

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Qua đây em xin chân thành cảm ơn toàn thể các thầy cô trong khoa đào tạo sau đại học trường Đại học Công nghệ Thông tin và Truyền thông và đặc biệt là Thầy TS.Vũ Đức Thái, đã tạo điều kiện thuận lợi và hướng dẫn em để hoàn thành luận văn này.

Thái Nguyên, tháng 12 năm 2015

Tác giả

Chanthiva MANIVONG

LỜI CẢM ƠN

Để hoàn thành chương trình cao học và viết luận văn, tôi đã nhận được sự hướng dẫn, giúp đỡ góp ý nhiệt tình của quý thầy cô trường Đại học Công nghệ Thông tin và Truyền thông – Đại học Thái Nguyên.

Tôi xin chân thành cảm ơn đến quý thầy cô trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên, quý thầy cô Viện CNTT, đặc biệt là những thầy cô đã tận tình dạy bảo cho tôi trong suốt thời gian học tập tại trường.

Tôi xin gửi lời cảm ơn sâu sắc đến TS. VŨ ĐỨC THÁI đã dành rất nhiều thời gian và tâm huyết, chỉ bảo tận tình hướng dẫn nghiên cứu và giúp tôi hoàn thành luận văn này.

Nhân đây, tôi xin chân thành cảm ơn Ban giám hiệu, Phòng Đào tạo Sau Đại học trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên đã tạo mọi điều kiện để tôi học tập và hoàn thành tốt khóa học.

Mặc dù đã cố gắng hoàn thành luận văn bằng tất cả sự nhiệt tình tâm huyết và năng lực của mình. Tuy nhiên, luận văn không thể tránh những thiếu sót, tôi rất mong nhận được những đóng góp quý báu của quý thầy cô và các bạn.

Lời cảm ơn sau cùng tôi xin dành cho gia đình và những người bạn đã hết lòng quan tâm giúp đỡ, tạo điều kiện tốt nhất để tôi hoàn thành luận văn tốt nghiệp này.

Tôi xin chân thành cảm ơn!

Thái Nguyên, tháng 12 năm 2015

Học viên thực hiện

Chanthiva MANIVONG

MỤC LỤC

LỜI CAM ĐOAN	i
MỤC LỤC.....	ii
DANH MỤC CÁC KÝ HIỆU.....	v
DANH MỤC CÁC CHỮ VIẾT TẮT	vi
DANH MỤC BẢNG BIỂU	vii
DANH MỤC HÌNH ẢNH	viii
LỜI NÓI ĐẦU	1
CHƯƠNG 1. NHỮNG VẤN ĐỀ CƠ BẢN VỀ CSDL VÀ CÂU LỆNH SQL.....	2
1.1. Cơ sở dữ liệu quan hệ	2
1.1.1. Vai trò của hệ thống thông tin	2
1.1.2 Các chức năng của một hệ quản trị CSDL	3
1.1.3. Các thành phần của một hệ QTCSDL	4
1.2. Giới thiệu ngôn ngữ SQL	4
1.3. Ngôn ngữ định nghĩa dữ liệu	5
1.4. Ngôn ngữ truy vấn dữ liệu SQL (Structured Query Language)	6
CHƯƠNG 2. CÁC GIẢI PHÁP VÀ THUẬT TOÁN TỐI ƯU CÂU LỆNH SQL	13
2.1. Tối ưu hóa câu lệnh SQL.....	13
2.2. Một số nguyên tắc trong tối ưu.....	13
2.3. Một số chiến lược tối ưu.....	15
2.4. Một số thuật toán	17
2.4.1.Thuật toán Ingres	17
2.4.2. Thuật toán INGRES- QOA	23
2.4.3. Thuật toán System	23
2.4.4. Thuật toán R-QOA	25
2.5. Phương pháp phân rã SQL.....	28
2.5.1. Nội dung của phương pháp phân rã SQL.	28
2.5.2. Những ưu và nhược điểm của phương pháp phân rã SQL	33
2.5.3. Ứng dụng của phương pháp phân rã SQL.....	34
2.6. Phương pháp thay thế SQL.....	34
2.6.1. Bản chất của vấn đề thay thế SQL.....	35

2.6.2. Những ưu và nhược điểm của của phương pháp thay thế SQL	36
2.6.3. Ứng dụng của phương pháp thay thế SQL	37
2.7. Thuật toán xử lý tiên tối ưu câu SQL	38
2.8. Một số kỹ thuật tối ưu.....	42
2.8.1. Nguyên nhân làm giảm tốc độ xử lý	42
2.8.2. Các câu truy vấn được viết nghèo nàn và một số cách khắc phục: ...	42
CHƯƠNG 3. CHƯƠNG TRÌNH THỬ NGHIỆM TỐI ƯU MỘT SỐ CÂU LỆNH	49
3.1. Mô tả bài toán	49
3.2. Thiết kế các bảng dữ liệu cho hệ thống	50
3.3 Thử nghiệm tiên tối ưu câu lệnh SQL	53
3.3.1 Xây dựng chương trình thử nghiệm.	53
3.3.2. Thực hiện thử nghiệm.....	54
KẾT LUẬN	63
TÀI LIỆU THAM KHẢO.....	64

