

以固态硬盘为缓存的存储技术研究

梁俊杰, 熊亚军

(湖北大学 计算机与信息工程学院, 湖北 武汉 430062)

摘 要: 在传统数据存储技术的基础上提出以固态硬盘为缓存的存储技术, 结合固态硬盘的写缩减技术, 实现热点数据与冷数据在内存-固态硬盘-机械磁盘之间的分级缓存与存储机制, 从而能充分利用内存资源, 增强与 CPU 的数据交互能力, 进而有效提升大数据环境下的数据处理速度。

关键词: 大数据; 多级缓存机制; 写缩减技术; 固态硬盘

中图分类号: TP303

文献标识码: A

文章编号: 1000-7180(2015)01-0040-05

Storage Technology Research Based on SSD-Buffer

LIANG Jun-jie, XIONG Ya-jun

(Computer and Information Engineering College, Hubei University, Wuhan 430062 China)

Abstract: Proposing a kind of storage technology which regards SSDs as storage devices based on traditional data storage technology, combining SSD's write-reduce technology to achieve classified buffering and storage mechanism of hotspot data and cold data in memory - solid state disk - mechanical hard disk, and this mechanism can make full use of memory resources. Application of this technology enhances data exchange capacity with the CPU, can effectively enhance data processing speed in big data environment.

Key words: big data; multi-level buffering mechanism; Write-Reduce; SSD

1 引言

近年来, 随着互联网、物联网、云计算、三网融合等 IT 与通信技术的迅猛发展, 数据的快速增长成为许多行业共同面对的严峻挑战和宝贵机遇, 因而信息社会已经进入了大数据时代。大数据的出现对传统的数据收集、数据存储、数据处理以及数据挖掘提出了新的挑战^[1], 而数据存储作为大数据处理过程中一个非常重要的环节, 利用现有技术优化和提升大数据的存储性能乃至彻底解决大数据的存储问题, 更好地进行大数据处理以及进一步推动大数据技术的研究具有重要意义^[2]。本文综合利用固态硬盘和机械硬盘的优势, 提出以固态硬盘为缓存的存储技术, 即在传统的内存-机械硬盘缓存和存储结构中增加固态硬盘一级作为二级缓存来提升大数据环境下的海量数据的缓存和存储性能, 对提高大数据

的处理效率具有重要作用。

2 大数据对传统存储技术的挑战

2.1 传统的数据存储技术

在大数据时代到来之前, 传统的数据存储技术采用内存-机械硬盘的缓存和存储结构。在传统的数据存储技术条件下, 有限数目的内存和机械硬盘直接进行数据交互。在这种环境下, 机械硬盘作为唯一的永久存储设备, 内存作为 CPU 和机械硬盘之间唯一的缓存设备, 其优点是结构简单, 可靠性较高, 缺点是计算机数目有限, 导致无论是缓存还是存储设备, 扩展能力都非常有限。

在大数据环境中, 从存储技术层面看, 传统技术的本质特点是缓存和存储结构单一, 随着大数据对处理性能要求的提升, 内存和机械磁盘二者之间的数据交互能力将会成为数据处理性能的瓶颈之一。

收稿日期: 2014-03-17; 修回日期: 2014-05-11

基金项目: 国家自然科学基金项目(61272111, 61202031); 湖北省自然科学基金项目(2013CFB002)

2.2 大数据及对传统存储技术的挑战

随着互联网、物联网、云时代的来临,大数据(Big data)也吸引了越来越多的关注.所谓大数据,或称巨量资料,指的是所涉及的资料量规模巨大到无法透过目前主流软件工具,在合理时间内达到撷取、管理、处理,并整理成为帮助企业经营决策更积极目的资讯^[3].大数据不同于传统的数据,目前业界普遍认为大数据具有4V+1C特性,即Volume(大量化)、Velocity(快速化)、Variety(多样化)、Veracity(价值密度低)以及Complexity(复杂度).

