

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

NGUYỄN MINH PHÚC

**MỘT SỐ KỸ THUẬT TRUY CẬP
TRONG CƠ SỞ DỮ LIỆU TRÊN BỘ NHỚ**

Chuyên ngành: Khoa Học Máy Tính

Mã số: 60 48 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2015

LỜI CẢM ƠN

Để hoàn thành luận văn này, bên cạnh sự nỗ lực của bản thân, tôi đã nhận được rất nhiều sự giúp đỡ, động viên và hướng dẫn của các thầy cô giáo, bạn bè, đồng nghiệp và gia đình trong suốt khoá học cũng như thời gian nghiên cứu đề tài luận văn.

Tôi xin bày tỏ lòng biết ơn chân thành tới TS. Lê Quang Minh, người đã tận tình hướng dẫn và giúp đỡ tôi trong quá trình nghiên cứu và viết đề tài.

Tôi xin chân thành cảm ơn các thầy cô trong Hội đồng đã cho tôi những lời nhận xét cũng như những ý kiến đóng góp quý báu, giúp tôi hoàn thiện luận văn này.

Tôi xin cảm ơn tới các thầy cô giáo về những bài giảng thú vị và hữu ích; cảm ơn bạn bè đồng nghiệp về sự cổ vũ tinh thần lớn lao; cảm ơn gia đình đã có sự trợ giúp về mọi mặt.

Đây là một đề tài liên quan đến lĩnh vực cơ sở dữ liệu trên bộ nhớ, lĩnh vực còn khá mới mẻ ở Việt Nam. Vì vậy, luận văn không thể tránh khỏi thiếu sót và hạn chế nhất định. Tôi rất mong nhận được ý kiến đóng góp của mọi cá nhân, tổ chức quan tâm đến đề tài để đề tài được hoàn thiện hơn nữa.

Xin chân thành cảm ơn!

Thái Nguyên, 2015

MỤC LỤC

LỜI CẢM ƠN	i
MỤC LỤC.....	iii
DANH SÁCH HÌNH, ẢNH.....	iv
DANH SÁCH BẢNG	v
BẢNG DANH MỤC CÁC CHỮ VIẾT TẮT	v
MỞ ĐẦU.....	1
Chương 1. KHÁI QUÁT VỀ CƠ SỞ DỮ LIỆU THỜI GIAN THỰC VÀ VẤN ĐỀ TRUY CẬP	4
1.1. Khái quát về xử lý dữ liệu trong cơ sở dữ liệu thời gian thực	4
1.2. Vấn đề truy cập và xử lý nhanh trong hệ thống cơ sở dữ liệu	6
1.3. Giải pháp sử dụng cơ sở dữ liệu trong bộ nhớ	8
1.3.1. Giải pháp.....	8
1.3.2. Kiến trúc hệ thống cơ sở dữ liệu trong bộ nhớ	12
1.3.3. Lợi thế khi sử dụng cơ sở dữ liệu trong bộ nhớ.....	22
Chương 2. MỘT SỐ KỸ THUẬT TRUY CẬP TRONG CSDL TRÊN BỘ NHỚ.....	26
2.1. Tổ chức cấu trúc chỉ mục trong cơ sở dữ liệu trên bộ nhớ.....	26
2.2. Kỹ thuật truy cập trên T- Tree.....	30
2.3. Kỹ thuật khôi phục và kiểm soát đồng thời trong cơ sở dữ liệu trên bộ nhớ	38
2.3.1. Kỹ thuật khôi phục.....	38
2.3.2. Kiểm soát đồng thời.....	41
Chương 3: THỬ NGHIỆM CƠ SỞ DỮ LIỆU TRÊN BỘ NHỚ.....	53
3.1. Phát biểu bài toán	53
3.2. Lựa chọn giải pháp.....	53
3.2.1. Kết nối trực tiếp hay client/server?.....	54
3.2.2 Lựa chọn loại cache group sử dụng	55
3.2.3. Lựa chọn cơ chế aging	57
3.3. Lựa chọn hệ thống phần mềm để thử nghiệm.....	58
3.4. Kết quả thử nghiệm.....	62
3.4.1. Kết quả thử nghiệm cho hệ thống CustomerCare.....	62
3.4.2 Kết quả thử nghiệm cho hệ thống BCCS_Rating.....	64