DANH MỤC CÁC KÝ HIỆU

\cap	<i>Phép giao</i>
\cup	<i>Phép hợp</i>
\in	<i>Ký hiệu thuộc</i>
\notin	<i>Ký hiệu không thuộc</i>
$-$	<i>Phép trừ</i>
\times	<i>Tích đề các</i>
\bowtie	<i>Phép nối</i>
π	<i>Phép chiếu</i>
θ	<i>Tê ta</i>
$*$	<i>Kết nối tự nhiên</i>
$>$	<i>Phép so sánh lớn hơn</i>
$<$	<i>Phép so sánh bé hơn</i>
\div	<i>Phép chia</i>
\wedge	<i>Phép và</i>
\vee	<i>Phép hoặc</i>
\emptyset	<i>Tập rỗng</i>
\neg	<i>Phủ định</i>
$=$	<i>Phép bằng</i>
\Rightarrow	<i>Lớn hơn hoặc bằng</i>
\Leftarrow	<i>Nhỏ hơn hoặc bằng</i>
σ	<i>Phép chọn</i>
Π	<i>Pi</i>

DANH MỤC CÁC CHỮ VIẾT TẮT

- AQL Algebraic Query Language (*Ngôn ngữ vấn tin đại số*)
CSDL Cơ sở dữ liệu
CPU Central Processing Unit (*Bộ xử lý trung tâm*)
OO Object Oriented Programming (*Lập trình hướng đối tượng*)
QEP Query Execution Plan (*Chiến lược thực thi vấn tin*)
SQL Structured Query Language (*Ngôn ngữ truy vấn có cấu trúc*)
OVQP One Variable Query processor (*thể xử lý truy vấn một biến*)
ORQ Đơn quan hệ
MRQ Đa quan hệ

DANH MỤC BẢNG BIỂU

Bảng 2.1. NHÂN VIÊN	19
Bảng 2.2. PHAN NHIỆM VỤ.....	19
Bảng 2.3. Du An.....	20
Bảng 2.4. LUONG	20
Bảng 2.5. Số liệu của quan hệ DanhHieu	30
Bảng 2.6. Số liệu quan hệ ThamGiaCM	31
Bảng 2.7. Số liệu quan hệ KhenThuong	31
Bảng 2.8. Kết quả danh sách người nhận huân chương.....	36
Bảng 2.9. Số liệu của quan hệ DanhHieu	39
Bảng 2.10. Số liệu quan hệ ThamGiaCM	39
Bảng 2.11. Số liệu quan hệ KhenThưởng	40
Bảng 2.12. Kết quả sau khi thực hiện thuật toán tối ưu	42
Bảng 3.1. Sinh viên	50
Bảng 3.2 Lớp.....	51
Bảng 3.3. Môn học	51
Bảng 3.4. Điểm	52
Bảng 3.5. Điểm rèn luyện	52
Bảng 3.6. Điểm học kỳ.....	52

DANH MỤC HÌNH ẢNH

Hình 2.1. Đồ thị kết nối của câu truy vấn	27
Hình 2.2 Các thứ tự kết nối lựa chọn	27
Hình 3.1. Cơ sở dữ liệu thực nghiệm	53
Hình 3.2. Giao diện chính chương trình.....	53
Hình 3.3. Giao diện phần chèn dữ liệu tự động	54
Hình 3.4. Thực hiện truy vấn lấy danh sách điểm sinh viên đã học môn Toán cao cấp 1	55
Hình 3.5. Thực hiện truy vấn lấy danh sách điểm sinh viên sau khi tối ưu.....	57
Hình 3.6 Thực hiện truy vấn lấy thông tin lý lịch sinh viên	57
Hình 3.7. Thực hiện truy vấn lấy thông tin sinh viên sau khi tối ưu SQL.....	58

LỜI NÓI ĐẦU

Trong tình hình hiện nay, các tổ chức, doanh nghiệp có nhiều chi nhánh và đặt ở nhiều nơi với khối lượng dữ liệu lớn. Người dùng cũng phân tán nhiều địa điểm với khoảng cách không gian lớn tần suất truy cập sử dụng cơ sở dữ liệu khác nhau. Do đó hiểu và nghiên cứu một số quy trình xây dựng phần mềm thì quá trình tối ưu hóa câu lệnh của các chương trình rất có ý nghĩa vì kinh phí đầu tư của tổ chức, doanh nghiệp cho các thiết bị phần cứng bước đầu còn rất hạn chế. Để tối ưu câu lệnh trong chương trình phần mềm ta cần thử nghiệm các câu lệnh đơn lẻ đảm bảo các yêu cầu sau đó mới có thể áp dụng cho một chương trình tổng hợp. Các giải pháp, thuật toán tối ưu hóa đến từng câu lệnh cần được nghiên cứu phân tích và thử nghiệm cẩn thận.

