

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

DƯƠNG HỒNG MINH

**NGHIÊN CỨU MỘT SỐ KỸ THUẬT QUẢN LÝ,
TRUY VẤN DỮ LIỆU LƯỢNG LIÊN TỤC**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN 2015

ố hóa bởi STrung tâm Học liệu – ĐHTN

<http://www.lrc.tnu.edu.vn>

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

DƯƠNG HỒNG MINH

**NGHIÊN CỨU MỘT SỐ KỸ THUẬT QUẢN LÝ,
TRUY VẤN DỮ LIỆU LƯỒNG LIÊN TỤC**

Mã số : 60 48 01 01

Chuyên ngành: Khoa học máy tính

LUẬN VĂN THẠC SỸ: KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. HOÀNG ĐỖ THANH TÙNG

Thái Nguyên, 11/2015

LỜI CAM ĐOAN

Tôi xin cam đoan, đây là công trình nghiên cứu của tôi. Trong đó có sự giúp đỡ tận tình của thầy giáo hướng dẫn **TS. Hoàng Đỗ Thanh Tùng**, các thầy cô tại Viện CNTT, thầy cô giáo Trường Đại học Công nghệ Thông tin và Truyền thông. Các nội dung nghiên cứu và kết quả trong đề tài này là hoàn toàn trung thực.

Trong luận văn, tôi có tham khảo tài liệu của một số tác giả đã được liệt kê tại phần Tài liệu tham khảo ở cuối luận văn.

Thái Nguyên, tháng 11 năm 2015

Tác giả

Dương Hồng Minh

LỜI CẢM ƠN

Để hoàn thành chương trình 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. Hoàng Đỗ Thanh Tùng** đã dành rất nhiều thời gian, 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 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.

Tôi xin chân thành cảm ơn!

Thái Nguyên, tháng 11 năm 2015

Học viên thực hiện

Dương Hồng Minh

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC	iii
DANH MỤC CÁC HÌNH	vi
DANH MỤC CÁC BẢNG	vii
PHẦN MỞ ĐẦU	1
Chương 1: TỔNG QUAN VỀ DỮ LIỆU LUỒNG VÀ HỆ THỐNG QUẢN LÝ DỮ LIỆU LUỒNG	3
1.1. Tổng quan về dữ liệu luồng	3
1.1.1. Giới thiệu	3
1.1.2. Sự khác biệt giữa dữ liệu luồng và cơ sở dữ liệu truyền thống	5
1.1.3. Một số ứng dụng	11
1.2. Những thách thức trong xây dựng hệ thống quản lý dữ liệu luồng	14
1.2.1. Yêu cầu bộ nhớ không hạn chế	14
1.2.2. Truy vấn xấp xỉ	17
1.2.3. Truy vấn cửa sổ trượt	17
1.2.4. Truy vấn khối	18
1.2.5. Xây dựng hệ thống quản lý dữ liệu luồng	19
Chương 2: MỘT SỐ KỸ THUẬT TRONG QUẢN LÝ VÀ TRUY VẤN DỮ LIỆU LUỒNG	22
2.1. Kỹ thuật đánh chỉ số dữ liệu luồng R-Tree	22
2.1.1. Khái niệm	23
2.1.2. Cấu trúc của một R-Tree	25
2.1.3. Thuật toán R-Tree	26
2.2. Kỹ thuật truy vấn dữ liệu luồng	35

2.2.1. Ngữ nghĩa truy vấn.....	35
2.2.2. Ngôn ngữ truy vấn	36
2.2.3. Kiến trúc truy vấn.....	37
2.2.4. Thuật toán.....	39
2.3. Phân tích lựa chọn công cụ	42
2.3.1. Microsoft StreamInsight	42
2.3.2. Ứng dụng Web thời gian thực sử dụng thư viện SignalR.....	49
Chương 3: CHƯƠNG TRÌNH THỬ NGHIỆM.....	55
3.1. Bài toán Web thời gian thực	55
3.1.1. Bài toán	55
3.1.2. Vấn đề cần giải quyết.....	55
3.1.3. Giải pháp đề xuất	56
3.2. Một số kết quả.....	59
3.2.1. Kết quả chương trình Streaming sử dụng StreamInsight.....	60
3.2.2. Kết quả ứng dụng chạy trên Web thời gian thực	63
PHẦN KẾT LUẬN	68
TÀI LIỆU THAM KHẢO.....	69

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Nghĩa tiếng Anh
DBMS	Data Base Management System
DSMS	Data Stream Management System
CPU	Central Processing Unit
SQL	Structured Query Language
CQL	Continuous Query Language
SPE	Stream Processing Engine
URL	Uniform Resource Locator
HTTP	Hypertext Transfer Protocol