KẾT LUẬN VÀ KIẾN NGHỊ.....	67
TÀI LIỆU THAM KHẢO.....	68

DANH SÁCH HÌNH, ẢNH

Hình 1. 1. Hội hợp trong tương lai.....	5
Hình 1. 2. Hệ thống thi trực tuyến.....	6
Hình 1. 3. Cấu trúc phân đoạn.....	15
Hình 1. 4. Cấu trúc phân vùng quan hệ.....	16
Hình 1. 5. Quan hệ và Thiết kế chỉ mục.....	19
Hình 1. 6. Ví dụ liên kết.....	21
Hình 2. 1. Các chỉ mục được cấu trúc cây.....	27
Hình 2. 2. Hàm băm dựa trên các chỉ mục.....	29
Hình 2. 3. T- Tree.....	31
Hình 2. 4. Giá trị giới hạn của Nút A.....	32
Hình 2. 5. Các phép tái cân bằng T- Tree.....	36
Hình 2. 6. Các phép tái cân bằng T Tree đặc biệt.....	37
Hình 2. 7. Cấu trúc khối kiểm soát khóa quan hệ.....	44
Hình 2. 8. Cấu trúc khối kiểm soát khóa giao dịch.....	49
Hình 3. 1. Mô hình CC trước và sau khi áp dụng TimesTen.....	63
Hình 3. 2. Mô hình Rating trước và sau khi áp dụng TimesTen.....	65

DANH SÁCH BẢNG

Bảng 1. 1. Bảng thống kê thời gian xử lý một bản ghi cước	59
Bảng 1. 2. Bảng lựa chọn TimesTen.....	60
Bảng 1. 3. Bảng lựa chọn tính năng TimesTen.....	61
Bảng 1. 4. Bảng so sánh response time giữa TimesTen và Oracle	63
Bảng 1. 5. Bảng so sánh thời gian xử lý khi dùng Oracle và TimesTen	65

BẢNG DANH MỤC CÁC CHỮ VIẾT TẮT

Thuật ngữ	Chi tiết	Ý nghĩa
ACID	Atomicity, Consistency, Isolation, và Durability.	Tính nguyên tử, tính nhất quán, tính tách biệt và tính bền vững.
CSDL	Cơ sở dữ liệu	
DBMS	Database Management System	Hệ quản trị cơ sở dữ liệu
IMDB	In-Memory Database	Cơ sở dữ liệu trong bộ nhớ
MMDB	Main Memory Database	Cơ sở dữ liệu trong bộ nhớ
RDBMS	Relational Database Management	Hệ(thống) quản lý cơ sở dữ liệu quan hệ

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 (CSDL) trực tuyến (online) phục vụ tức thời (real time) có thể đáp ứng hàng triệu người dùng cuối ở khắp mọi nơi. Để đáp ứng xu thế này, điện toán đám mây (cloud mputing) đang là mô hình công nghệ thông tin hứa hẹn nhất.

Khác với mô hình tính toán lưới (Grid computing) trong đó yêu cầu thực thi có được phép “đợi lúc CPU rảnh” thì Mô hình tính toán trong Clouds hoàn toàn khác, nhiều người dùng được sử dụng tài nguyên để giao dịch đồng thời và đảm bảo chất lượng dịch vụ tức thời nhưng lại phải đúng thứ tự. Hay nói cách khác, bắt buộc hệ thống phải xử lý tuần tự và đáp ứng thời gian thực cho hàng triệu giao dịch. Vì vậy, vấn đề này đang là nỗi lo lắng nhất cho các hệ thống trực tuyến đang phát triển nhanh chóng, khi mà các giải pháp xử lý song song không còn nhiều tác dụng. Ví như giao dịch ngân hàng trực tuyến, giao dịch chứng khoán, hệ thống chăm sóc khách hàng, hệ thống tính cước, mua hàng trực tuyến, kiểm soát các phương tiện giao thông...

Vì vậy, với hệ thống phải xử lý nhiều giao dịch với số lượng vô cùng lớn mà vẫn đảm bảo thời gian thực thì có 4 xu hướng cải tiến kỹ thuật cho CSDL là xử lý song song, thay đổi kiến trúc CSDL, thay đổi cơ chế hoạt động, và sử dụng bộ nhớ trong(In Memory).

Trong luận văn này tôi tập trung nghiên cứu, tìm hiểu CSDL trong bộ nhớ đảm giải quyết bài toán cho các hệ thống có số lượng giao dịch lớn đáp ứng truy vấn thời gian thực. Thử nghiệm với hệ thống Oracle TimesTen để đưa ra đánh giá đối với CSDL trong bộ nhớ.

Vì vậy, “*Một số kỹ thuật truy cập trong cơ sở dữ liệu trên bộ nhớ*” được em chọn làm đề tài.

