

# 浙江工业大学

## 本科毕业设计文献综述

(2013 届)



论文题目 基于内存数据库的大数据应用系统  
的设计与实现

作者姓名 陈佳鹏

指导教师 陈 波

学科(专业) 软件工程

所在学院 计算机科学与技术学院

提交日期 2013 年 06 月

## 基于内存数据库的大数据应用系统的设计与实现

**摘要：** 本文是基于内存数据库的大数据应用系统的设计与实现的一篇文献综述，先介绍项目的由来及其研究意思，然后介绍项目的国内外研究现状及难点以定位项目开发的一个大环境，明确当前同类项目的研究情况。接着本文简述内存数据库系统的结构，紧接着介绍系统开发中所需的关键技术。

**关键字：** 内存数据库，索引结构，并发控制，T 树，数据恢复，影子内存，混合日志

### 一、引言

在信息化建设中用于对数据进行维护和管理数据库扮演了一个十分重要的角色，一个合适的数据库系统往往是信息化改造成功与否的关键所在。传统的磁盘数据库(Disk Resident Database, DRDB)由于经过几十年的发展其功能完备，稳定性好，因而经常应用于电信，银行等公司进行客户资料管理，认识部门档案管理等方面，并且一直以来有着令人满意的表现。然而伴随着科技的发展，涌现出一批新兴行业或传统行业需要新的应用，例如工业控制，数据通信，证券交易，电力调度，航空航天等。这些新的产业所要求的数据库系统通常并不需要数据库具有强大而完备的功能和进行非常复杂的事物处理的能力，却需要数据库能在指定的时刻或时间内对大量的数据进行采集，处理并能正确的响应的高速的性能。由于“I/O 瓶颈”问题，基于磁盘的数据库系统(Oracle, SQL Server 等)不能满足现代应用对数据库的实耐性处理的要求。内存数据与磁盘数据在访问时间上相差 5 个数量级，内存的速度具有明显的优势。内存容量增大，而价格却在不断地下降。从 20 世纪 80 年代开始，数据库研究人员考虑把整个或者大部分数据库放在内存中。内存数据库(MMDB)是实时数据库研究的基础，并成为了研究的热点。

MMDB 与 DRDB 最主要的区别就是数据主版本的驻留位置不同，前者驻留在内存，后者驻留在磁盘。由于内存与磁盘在访问速度、易失性、访问模式等方面存在很大的差异，内存数据库与磁盘数据库中的数据是常驻内存，处理前不需要从磁盘读取数据，数据库也存在差别。如表 1.1 所示。

### 二、研究意义

内存数据库因为其快速的数据访问能力，使其能比磁盘数据库(DRDB)更适合于需要快速响应和高事务吞吐量的应用环境。对于那些需要在严格要求的时间段内完成事务请求的实时应用系统，和需要支持大数据量并发访问的高性能事务处理平台来讲，内存数据库都是一个理想的选择。此外，在实际生产中，常常出现不能互相访问其内置实时数据库的信息，从而使大量信息冗余重复存在于各系统中，也就是出现数据孤岛。为了解决这个问题，必须对实时数据库的数据管理

表 1.1: MMDB 和 DRDB 的比较

性能	MMDB	DRDB
数据存储	行、列级储存模型以及段-分区式储存模型, 不压球模型在内存中连续存放	Sybase IQ 采用列级储存, 其他数据库系统采用行级存储, 在磁盘上连续存放
缓冲管理	不需要	需要
并发控制	采用封锁机制、多版本等方式, 一般采用较大粒度的锁, 如库级锁、表级锁或采用乐观封锁机制	采用封锁机制、时间戳、多版本等方式, 为了提高事务的并发度, 一般支持多粒度和多种类型的锁
恢复机制	备份、日志和检查点技术, 采用预提交、组提交等提交方式; 用稳定内存来存储日志记录	采用备份、日志、检查点、保存点等技术
索引结构	T 树索引、hash 索引	B 树索引、hash 索引
查询优化	基于 CPU 代价以及 cache 代价	基于 I/O 代价

