, nguyên tắc bảo mật trong trao đổi hóa đơn điện tử, mô hình áp dụng và kết quả thực nghiệm

3. Hướng nghiên cứu của đề tài

- Xu thế của dữ liệu phân tán
- Các nguyên tắc thiết kế cơ bản, kiến trúc mô hình dữ liệu phân tán
- Các kỹ thuật áp dụng trong việc xử lý cơ sở dữ liệu phân tán
- Tổng quát hóa lớp bài toán áp dụng, kiến trúc có thể áp dụng trên cơ sở dữ liệu phân tán
- Đưa ra những đề xuất, thiết kế ứng dụng, phân tích ưu nhược điểm, so sánh với hệ thống không áp dụng kiến trúc cơ sở dữ liệu phân tán

4. Bố cục của luận văn

- Mở đầu
- Chương 1. Tổng quan về cơ sở dữ liệu và cơ sở dữ liệu phân tán
- Chương 2. Một số kỹ thuật xử lý trong cơ sở dữ liệu phân tán
- Chương 3. Xây dựng hệ thống quản lý hóa đơn điện tử ngành điện cho CHDC ND Lào
- Kết luận

5. Phương pháp nghiên cứu

- Nghiên cứu lý thuyết: Tìm hiểu lý thuyết về cơ sở dữ liệu phân tán trên các sách, bài báo, luận văn và các diễn đàn
- Thực nghiệm: Nghiên cứu bài toán hóa đơn điện tử, áp dụng kỹ thuật xử lý đã nghiên cứu trong lý thuyết vào bản toán cụ thể

6. Ý nghĩa khoa học của đề tài

- Ý nghĩa khoa học
- ✓ Tổng hợp, phân tích ưu nhược điểm các kỹ thuật xử lý trong cơ sở dữ liệu phân tán