Nếu câu vấn tin SQL (Structured Query Language) đã được tối ưu trước khi thực hiện truy vấn sẽ nâng cao hiệu quả của chương trình và tiết kiệm chi phí. Nhận thấy được mức độ cấp thiết của vấn đề và được sự gợi ý của thầy giáo hướng dẫn em đã chọn đề tài: **“Tối ưu một số câu lệnh truy vấn SQL và ứng dụng”** làm luận văn tốt nghiệp của mình.

Trong phạm vi luận văn của mình, học viên tìm hiểu câu lệnh truy vấn SQL, nghiên cứu phương pháp tối ưu câu SQL để nâng cao hiệu quả của chương trình. Trong quá trình nghiên cứu học viên đã đề xuất phương pháp phân rã và thay thế câu lệnh SQL. Nội dung luận văn gồm 03 chương:

Chương 1: Giới thiệu vấn đề cơ bản về CSDL và câu lệnh truy vấn SQL

Chương 2: Các giải pháp và thuật toán tối ưu câu lệnh SQL

Chương 3: Xây dựng chương trình thử nghiệm

Trong đó, tập trung vào các thuật toán được thử nghiệm phân tích ở chương 2. Sau khi thử nghiệm thành công có thể khuyến khích các lập trình viên sử dụng linh hoạt trong các bài toán cụ thể. Từ đó khi xây dựng các phần mềm ứng dụng các nhà phát triển có thể ứng dụng để chọn lựa sử dụng trong một chương trình cụ thể để có thể tối ưu cho nhiều đoạn mã tương tự.

CHƯƠNG 1

NHỮNG VẤN ĐỀ CƠ BẢN VỀ CSDL VÀ CÂU LỆNH SQL

1.1. Cơ sở dữ liệu quan hệ

1.1.1. Vai trò của hệ thống thông tin

Mọi hoạt động kinh doanh của một doanh nghiệp đều phải lưu trữ chứng từ, tài liệu, số liệu phục vụ công tác kinh doanh, hạch toán kinh tế, theo dõi quá trình kinh doanh đảm bảo hiệu quả. Những thông tin đó gọi là cơ sở dữ liệu kinh doanh. Trước khi có công nghệ thông tin hỗ trợ, các nhà doanh nghiệp phải lưu trữ, xử lý thông tin bằng các công cụ thủ công, hoặc bán thủ công. Khi công nghệ thông tin với máy tính điện tử phát triển hỗ trợ việc xử lý dữ liệu đóng góp to lớn vào hiệu quả kinh doanh của doanh nghiệp. Những thông tin phục vụ kinh doanh có những biến đổi quan trọng với những cấu trúc phức tạp, tinh vi, khối lượng lớn, chất lượng cao ngày càng có giá trị trong hoạt động kinh doanh.

Từ những thông tin, tài liệu thô sơ ban đầu đã hình thành nên hệ cơ sở dữ liệu bao gồm thiết bị phần cứng để lưu trữ, xử lý, truyền thông và cung cấp thông tin đến dữ liệu và các phần mềm thao tác dữ liệu với nhiều mô hình đa dạng phong phú có nhiều quy mô, nhiều lớp.

Mọi doanh nghiệp đều có chức năng kinh doanh để thực hiện những công việc sinh ra lợi nhuận. Để thực hiện chức năng kinh doanh hiệu quả doanh nghiệp cần xử lý các thông tin liên quan đến quá trình kinh doanh. Nếu tổ chức tốt thông tin việc kinh doanh của doanh nghiệp sẽ có hiệu quả, ổn định và giúp doanh nghiệp phát triển. Thông tin phục vụ cho kinh doanh chính là cơ sở dữ liệu về toàn bộ quá trình kinh doanh được lưu trữ theo một cấu trúc thích hợp, đáp ứng mọi thông tin cho việc điều hành, ra quyết định và định hướng cho doanh nghiệp.

Mô hình khái niệm có vai trò rất lớn để xây dựng nên một cơ sở dữ liệu nhất là trường hợp doanh nghiệp bắt đầu hoạt động. Nó mô tả đầy đủ chính xác thông tin cần cho doanh nghiệp, tránh được sự nhầm lẫn thiếu hụt thông tin có thể dự phòng cho sự phát triển của doanh nghiệp trong một tương lai lâu dài. Là nền tảng vững chắc cho việc cài đặt thực thi vận hành. Việc xây dựng mô hình