2. Đối tượng và phạm vi nghiên cứu

- Đối tượng nghiên cứu: Một số kỹ thuật truy cập cơ sở dữ liệu trên bộ nhớ.
- Phạm vi nghiên cứu:
 - ✓ Khái quát về xử lý dữ liệu trong thời gian thực, vấn đề về truy cập và xử lý nhanh trong hệ thống cơ sở dữ liệu, giải pháp xử dụng cơ sở dữ liệu trong bộ nhớ.
 - ✓ Các tổ chức chỉ mục trong cơ sở dữ liệu trên bộ nhớ và kỹ thuật khôi phục, kiểm soát đồng thời trong cơ sở dữ liệu trên bộ nhớ.
 - ✓ Mô hình áp dụng và kết quả thực nghiệm.

3. Hướng nghiên cứu của đề tài

- Tìm hiểu xu hướng, tình hình nghiên cứu cơ sở dữ liệu trong bộ nhớ,
- Nghiên cứu các kỹ thuật được áp dụng trên Cơ sở dữ liệu trong bộ nhớ
- Nghiên cứu ưu nhược điểm của Cơ sở dữ liệu trên bộ nhớ.
- Tổng quát hóa lớp bài toán áp dụng, kiến trúc có thể áp dụng Cơ sở dữ liệu trên bộ nhớ; Kết quả thử nghiệm.
- Đưa ra những đề xuất, ý tưởng ứng dụng, cải tiến với CSDL trên bộ nhớ.

4. Bố cục của luận văn

- Mở đầu.
- Chương 1. Khái quát về cơ sở dữ liệu thời gian thực và vấn đề truy cập
- Chương 2. Một số kỹ thuật truy cập trong CSDL trên bộ nhớ
- Chương 3. Thử nghiệm cơ sở dữ liệu trên bộ nhớ
- Kết luận.

5. Phương pháp nghiên cứu

- Nghiên cứu lý thuyết: tìm hiểu lý thuyết về Cơ sở dữ liệu trên bộ nhớ trên các báo, diễn đàn nổi tiếng, các website chính thức của các công ty có sản phẩm hoặc giải pháp liên quan Cơ sở dữ liệu trên bộ nhớ.
- Thử nghiệm: tổng quát hóa loại bài toán có thể áp dụng Cơ sở dữ liệu trên bộ nhớ, chọn lựa một sản phẩm Cơ sở dữ liệu trên bộ nhớ và hệ thống để áp dụng, rút ra kết quả.

6. Ý nghĩa khoa học của đề tài

- Ý nghĩa khoa học:
 - ✓ Khái quát về Cơ sở dữ liệu trên bộ nhớ: khái niệm, mô hình kiến trúc, ưu nhược điểm.
 - ✓ Phân tích các kỹ thuật truy cập trong Cơ sở dữ liệu trên bộ nhớ.
- Ý nghĩa thực tiễn:
 - ✓ Tổng quát hóa lớp bài toán có thể áp dụng Cơ sở dữ liệu trên bộ nhớ để tăng hiệu năng.
 - ✓ Kết quả thử nghiệm

Chương 1. KHÁI QUÁT VỀ CƠ SỞ DỮ LIỆU THỜI GIAN THỰC VÀ VẤN ĐỀ TRUY CẬP

1.1. Khái quát về xử lý dữ liệu trong cơ sở dữ liệu thời gian thực

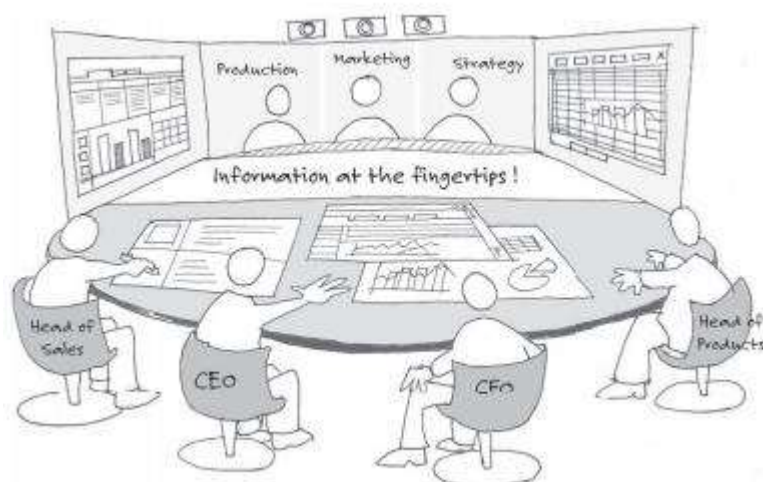
Cơ sở dữ liệu (CSDL) truyền thống đã hình thành từ rất sớm có một vai trò quan trọng trong các ứng dụng trên thực tế. Đặc điểm của các giao dịch trong CSDL truyền thống là yêu cầu về tính nhất quán. Điều này thể hiện ở trong việc lưu trữ dữ liệu thống nhất, nghĩa là tránh được tình trạng trùng lặp thông tin và có cơ chế điều khiển truy xuất dữ liệu hợp lý, đó là tránh được việc tranh chấp dữ liệu và bảo đảm dữ liệu luôn đúng đắn tại mọi thời điểm. Thông thường các thực hiện giao dịch và xử lý truy vấn trong CSDL là cố gắng nhằm đạt được một số lượng lớn giao dịch được thực thi và thời gian đáp ứng nhanh nhất có thể.

