

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

**TÌM CÁC CÂU SQL TỪ CHƯƠNG TRÌNH NGUỒN CHUYỂN SANG
AQL VÀ VẼ CÂY TOÁN TỬ AQL**

ĐỖ QUANG VINH

THÁI NGUYÊN, 2015

LỜI CẢM ƠN

Tôi xin gửi lời cảm ơn chân thành toàn thể các thầy cô trong khoa đào tạo sau đại học trường ĐH Công nghệ thông tin và Truyền thông Thái Nguyên và đặc biệt là PGS.TS. Lê Huy Thập đã tận tình chỉ dạy và giúp đỡ để hoàn thành được luận văn này.

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của tôi, được xây dựng từ những số liệu và kết quả thực tiễn khi tôi học tập tại trường ĐH CNTT & TT Thái Nguyên và ứng dụng dựa trên số liệu của công ty IMAX, chưa từng được công bố hay xuất bản trong bất cứ công trình nào khác.

DANH MỤC CÁC CHỮ VIẾT TẮT

AQL	Algebraic Query Language (<i>Ngôn ngữ vấn tin đại số</i>)
CSDL	Cơ sở dữ liệu
CPU	Central Processing Unit (<i>Bộ xử lý trung tâm</i>)
OOP	Object Oriented Programming (<i>Lập trình hướng đối tượng</i>)
QEP	Query Execution Plan (<i>Chiến lược thực thi vấn tin</i>)
SQL	Structured Query Language (<i>Ngôn ngữ truy vấn có cấu trúc</i>)

BẢNG DANH MỤC CÁC KÍ HIỆU

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

MỤC LỤC

LỜI CẢM ƠN.....	i
LỜI CAM ĐOAN	iii
DANH MỤC CÁC CHỮ VIẾT TẮT	iv
BẢNG DANH MỤC CÁC KÍ HIỆU	v
MỤC LỤC.....	1
LỜI MỞ ĐẦU.....	3
Chương 1 : TỔNG QUAN.....	4
1.1. Giới thiệu về một số phần mềm tìm kiếm cơ bản và một số hàm tìm kiếm	4
1.1.1. Một số công cụ tìm kiếm cơ bản	5
1.1.2. Giới thiệu hàm tìm kiếm cơ bản trong các ngôn ngữ lập trình.....	7
1.2. Tổng quan về cơ sở dữ liệu phân tán.....	9
1.2.1. Định nghĩa Hệ CSDL phân tán và hệ quản trị CSDL phân tán.....	9
1.2.2. Một số dạng câu vấn tin SQL cơ bản	28
1.2.3. Định nghĩa AQL và cây toán tử.....	30
1.3 Kết luận chương.....	32
Chương 2: TÌM KIẾM CÂU SQL TỪ CHƯƠNG TRÌNH NGUỒN, CHUYỂN SANG AQL VÀ VẼ CÂY TOÁN TỬ AQL.....	33
2.1 Quá trình tối ưu hóa và một số thuật toán tối ưu vấn tin.	33
2.1.1. Quá trình tối ưu hóa.	33
2.1.2. Giới thiệu một số tối ưu hóa câu vấn tin cơ sở dữ liệu phân tán	34
2.1.2.1. Thuật toán INGRES	35
2.1.2.2. Thuật toán System R*	38
2.1.2.3. Thuật toán SDD-1	42
2.2. Thuật toán tìm câu vấn tin SQL, tạo câu vấn tin AQL và vẽ cây toán tử....	43
2.2.1. Thuật toán tìm câu vấn tin SQL từ chương trình nguồn	44

2.2.2. Thuật toán chuyển đổi từ SQL sang AQL : CH_SQL_AQL.	48
2.2.3. Thuật toán tạo cây AQL.....	50
2.2.4. Thuật toán vẽ cây toán tử.....	51
2.2.5. Ví dụ.....	53
2.3. Kết luận chương.....	54
Chương 3: CHƯƠNG TRÌNH ỨNG DỤNG.....	56
3.1. Tìm câu SQL từ chương trình quản lý hàng hóa tại công ty Imax.....	56
3.2. Chuyển lệnh SQL sang AQL.....	60
3.3. Vẽ cây toán tử AQL.....	62
KẾT LUẬN.....	63
HƯỚNG PHÁT TRIỂN CỦA ĐỀ TÀI.....	64
TÀI LIỆU THAM KHẢO.....	65

LỜI MỞ ĐẦU

Nhu cầu quản lí dữ liệu trong lĩnh vực khoa học công nghệ ngày càng cao và trở thành một thách thức lớn. Hệ thống xử lí phần cứng dần dần không theo kịp nhu cầu trên do các nguồn dữ liệu trở lên khổng lồ. Các giải pháp nhằm tăng tốc độ xử lí liên tục ra đời , trong số đó thì xử lí song song và phân tán ngày càng được quan tâm nghiên cứu và ứng dụng để giải quyết các vấn đề lớn và phức tạp.