基于大数据的4V+1C特性,传统的缓存和存储技术已不能满足大数据分析和处理的需求,以下主要就应对大数据的Volume(大量化)和Velocity(快速化)特性来说明大数据对传统的缓存和存储技术带来的挑战.

大数据的大量化特性指数据规模通常可达到PB级别或者更大,虽然可以利用Google的GFS或Hadoop的HDFS等分布式文件系统对该级别的海量数据进行存储.但是在对这些已经存储的海量数据进行分析处理的过程中,虽然可以基于MapReduce模型^[4]将每个数据处理任务分层次划分若干个小任务,实现处理并行化,然而MapReduce并行处理的思想并没有解决传统存储技术中内存和机械硬盘之间速度不匹配的问题,即机械硬盘的I/O速度始终是瓶颈.并且大数据环境中,在数据处理平台配备海量内存的情况下,海量内存与基于机械硬盘的分布式文件系统之间的速度不匹配问题急剧累积,瓶颈效应将会更加明显^[5-8].

在这样的背景下,大数据处理平台的内存资源就得不到充分利用,CPU与机械硬盘的数据交互能力也因此受限,进而影响到大数据的分析处理速度^[9-10],即大数据处理的快速化特性得不到保证,甚至超出了大数据处理者所能忍受的极限^[11].

3 以固态硬盘为缓存的存储技术

3.1 固态硬盘简介

闪存固态硬盘(SSD)采用固态电子存储的Flash芯片阵列而制成的硬盘,由控制单元和存储单元组成,该类固态硬盘最大的优点就是可以移动,且数据保护不依赖电源^[12].

闪存固态硬盘和与机械硬盘存在显著的差别,它们的主要参数比较如表1所示.

通过以上参数对比可知,闪存固态硬盘具有读写速度快、重量轻和功耗低的优势,但受容量价格

比、使用寿命和数据可恢复性这些因素的限制,在目前条件下,一些应用环境中可以使用固态硬盘代替机械硬盘,但不是完全代替,即在固态硬盘和机械硬盘混合使用环境中综合利用二者的优势,提升数据的存储和读取性能^[13].

表1 固态硬盘和机械硬盘主要参数对比

对比项	固态硬盘	机械硬盘
存储和读取速度	快速存储和随机读取	存储和读取速度比SSD慢
容量价格比	相对较低	相对较高
使用寿命	短	长
数据可恢复性	基本不可回复	可恢复部分或全部
重量/功耗	相对较轻/高	相对较重/低

为了充分利用固态硬盘随机读取速度快的优势,同时兼顾闪存具有擦写次数限制的问题,本文提出了基于固态硬盘的写缩减技术.所谓写缩减技术,一方面利用固态硬盘进行数据缓存时,应尽量将机械硬盘中复用次数比较多的数据页写入固态硬盘上,反之则没有必要写入固态硬盘,调用时可直接从机械硬盘读入内存;另一方面,修改数据页的物理日志通常都是小数据量的,不适宜采用固态硬盘存储,因此通过仅将改变数据的物理日志保存到机械硬盘,而数据页始终在固态硬盘上保持原始状态,等待脏页需要从固态硬盘上淘汰时,才一次性将物理日志应用于数据页,写回机械磁盘,这样可以大大减少固态硬盘的写次数^[14];另外,当需要从固态硬盘调页进入内存缓存时,需要将与数据页相关的物理日志全部应用于数据,形成最新版本的数据页后,写入内存.

3.2 以固态硬盘为缓存的分层管理机制

固态硬盘在性能上相对传统机械硬盘有了明显的提升,但是在容量价格比、寿命等方面还有不足之处,并且传统硬盘通过RAID等技术在顺序读性能上也能达到要求.因此,在目前这一阶段,大数据存储需要考虑固态硬盘与机械硬盘相结合的缓存和存储技术^[15].

本文提出的多级缓存机制支持大内存环境与固态硬盘设备,能够在海量内存及配备固态硬盘的硬件环境下充分利用内存资源^[16],支持实现热点数据与冷数据在内存-固态硬盘-机械磁盘之间的分级缓存与存储机制,具体管理机制如图1所示.

