

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

NGUYỄN VĂN TÙNG

**NGHIÊN CỨU CƠ SỞ DỮ LIỆU TRONG BỘ NHỚ
GIẢI QUYẾT BÀI TOÁN ĐA GIAO DỊCH**

LUẬN VĂN THẠC SỸ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2012

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

NGUYỄN VĂN TÙNG

**NGHIÊN CỨU CƠ SỞ DỮ LIỆU TRONG BỘ NHỚ
GIẢI QUYẾT BÀI TOÁN ĐA GIAO DỊCH**

Chuyên ngành: Khoa học máy tính

Mã số: 60.48.01

LUẬN VĂN THẠC SỸ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. Hoàng Đỗ Thanh Tùng

Thái Nguyên - 2012

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn “Nghiên cứu Cơ sở dữ liệu trong bộ nhớ giải quyết bài toán đa giao dịch” này là công trình nghiên cứu của riêng tôi. Luận văn này do tôi tự tìm hiểu, thử nghiệm và thu thập từ nhiều nguồn dưới sự hướng dẫn chỉ bảo tận tâm của thầy giáo TS. Hoàng Đỗ Thanh Tùng. Các số liệu trong luận văn là trung thực với thử nghiệm. Luận văn này chưa từng được công bố trước đây.

Học viên

Nguyễn Văn Tùng

MỤC LỤC

MỤC LỤC.....	i
LỜI CẢM ƠN	iii
BẢNG DANH MỤC CÁC CHỮ VIẾT TẮT.....	iv
DANH MỤC CÁC HÌNH VẼ.....	v
MỞ ĐẦU.....	1
Chương 1. GIẢI PHÁP THỜI GIAN THỰC CHO GIAO DỊCH CƠ SỞ DỮ LIỆU	2
1.1. Xu hướng giao dịch cơ sở dữ liệu thời gian thực	2
1.2. Khái quát về giao dịch cơ sở dữ liệu	4
1.2.1. Khái niệm	4
1.2.2. Trạng thái của giao dịch	6
1.3. Bài toán đa giao dịch trong ứng dụng trực tuyến	7
1.4. Giải pháp sử dụng cơ sở dữ liệu trong bộ nhớ.....	8
1.5. Lợi thế khi sử dụng cơ sở dữ liệu trong bộ nhớ.....	12
1.6. Kết luận chương 1.....	14
Chương 2. CƠ SỞ DỮ LIỆU TRONG BỘ NHỚ	15
2.1. Kiến trúc hệ thống cơ sở dữ liệu trong bộ nhớ	15
2.1.1. Lợi thế của đánh địa chỉ trực tiếp và thường trú bộ nhớ	15
2.1.2. Các vấn đề với đánh địa chỉ trực tiếp	16
2.1.3. Tổ chức dữ liệu thường trú bộ nhớ.....	17
2.2. Các cấu trúc chỉ mục.....	25
2.2.1. Các cấu trúc chỉ mục đã có.....	25
2.2.2. T Tree	29
2.3. Xử lý truy vấn	36
2.4. Khôi phục.....	37
2.5. Kiểm soát đồng thời.....	41
2.5.1. Cấu trúc khóa.....	42
2.5.2. Các khóa quan hệ.....	44
2.5.3. Các khóa bản ghi	44

2.5.4. Các khóa vị từ.....	45
2.5.5. Các khóa thành phần chỉ mục.....	46
2.6. Kết luận chương 2.....	46
Chương 3. THỬ NGHIỆM VÀ ĐÁNH GIÁ.....	48
3.1. Cơ sở dữ liệu trong bộ nhớ của Oracle TimesTen.....	48
3.1.1. So sánh TimesTen và các CSDL thông thường.....	48
3.1.2. Kiến trúc của TimesTen.....	50
3.1.3. Ứng dụng kết nối đến Data Store.....	52
3.1.4. Data Availability và Integrity.....	53
3.2. Thử nghiệm và đánh giá.....	60
3.2.1. Môi trường thử nghiệm.....	60
3.2.2. Thử nghiệm.....	62
3.2.3. Đánh giá.....	65
3.3. Kết luận chương 3.....	66
KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU.....	67
TÀI LIỆU THAM KHẢO.....	68

LỜI CẢM ƠN

Lời đầu tiên, tôi xin chân thành cảm ơn Ban giám hiệu và các thầy cô giáo Trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên, Viện Công nghệ thông tin - Viện Khoa học và Công nghệ Việt Nam vì những hỗ trợ, hướng dẫn và các bài giảng hết sức ý nghĩa, quan trọng trong suốt quá trình tôi được đào tạo tại Trường.