DANH MỤC CÁC HÌNH

Hình 1.1: Xử lý dữ liệu trong quản trị cơ sở dữ liệu.....	6
Hình 1.2: Xử lý dữ liệu trong hệ thống dữ liệu luồng	8
Hình 2.1: Ví dụ về R-Tree.....	23
Hình 2.2: Trường hợp phân chia node (a) bad split, (b) good split	31
Hình 2.3: Phân chia phần tử thành các nhóm node mới	32
Hình 2.4: Các kiểu dữ liệu và các lớp hàm trong ngữ nghĩa trừu tượng	35
Hình 2.5: Mô hình kiến trúc xử lý truy vấn liên tục	37
Hình 2.6: Kiến trúc truy vấn liên tục trong luồng dữ liệu.....	41
Hình 2.7: Nền tảng ứng dụng StreamInsight	43
Hình 2.8: Truy vấn ràng buộc	49
Hình 2.9: Mô hình gửi thông tin của SignalR.....	51
Hình 2.10: Phương thức kết nối của SignalR	53
Hình 3.1: Kiến trúc tạo báo cáo Web sử dụng thuật toán truy vấn liên tục	55
Hình 3.2: Sơ đồ bài toán.....	57
Hình 3.3: Sơ đồ tổng thể bài toán	58
Hình 3.4: Lớp SimpleEventType	60
Hình 3.5: Các lớp trong gói SimulatedInputAdapter	60
Hình 3.6: Các lớp trong gói ConsoleOutput (hiển thị thông tin)	61
Hình 3.7: Báo cáo theo dòng.....	62
Hình 3.8: Chương trình Web.....	67
Hình 3.9: Kết quả hiển thị trên Web	67

DANH MỤC CÁC BẢNG

Bảng 1.1: So sánh sự khác nhau giữa DBMS và DSMS	10
Bảng 1.2: Ví dụ truy vấn trên dữ liệu luồng S(A, B, C) và T(D, E).....	15
Bảng 1.3: Ký hiệu một số truy vấn Q và luồng S	16

PHẦN MỞ ĐẦU

Công nghệ thông tin ngày càng phát triển, ứng dụng rộng rãi trong nhiều lĩnh vực của đời sống kinh tế xã hội. Theo thời gian, lượng dữ liệu thông tin cần lưu trữ của các cơ quan tổ chức, đơn vị, doanh nghiệp, người sử dụng... ngày càng tăng. Do dung lượng dữ liệu gia tăng không ngừng, yêu cầu ngày càng cao về hiệu năng truy xuất, tính ổn định và sự sẵn sàng của dữ liệu; việc lưu trữ đã và đang trở nên rất quan trọng. Lưu trữ dữ liệu không còn đơn giản là cung cấp các thiết bị lưu trữ dung lượng lớn mà còn bao gồm cả khả năng quản lý, chia sẻ cũng như sao lưu và phục hồi dữ liệu trong mọi trường hợp.

Mặt khác trong môi trường kinh tế đang trên đà hội nhập, phát triển và cạnh tranh người ta ngày càng cần có nhiều thông tin cần truy vấn nhanh để hỗ trợ giúp đưa ra quyết định và ngày càng có nhiều câu hỏi, vấn đề cần phải trả lời, giải quyết dựa trên một khối lượng thông tin khổng lồ. Các phương pháp lưu trữ, truy vấn cơ sở dữ liệu truyền thống ngày càng không đáp ứng được yêu cầu thực tế đề ra, đã phát triển một khuynh hướng kỹ thuật mới đó là truy vấn dữ liệu luồng liên tục (Data Streaming).

Streaming là một kỹ thuật xử lý truyền dữ liệu theo luồng một cách ổn định và liên tục. Trong mô hình luồng dữ liệu, hoạt động của một hoặc tất cả các dữ liệu đầu vào không được truy cập ngẫu nhiên từ ổ đĩa hoặc bộ nhớ mà truy cập từ một hoặc từ nhiều luồng dữ liệu liên tục. Ứng dụng trực tuyến không chỉ nhấn mạnh đến cơ chế xử lý của một hệ thống quản lý dữ liệu luồng, mà còn đưa ra những thách thức mới cho việc quản lý dữ liệu: Lớn, nhanh chóng thay đổi bộ dữ liệu cần phải được xử lý với chi phí thấp về bộ nhớ và CPU trong khi cung cấp truy cập hiệu quả và thời gian đáp ứng đảm bảo thấp. Giải quyết và khắc phục những thách thức về lưu trữ và quản lý dữ liệu, truy vấn dữ liệu nhanh từ đó hỗ trợ cá nhân, doanh nghiệp đưa ra quyết định nhanh chóng, kịp thời, hiệu quả, tăng doanh thu. Luận văn tập trung tìm