在本分层缓存管理机制中,在机械硬盘和CPU之间的缓冲区由传统的单一的内存作为缓存区转变为由内存和固态硬盘(SSD)共同组成的分层缓冲区,其中,由于内存存储器有着比固态硬盘更快的数

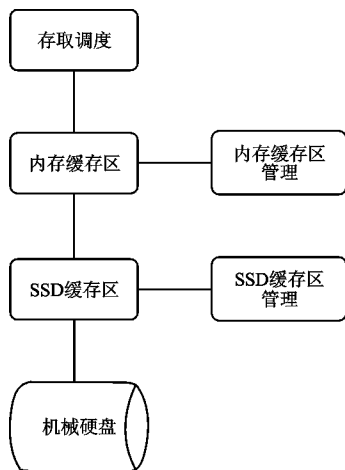


图1 分层缓存管理示意图

据读取速度,所以内存作为直接和CPU交互的缓存,称之为一级缓冲区,固态硬盘主要通过内存与CPU进行间接交互,称之为二级缓冲区。在这样的缓存机制中,结合固态硬盘随机读速度快及写入次数有限,使用RAID技术^[17]的机械磁盘读速度与固态硬盘大致相当的特性,可将访问频繁的热点数据和随机读数据从机械硬盘读出入SSD缓冲区,需要访问这些数据时再从SSD缓冲区读入内存与CPU进行交互,而将访问频率低的冷数据和顺序读数据直接存放在机械硬盘上,需要访问这些数据时直接将数据读入内存与CPU进行交互。

3.3 以固态硬盘为缓存的页操作策略

相对于传统的存储技术,本文提出的多级缓存机制在内存-机械硬盘结构的基础上增加固态硬盘作为二级缓冲区,即采用内存-固态硬盘-机械硬盘缓存和存储技术,为充分发挥出这种固态硬盘与机械硬盘相结合的缓存和存储技术的优势^[18],在缓冲区的页操作的过程中需要运用不同于传统存储技术体系中的页操作策略。

(1) 缓冲区页读取策略

数据页的存储位置有三个,即内存缓冲区、SSD缓冲区、机械硬盘,相应的对于同一页数据也就有三个版本,即内存缓存页、SSD缓存页+物理日志、机械磁盘页。但在任何时候,前面位置中存放的页为数据的最新版本,并且对于SSD缓存和机械磁盘页的更新都遵循先写日志原则^[19]。

结合固态硬盘的写缩减技术,当内存缓存缺页时,页读取策略如图2所示,其中 $p_1, p_2, p_3, \dots, p_m$ 表示编号从1到 m 的 m 个数据页, $L_{p_1 t_1}$ 表示数据页 p_1 在 t_1 时刻由于更新而写到机械硬盘的物理日志,

依此类推……(下同)。

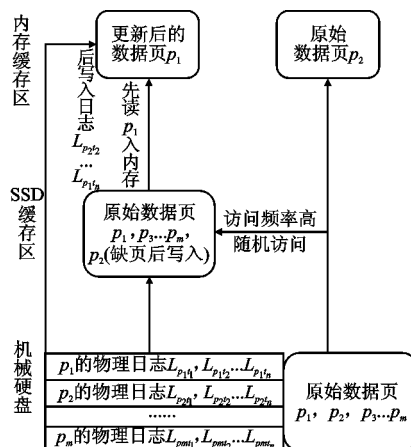


图2 缓冲区页读取策略

先检查SSD缓存区,如果存在该数据页,则从SSD缓存区读入该页数据的原始版本到内存缓存区,并且根据关联的物理日志对内存页数据进行修改,以保证内存缓存区中的数据页始终是最新版本。

如果SSD缓存中不存在该数据页,则从机械硬盘中直接将数据页读入内存缓存区,同时根据该页的数据特点即数据访问的频率,随机访问还是顺序访问来决定该数据页是否需要写入到SSD缓存区中,如果访问频率比较高或者是随机访问则将该数据页写入SSD缓存区中,反之则不写。