Tôi xin bày tỏ lòng biết ơn sâu sắc đến thầy giáo TS. Hoàng Đỗ Thanh Tùng đã tận tình hướng dẫn và định hướng tôi trong nghiên cứu, đưa ra những chỉ bảo cặn kẽ trong quá trình thực hiện luận văn này.

Lời cảm ơn chân thành của tôi xin được gửi tới Lãnh đạo và đồng nghiệp của tôi tại trường Trung cấp nghề Công nghiệp tàu thủy Phà Rừng đã tạo điều kiện thuận lợi, hỗ trợ tôi giải quyết công việc để tôi có thể theo học chương trình này. Đồng cảm ơn các bạn bè trong lớp Cao học CK09C đã cùng thảo luận, cung cấp cho tôi tài liệu, tư liệu rất quý giá để xây dựng nội dung luận văn và học tập.

Tôi xin cảm ơn sự giúp đỡ, động viên, chia sẻ của gia đình, bạn bè đã dành cho tôi trong quá trình học tập và nghiên cứu.

Học viên

Nguyễn Văn Tùng

BẢNG DANH MỤC CÁC CHỮ VIẾT TẮT

Thuật ngữ	Chi tiết	Ý nghĩa
ACID	Atomicity, Consistency, Isolation, Durability	Tính nguyên tử, tính nhất quán, tính cô lập, tính bền
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 System	Hệ quản trị cơ sở dữ liệu quan hệ

DANH MỤC CÁC HÌNH VẼ

TT	Tên hình	Trang
1	Hình 1.1. Hội họp trong tương lai	3
2	Hình 1.2. Hệ thống thi trực tuyến	4
3	Hình 1.3. Các bước thực hiện giao dịch chuyển khoản từ tài khoản A sang tài khoản B	5
4	Hình 1.4. Trạng thái của giao dịch	6
5	Hình 2.1. Cấu trúc phân đoạn	18
6	Hình 2.2. Cấu trúc phân vùng quan hệ	19
7	Hình 2.3. Quan hệ và Thiết kế chỉ mục	21
8	Hình 2.4. Ví dụ các liên kết	23
9	Hình 2.5. Các chỉ mục được cấu trúc cây	26
10	Hình 2.6. Hàm băm dựa trên các chỉ mục	28
11	Hình 2.7. T Tree	30
12	Hình 2.8. Giá trị giới hạn của Nút A	31
13	Hình 2.9. Các phép tái cân bằng T Tree	35
14	Hình 2.10. Các phép tái cân bằng T Tree đặc biệt	36
15	Hình 2.11. Cấu trúc khối kiểm soát khóa quan hệ	43
16	Hình 2.12. Cấu trúc khối kiểm soát khóa giao dịch	44
17	Hình 3.1: So sánh giữa các công cụ disk-based RDBMS và TimesTen	49
18	Hình 3.2. Các thành phần trong TimesTen	50
19	Hình 3.3. TimesTen Data Store	51
20	Hình 3.4. TimesTen Data Manager	51
21	Hình 3.5. TimesTen ODBC và JDBC APIs	52

TT	Tên hình	Trang
22	Hình 3.6. Two-tiered logging	53
23	Hình 3.7. Các hoạt động trong Checkpoint	55
24	Hình 3.8. Khôi phục dữ liệu từ log và checkpoint files	56
25	Hình 3.9. Master và Subscriber Data Store	56
26	Hình 3.10. Master và Subscriber Replication Agents	57
27	Hình 3.11. Ví dụ về các cấu hình replication	58
28	Hình 3.12. Replication trong Hot standby và load balancing	59
29	Hình 3.13. Active standby pair	59
30	Hình 3.14. Insert dữ liệu vào TimesTen	63
31	Hình 3.15. Insert dữ liệu vào MS SQL Server 2008 trên ổ cứng SSD	63
32	Hình 3.16. Insert dữ liệu vào MS SQL Server 2008 trên ổ cứng HDD	64
33	Hình 3.17. Dữ liệu trong TimesTen	64
34	Hình 3.18. Dữ liệu trong MS SQL Server 2008	65
35	Hình 3.19. Biểu đồ so sánh thời gian insert 1 triệu bản ghi	66

MỞ ĐẦU

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, các 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 computing) 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 thể 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 dụ như giao dịch ngân hàng trực tuyến, giao dịch chứng khoán, mua hàng trực tuyến, kiểm soát các phương tiện giao thông...

Vì vậy, với hệ thống phải xử lý đa 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 nhằm 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.

Luận văn được chia làm 3 chương:

Chương 1: Giải pháp thời gian thực cho giao dịch cơ sở dữ liệu

Chương 2: Cơ sở dữ liệu trong bộ nhớ

Chương 3: Thử nghiệm và đánh giá