进行合理规划以建立开放的实时数据库系统, 使之能够提供高速、及时的实时数据服务。

### 三、国内外研究现状

内存数据库的研究始于 20 世纪 80 年代, 并逐渐吸引了越来越多的研究者的研兴趣。经过 20 多年的发展, 时至今日研究者们已对它的体系结构, 数据组织与存取方法, 事务处理, 并发控制和恢复备份技术进行了大量的探讨和研究, 针对硬件, 软件和算法设计提出了许多不同的策略和实现方案, 取得了丰硕的成果。目前内存数据库的研究主要集中在一下几个方向:

1. 内存数据库的体系结构, 内存数据库的体系结构包括系统的习题结构和存储体系结构两部分内容, 系统体系结构方面主要侧重于研究多处理器在 MMDB 中的应用, 而存储体系结构方面则侧重于研究非易失性内存(NV-RAM, UPS RAM 等)在 MMDB 中的应用。

2. 事务处理, 事务的处理主要对事务的提交及与之相联的日志记录、查询优化, 尤其是联机查询的优化进行了多方面的研究, 开发了一些有效的技术。如开发了“提前提交”等策略以优化事务的初始如对事务的初始处理、提交处理、并发控制、完整性和安全性检验等, 加速并发事务的响应时间, 查询的优化主要在于对减少查询中中间关系的算法的研究。

3. 数据组织与存取方法, 这方面主要针对 MMDB 存储介质的特性设计了许多适合内存特性的数据存储组织结构和存取方式, 索引结构和存储策略, MMDB 的压缩, 如 AVL 树索引结构, 区段式组织结构, 位图分配法等。

4. MMDB 的备份和恢复, 备份方面更是提出了多种方案, 为了预防 MMDB 的崩溃主要是结合检查点和日志来保证 MMDB(MMDB)在崩溃后的可恢复性, 提出了许多具体的算法如 FUZZY(模糊)检验点策略, BLACK/WHITE(黑/白)策

略, COPY-ON-UPDATE(变更拷贝)检验点等,同时研究了非易失性内存在 MMDB 备份中的应用,如“影子内存”技术等。而备份恢复方面则是对 MMDB 重启之后的装入策略的研究,结合内存数据中数据使用的频率和优先级提出一种最优的装入策略,目前提出了“部分重装”等策略,MMDB 的目录恢复后就启动系统,然后根据要求再继续重装,从而提高 CPU 的使用率和事务的吞吐量。

5. MMDB 的并发控制,在短小事务时,往往采用大力度的锁,MMDB 近似于串行处理,串行化执行事务一方面使得并发控制的代价几乎完全消除,同时也能减少 CPU 的 cache 缓存和虚拟内存页表 TLB 的刷新频率;而对于长事务和多处理机环境串行处理明显不合适,这样就必需应用而合适的锁机制来支持并发操作,目前提出了二级层次封锁协议方案、乐观并发控制方法、使用可扩展的哈希技术的方法等。

近些年来随着大容量廉价的内存投入市场,使得上诉的各种技术也逐渐在内存数据库的设计和实现上得到了应用,内存数据库也不在停留在理论研究阶段,内存数据库走向了实际应用阶段,各高校和研究机构发布了研究模型系统,一些公司也推出了用于工业的商用内存数据库系统。目前工业应用上比较流行的商用系统有 Oracle 公司的内存数据库 Berkeley DB,内存数据库 SQLite,开源的内存数据库 FastDB,以及 McObject 公司的 eXtremDB。他们的特点如表 3.1 所示:

表 3.1: 主流内存数据库

内存数据库名	厂商	特点
Berkeley DB	Oracle	Berkeley DB 数据库系统简单、小巧、可靠、高性能,提供了一系列应用程序接口(API),应用程序和 Berkeley DB 所提供的库在一起编译成为可执行程序。每一个记录由关键字和数据(KEY/VALUE)组成的键值对构成。
SQLite	开源	SQLite 是一款轻型的数据库,它的设计目标是针对嵌入式系统,提供了很多语言的接口,支持 SQL 语句,支持事务处理,它占用资源非常的低,支持跨平台操作,操作使用简单。每完成一次操作需要进行内存和磁盘之间的同步。
ExtremeDB	McObject	ExtremDB 是为实时系统及嵌入式系统而特别设计的,完全工作在主内存中,不基于文件系统,支持事务,支持部分 SQL,支持稳定的 RAM。通过数据库定义语言为应用系统各自的 API。具有工业应用强度。
FastDB	开源	FastDB 是一个高效率的内存数据库系统,支持事务、在线备份和系统崩溃之后的自动恢复,支持类似 SQL 语言并提供了 C++ 接口,fastDB 算法和结构的优化都是基于数据存放在内存中这个假设上,但物理内存较小时也可使用。