Ngày nay, với sự phát triển của công nghệ, ngoài việc đảm bảo các tính chất của giao dịch, giao dịch với số lượng lớn thì phải đảm bảo được thời gian đáp ứng của các giao dịch, đó là đảm bảo tính thời gian thực cho hệ thống. Các lĩnh vực cần đáp ứng thời gian thực bao gồm cơ sở hạ tầng mạng, viễn thông, thị trường tài chính, chứng khoán, hàng không, thương mại điện tử, doanh nghiệp... với các ứng dụng điển hình như hệ thống điều khiển không lưu, giao dịch cổ phiếu chứng khoán, đặt chỗ máy bay, du lịch, hệ thống điều khiển robot, thi trực tuyến, hội họp trực tuyến... Dưới đây là một số lĩnh vực có xu hướng giao dịch CSDL thời gian thực.

Trong lĩnh vực hoạt động doanh nghiệp, việc ứng dụng một hệ quản trị CSDL (Database Management System - DBMS) là rất quan trọng. DBMS là một hệ thống phần mềm tạo điều kiện cho quá trình xác định, xây dựng, thao tác, và chia sẻ CSDL giữa các người dùng khác nhau và các ứng dụng. Nó cũng cố tất cả các hoạt động trong một ứng dụng doanh nghiệp và hiệu suất của các ứng dụng doanh nghiệp phụ thuộc nhiều vào hiệu suất của các DBMS. Khi hệ quản trị CSDL quan hệ ra đời đã hỗ trợ giao dịch một cách hiệu quả và rất tin cậy. Nhưng khi khối lượng dữ liệu lớn hơn, hệ quản trị CSDL quan hệ không còn có thể hiệu quả phục vụ các yêu cầu của các ứng dụng doanh nghiệp. Đặc biệt nó thể xử lý

giao dịch trên toàn bộ CSDL một cách kịp thời. Lý do chính mà hệ quản trị CSDL quan hệ hiện tại không thể thực hiện các truy vấn yêu cầu đủ nhanh là dữ liệu phải được lấy ra từ đĩa cứng. Hệ thống hiện đại sử dụng rộng rãi của bộ nhớ đệm để lưu trữ dữ liệu thường xuyên truy cập trong bộ nhớ chính nhưng đối với các truy vấn xử lý lượng lớn dữ liệu, đọc ổ đĩa vẫn được yêu cầu. Đơn giản chỉ cần truy cập và đọc dữ liệu từ đĩa cứng có thể mất một số lượng đáng kể thời gian. Vì vậy đối với doanh nghiệp đòi hỏi phải có phương án thay thế để giảm các chi phí và thời gian xử lý của hệ thống quản lý dữ liệu trên đĩa là rất cần thiết.

Hình 1.1 là một ví dụ về đề xuất của Bill Gates vào năm 1994, Information at the fingertips, khi đó ông hình dung ra một tương lai mà trong đó bất kỳ thông tin nào đều có sẵn ở bất cứ đâu. Việc tham gia cuộc họp có thể ở nhiều địa điểm khác nhau, tất cả các trình duyệt, truy vấn, thao tác cùng thông tin đều trong thời gian thực. (<http://sapinsider.wispubs.com/Assets/Blogs/2010/June/Hasso-Plattner-on-In-Memory-Computing-OLAP-and-OLTP>).



Hình 1. 1. Hội họp trong tương lai

Trong lĩnh vực tài chính, những cán bộ kiểm soát tài chính là những người đặc biệt gặp khó khăn với khối lượng dữ liệu ngày càng gia tăng . Do thời gian trả lời dữ liệu chậm, họ bị buộc phải giới hạn khung thời gian phân tích trong vài ngày thay vì vài tháng hay vài quý . Điều này dẫn tới một loạt chậm trễ, bao gồm cả việc đóng