(2) 缓冲区页更新策略

当内存缓存区中页更新后,按照先写日志原则,事务提交时总会先更新机械硬盘上的物理日志,然后检查该数据页是否在SSD缓冲区中。如果不在,则直接用内存缓冲区中更新后的数据页写入机械硬盘替换该数据页的旧版本;如果在,则结合写缩减技术和该数据页的特点决定该数据页是继续以原始版本保留在SSD缓存中,还是直接将此页设置为失效,纳入空闲页管理。

如果该更新页在最近的一段时间内访问频率比较高或者是随机访问,则将该数据页继续保留在SSD缓冲区中,并且不做任何修改,并且将相关的更新日志写到机械硬盘上;如果该更新页访问频率不高或者是顺序访问,则将SSD缓冲区中的该页设置为失效并将该页所占空间纳入空闲区管理,内存缓冲区中更新后的数据页写入机械硬盘替换该数据页的旧版本。具体更新策略如图3所示,其中①表示 p_2 在SSD缓冲区中保持不变,②表示将 p_2 纳入空闲区管理,③表示 p_2 的物理日志应用到 p_2 并写入机械硬盘,④将某时刻 p_2 更新的物理日志写入机械硬

盘;而且①④是同时出现的,②③也是同时出现的,并且①④与②③不会同时进行。

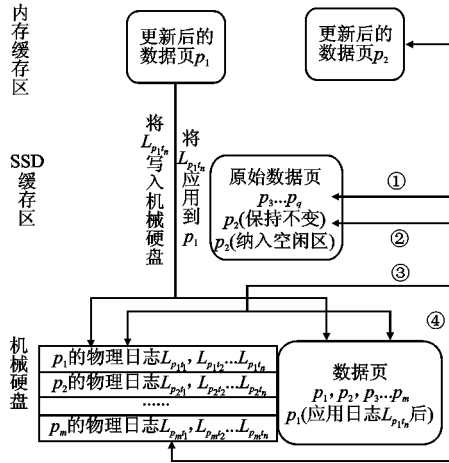


图 3 缓冲区页更新策略

(3) 缓冲区页淘汰策略

考虑内存与固态硬盘特性的不同,以及应用性质的不同,内存缓冲区管理器和 SSD 缓冲区管理器需要采用不同的脏页淘汰策略。

当内存缓冲区需要淘汰脏页时,如果 SSD 缓存区中存在该页,则 SSD 缓存区管理器不做任何操作,因为 SSD 缓存需要始终保存数据页的原始内容,此时它与机械硬盘上与该数据页相关的物理日志共同构成了数据的最新版本。