## 四、系统实现技术研究

### 4.1 数据存储方式

## 参考文献

- [1] Lehman T J, Carey M J. A recovery algorithm for a high-performance memory-resident database system[M]. Vol. 16.[S.l.]: ACM, 1987.
- [2] Lehman T J, Carey M J. A Study of Index Structures for Main Memory Database Management Systems[C]. Very large data bases: proceedings twelfth International Conference on Very Large Data Bases, Kyoto, Japan, August 15-18, 1986, 1986, 294:294.
- [3] Berkowitz B T, Simhadri S, Christofferson P A, et al. In-memory database system[缺文献类型标志代码]. US Patent 6,457,021.
- [4] Garcia-Molina H, Salem K. Main memory database systems: An overview[J]. Knowledge and Data Engineering, IEEE Transactions on, 1992, 4(6):509--516.
- [5] Gruenwald L, Eich M H. MMDB reload algorithms[C]. ACM SIGMOD Record, 1991, 20:397--405.
- [6] 杨朝辉, 王立松. pT-树: 高速缓存优化的主存数据库索引结构 [J]. 计算机科学, 2011, 38(10):161--165.
- [7] Lehman T J, Carey M J. Query processing in main memory database management systems[C]. ACM SIGMOD Record, 1986, 15:239--250.
- [8] Lu H, Ng Y Y, Tian Z. T-tree or b-tree: Main memory database index structure revisited[C]. Database Conference, 2000. ADC 2000. Proceedings. 11th Australasian, 2000:65--73.
- [9] 林鹏, 李航, 徐学洲. 关键业务中内存数据库的 T 树索引优化 [J]. 计算机工程, 2004, 30(017):75--76.
- [10] 鲍程锋, 杨小虎. 基于影子页面和混合日志的 MMDB 恢复方法 [J]. 计算机工程与设计, 2011, 32(7):2373--2376.
- [11] 王洪海, 潘朝华. 内存数据库的数据结构分析 [J]. 现代电子技术, 2004, 27(003):96--98.
- [12] 刘云生, 许贵平. 内存数据库的图论存取方法 [J]. 计算机学报, 2001, 24(10):1095--1101.
- [13] 王珊, 肖艳芹, 刘大为, et al. 内存数据库关键技术研究 [J]. 计算机应用, 2008, 27(10): 2353--2357.
- [14] 杨武军, 张继荣, 屈军锁. 内存数据库技术综述 [J]. 西安邮电学院学报, 2005, 10(3): 95--99.

- [15] 张延松, 王占伟, 孙妍, et al. 内存数据库可控的 page-color 优化技术研究 [J]. 计算机研究与发展, 2011, 48(z2).
- [16] 王晨. 内存数据库若干关键技术研究 [D].[S.1.]: 杭州: 浙江大学, 2006.
- [17] 赵艳梅, 郑鑫福, 徐立臻. 内存数据库系统 SwiftMMDB 索引机制的设计与实现 [J]. 计算机应用, 2011, 31(9):2395--2398.
- [18] 谷国栋. 内存数据库相关技术的研究与分析 [J]. 电脑知识与技术, 2007, 4(19):5--6.
- [19] 张琪. 实时内存数据库的设计与实现 [D].[S.1.]: 武汉科技大学, 2008.
- [20] 刘云生, 廖国琼, et al. 一个支持实时内存数据库的恢复系统 [J]. 小型微型计算机系统, 2003, 24(3):460--463.
- [21] 袁培森, 皮德常. 用于内存数据库的 Hash 索引的设计与实现 [J]. 计算机工程, 2007, 33(18):69--71.