Phạm vi nghiên cứu của luận văn này tập trung vào xử lí song song trên một chương trình nguồn có nhiều lệnh SQL có thể thỏa mãn điều kiện song song hóa và tối ưu hóa bởi phương pháp song song tự động. Việc đó tiến hành bằng cách tìm kiếm vét cạn các lệnh SQL sau đó chuyển sang AQL để tối ưu hóa vấn tin. Các vấn đề lần lượt được trình bày trong các chương như sau :

- Chương 1 : giới thiệu về một số công cụ tìm kiếm, tổng quan cơ sở dữ liệu (CSDL) phân tán, câu vấn tin SQL, toán tử Aql và cây toán tử.
- Chương 2 : trình bày các thuật toán tìm câu vấn tin SQL, thuật toán chuyển đổi câu vấn tin SQL sang AQL, thuật toán vẽ cây toán tử AQL
- Chương 3 : tạo chương trình mẫu và chạy thử dựa theo số liệu thực tế

Chương 1 : TỔNG QUAN

1.1. Giới thiệu về một số phần mềm tìm kiếm cơ bản và một số hàm tìm kiếm

Với việc phát triển càng ngày càng mạnh của mạng truyền thông và Internet, các yêu cầu về việc đáp ứng nhanh chóng khả năng lấy được thông tin hữu ích từ mạng. Các yêu cầu này đặt ra cho các nhà phát triển các hệ thống tìm kiếm nhằm đáp ứng nhu cầu tìm hiểu thông tin qua các cụm từ khóa được nhập vào.

Các công cụ tìm kiếm dựa trên chương trình tự động và các ứng dụng của điện toán đám mây nhằm lưu trữ và lấy thông tin một cách tốt nhất. Tuy nhiên một điều cơ bản đó là công cụ tìm kiếm không phải con người, chính vì thế có sự khác nhau căn bản giữa con người và công cụ tìm kiếm trong việc đánh giá nội dung của một trang web. Công cụ tìm kiếm nội dung theo định hướng, dù được trang bị công nghệ rất hiện đại nhưng công cụ tìm kiếm không đủ thông minh để có thể cảm nhận vẻ đẹp của một mẫu thiết kế, thưởng thức âm thanh hoặc thấy được sự chuyển động của một video nào đó. Công cụ tìm kiếm thu thập dữ liệu của trang web nó chỉ nhìn vào các văn bản cụ thể để cố gắng hiểu trang web này đang muốn nói về vấn đề gì.

Công cụ tìm kiếm có 3 bộ phận chính :

✓ Bộ phận thu thập dữ liệu

Bộ phận thu thập dữ liệu là một chương trình thu thập dữ liệu bằng cách đi từ trang web này sang trang web khác để khám phá nội dung và các liên kết trong trang web đó, đây là quá trình phát hiện các trang web mới, các liên kết mới, dữ liệu này sử dụng để cập nhật cho bộ phận lập chỉ mục.

✓ Bộ phận lập chỉ mục

Đây là quá trình xây dựng cơ sở dữ liệu của các từ khóa , cụm từ , các trang web và các trang liên quan đến một lĩnh vực nào đó.

✓ **Bộ phận xử lý tính toán**

Đây là quá trình tính toán nhằm cung cấp các kết quả cho người tìm kiếm. Bộ phận tính toán sẽ dựa trên các yếu tố khác nhau để phân loại và xếp hạng trang web. Các yếu tố này có tầm quan trọng khác nhau và do con người đặt ra, một trong những yếu tố quan trọng nhất để xếp hạng và phân loại là dựa trên chất lượng nội dung và chất lượng của những liên kết đến trang web của bạn.

1.1.1. Một số công cụ tìm kiếm cơ bản

- **Google**

Cơ chế hoạt động của Google :

Đầu tiên bộ phận thu thập dữ liệu của Google hay còn gọi là Google Spider xem xét trang web (định dạng HTML) nó lưu ý các từ bên trong trang web nơi nó tìm thấy các từ đó. Ví dụ các từ xuất hiện trong các thẻ tiêu đề, thẻ miêu tả... nó nhận định đây là phần quan trọng có liên quan đến việc tìm kiếm của người dùng sau này. Vì thế với mỗi website thì Google có nhiều phương pháp để index lại chỉ mục, liệt kê lại các từ khóa chính và có thể được cập nhật lại để cho hệ thống tìm kiếm diễn ra nhanh hơn, hiệu quả hơn hoặc cả hai.

Tiếp theo Google sẽ xây dựng chỉ mục, việc này làm cho các thông tin được tìm thấy một cách nhanh chóng. Sau khi tìm thông tin trên trang web, Google Spider nhận ra rằng việc tìm kiếm thông tin trên website là quá trình không bao giờ kết thúc bởi vì các nhà quản trị trang web luôn thay đổi, cập nhật thông tin trên website và điều đó có nghĩa các Spider sẽ liên tục thực hiện nhiệm vụ thu thập dữ liệu.

Cuối cùng Google sẽ xử lý, tính toán và mã hóa thông tin để lưu trữ trong cơ sở dữ liệu. Và khi có một truy vấn tìm kiếm thì hệ thống sẽ trả về các kết quả có chứa nội dung hữu ích tương ứng với các truy vấn tìm kiếm của người dùng.