如果 SSD 缓存区中不存在该页,则直接刷新写入机械硬盘替换原旧版数据页。内存缓冲区的脏页淘汰具体策略如图 4 所示。

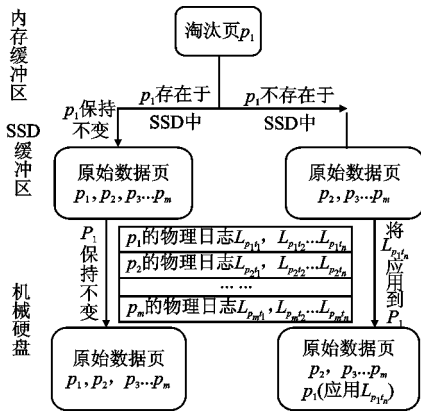


图 4 缓冲区页淘汰策略

当 SSD 缓存区需要淘汰脏页时,如果内存缓存区有此页,则直接将此页淘汰,因为内存中已经存在最新的数据版本。

如果内存缓存区中无此页,则先将 SSD 缓存区中的此页读入内存,然后读入物理日志,将累计的物

理日志一次性应用于该数据页,最后写入机械磁盘,最后再将此页从 SSD 缓冲区中淘汰。SSD 缓存区中淘汰脏页的策略如图 5 所示。

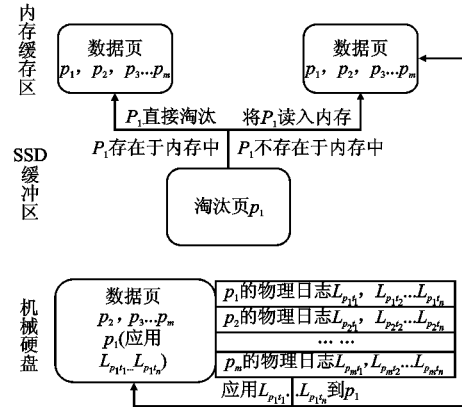


图 5 SSD 缓冲区页淘汰策略

利用以上技术,SSD 缓存中的数据页除了一次性从机械磁盘更新到固态硬盘上的 SSD 缓存外,不存在其他写入的操作,这样充分发挥了固态硬盘随机读能力强的优势,提升了大数据环境下对海量数据的处理速度,同时避免了频繁小数据量随机写请求的性能下降和对寿命的影响^[20]。

4 以固态硬盘为缓存的存储技术的优势

在以固态硬盘为缓存的存储技术中采用多级缓存机制,其中作为二级缓存的固态硬盘在一些页操作过程中能有效缓解内存和机械硬盘速度不匹配的问题,充分体现了以固态硬盘为缓存的存储技术的优势,具体分析如下:

在缓冲区的页读取策略中,当需要从机械硬盘中读取一数据页到内存中进行处理时,并且该数据页同时存在于 SSD 缓存区时,可直接从 SSD 缓存区将该数据页读入内存并将相应的物理日志写入即可使用数据。此种情况下,一是在不考虑数据页定位问题时数据从 SSD 到内存比从机械硬盘到内存速度要快得多;二是在不考虑 SSD 和机械硬盘分别读数据到内存的速度问题时,由于固态硬盘不使用磁头,数据页定位也比机械磁盘要快,此时页读取速度肯定要比传统技术快。同时记录机械硬盘中每个数据页的访问频率和访问方式(随机访问还是顺序访问)决定是否加入 SSD 缓冲区,保证了 SSD 缓冲区的更新。在本页读取策略中,最坏情况是要从机械硬盘中读入内存的数据页不在 SSD 中,读取时直接从机械硬盘到内存,此即传统存储技术^[21]。综合考虑两种情况,以固态硬盘为缓存的缓存和存储技术能有效

提升页读取速度.

在缓冲区页更新策略中,如果内存缓冲区中的更新页存在 SSD 缓存区中并且访问频率高或者随机访问,则直接将更新日志写到机械硬盘,当更新页最终需要写到机械硬盘上时,只需要一次将数据页写到机械硬盘同时将所有物理日志(在有日志的情况下)一次性应用于数据页即可完成一次更新.在本页更新策略中,最坏情况是内存缓冲区中的更新页访问频率低或者顺序访问,此时内存中更新一次数据页就需要写入机械硬盘一次,此即传统存储技术.综合考虑两种情况,以固态硬盘为缓存的缓存和存储技术能有效提升页更新速度.

综合以上分析可知,以固态硬盘为缓存的存储技术能够有效提升页的读取和更新速度,即提高数据的读写速度,进而提升了大数据环境下对海量数据的处理速度,能够有效缓解或者应对大数据对传统存储技术带来的挑战.

5 结束语

本文提出大数据环境下的以固态硬盘为缓存的缓存和存储技术,结合固态硬盘的写缩减技术^[22],能有效提升大数据环境下数据分析与处理效率,对推动大数据的理论研究和实际应用具有积极的意义.

参考文献:

- [1] Andrzej Jankowski, Andrzej Skowron. Dealing with data[J]. Science, 2011, 331(6018): 647-692.
- [2] Jeffrey Dean, Sanjay Ghemawat. MapReduce: A flexible data processing tool[J]. Communications of the ACM, 2010, 53(1): 74-76.
- [3] IBM. Bring big data to the enterprise [EB/OL]. [2013-02-05]. <http://www-01.ibm.com/software/data/bigdata>.
- [4] Jiang D, Ooi B C, Shi L, et al. The performance of MapReduce: An in-depth study[J]. PVLDB, 2010, 3(1): 464-453.
- [5] Floratou A, Patel J M, Shekita E J, et al. Column-oriented storage techniques for MapReduce[J]. PVLDB, 2011, 4(7): 411-427.
- [6] Afrati F N, Ullman J D. Optimizing joins in a map-reduce environment[C]//Proceedings of the 13th International Conference on Extending Database Technology. Switzerland, Lausanne, 2010: 80-92.
- [7] Stonebraker M, Abadi D J, DeWitt D J, et al. MapReduce and parallel DBMSs: Friends or foes? [J]. Communica-

tions of the ACM, 2010, 53(1): 53-68.

- [8] DeWitt D J, Gerber R H, Graefe G, et al. Gamma-A high performance dataflow database machine[C]//Proceedings of the 12th International Conference on Very Large Data Bases. Kyoto, Japan, 2010: 228-237.
- [9] Ginsberg J, Mohebbi M H, Patel R S, et al. Detecting influenza epidemics using search engine query data[J]. Nature, 2009, 457(19): 987-1003.
- [10] Ranger C, Ranghuraman R, Penmetsa A, et al. Evaluating mapreduce for multi-core and multiprocessor systems[C]//Proceedings of the IEEE 13th International Symposium on High Performance Computer Architecture. Phoenix, AZ, USA, 2009: 13-24.
- [11] Wegner D, Mock M, Adranale D, et al. Toolkit-based high-performance data mining of large data on MapReduce clusters[C]//Proceedings of the 9th IEEE International Conference on Data Mining. Miami, FL, USA, 2010: 296-301.
- [12] 肖韬. 基于 MapReduce 的信息检索相关算法并行化研究与实现[D]. 南京: 南京大学, 2012.
- [13] 王珊, 王会举, 覃雄派, 等. 架构大数据: 挑战、现状与展望[J]. 计算机学报, 2011, 34(10): 1744-1749.
- [14] 孟晓峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(1): 151-162.
- [15] 李佳琦, 李健, 党相甫, 等. 基于固态硬盘的闪存阵列并行结构设计[J]. 微电子学与计算机, 2010, 27(4): 85-88.
- [16] 王元卓, 靳小龙, 程学旗. 网络大数据: 现状与展望[J]. 计算机学报, 2013, 36(6): 1127-1134.
- [17] 周傲英, 金澈清, 王国仁, 等. 不确定性数据管理技术研究综述[J]. 计算机学报, 2009, 32(1): 3-8.
- [18] 马帅, 李建欣, 胡春明. 大数据科学与工程挑战与思考[J]. 中国计算机学会通讯, 2012, 8(9): 22-28.
- [19] 赵鹏, 白石. 基于随机游走的大容量固态硬盘磨损均衡算法[J]. 计算机学报, 2012, 35(5): 972-974.
- [20] 窦万春, 江澄. 大数据应用的技术体系及潜在问题[J]. 中兴通讯技术, 2013, 19(4): 8-14.
- [21] 姜承尧, 陈庆奎, 钱剑飞. 一种基于闪存固态硬盘的辅助缓冲池设计[J]. 计算机系统应用, 2011, 20(8): 2-5.
- [22] 谢健聪, 肖依, 褚瑞. 一种分布式磁盘缓存的设计与实现[J]. 微电子学与计算机, 2007, 24(9): 172-173.

作者简介:

梁俊杰 女, (1974-), 博士, 副教授, 硕士生导师. 研究方向为数据库管理系统、大数据和云计算技术等.

熊亚军 (通讯作者) 男, (1986-), 硕士研究生. 研究方向为大数据的存储和查询优化. E-mail: bearyajun@163.com.