基于HDD和SDD的混合存储系统研究综述

陈震 刘文洁 张晓 卜海龙

西北工业大学计算机学院

NCIS'2016

引言

大容量、低成本、高性能的存储系统设计一直是存储领域研究的热点，随着大数据云存储时代的到来，数据量成爆炸式的增长对存储系统提出了更高的要求。一方面存储系统既要能够在低成本前提下实现大容量存储，另一方面存储和计算之间的性能差距不断的扩大，这就需要具有在海量数据规模下与计算性能相匹配的高性能数据访问能力。我们将容量、成本、顺序性能、随机性能、寿命作为主要指标来观察不同类型存储其的性能。如图1 所示。

本文将从以下几个方面展开：首先从混合存储系统的概念、基于SSD与HDD的混合存储系统研究现状进行介绍。接下来以目前主流的基于SSD和HDD的混合存储系统为例，对其中涉及的关键技术进行对比分析，重点讨论不同的当前主流的设计方法而不讨论实验平台、测试集和电路设计。最后对当前混合存储系统发展所存在的问题进行探讨，并对今后该领域的研究重点和方向进行展望。

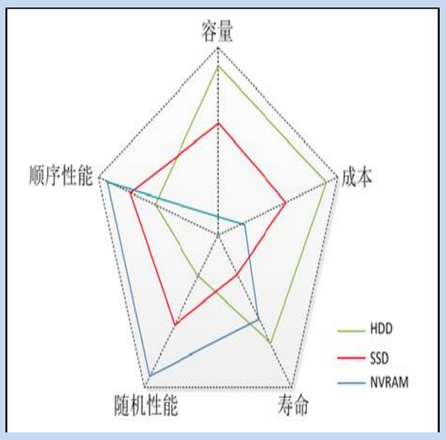

图 1 几种存储介质特性差异

1.混合存储系统简介

混合存储（hybrid storage）是一种数据存储方法，将具有不同特性的存储设备进行组合，根据数据访问特点和系统负载等情况，尽可能将数据请求交给最适合处理该请求的设备，进而提高整个系统的性价比、使用寿命、可靠性、容量等指标。

混合存储的主要目标是使存储的数据充分利用不同存储介质的特性，在保证存储系统容量的同时尽可能地提高性价比。混合存储所采用的介质可以是 NVRAM、不同转速的磁盘（SAS、SATA）、SSD、磁带等。需要说明的是本文主要针对将SSD与HDD进行混合的存储系统分析。

2. SSD作为磁盘的缓存架构

其基本原理是将SSD作为磁盘的cache，SSD中存放的数据是HDD存放数据的子集。其基本架构图如图2所示，混合存储系统的逻辑地址与磁盘中的物理地址是一一对应的，SSD中只是缓存了磁盘中的部分数据的拷贝。当有上层访问请求时，首先会在SSD中进行查找，如果该数据在SSD上，则直接返回数据。否则再去访问磁盘。目前国内外研究者已经对此做了大量的研究工作。

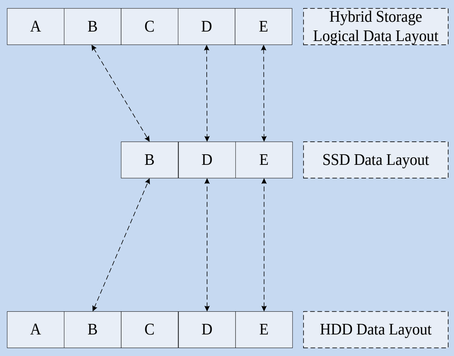

图 2 SSD的缓存存储架构

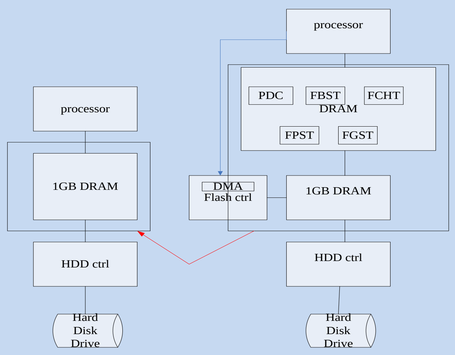

图 3 系统架构图

针对闪存磨损造成的系统降低的可靠性以及垃圾回收机制的问题，文献[16]在前文的研究基础上同样提出了一个使用闪存作为HDD的缓存存储架构，如图3所示。实验的结果表明其不仅降低了主存的能耗，还提高了整个系统的性能和可靠性。然而文献并没有对数据的热度进行分类，因此这会造成大量的冷数据占据SSD缓存。同时也没有考虑到Flash缓存的的利用率和缓存的命中率。

3. SSD与磁盘设备同层架构

目前将SSD和HDD用作同一级别的存储设备更为普遍，其主要原理是将SSD和HDD放在同一存储架构，对其进行统一编址，存储的总容量是SSD和HDD的容量之和，其存储架构图如图4所示，任何一份数据仅仅存储在SSD和HDD其中之一上。这种存储模型在数据库、Web应用等系统中运用非常广泛。

文献[24]首先提出了一个将SSD和HDD作为同级别存储的在线混合存储模型，其结构图如图5所示，模型主要聚焦于数据库系统，该模型根据页面的工作负载来决定将页面存放在哪一个磁盘。并根据迁移代价来决定是否将页面在SSD和HDD上进行迁移。具有读密集型的页放置在SSD上，而具有写密集型和频繁更新的页则放置在HDD上。该模型的特性如下：

- 充分利用了两种磁盘各自的优势，相比传统的存储模型在读、写和整个I/O开销上均有不少的提高。
- 提出了一个根据最近的页面访问决定页面放置的算法。并且该算法能够自适应页面负载的变化。
- 提出了一个缓冲区页面替换算法。每个页面被放置在闪存读队列、闪存写队列、磁盘读队列、磁盘写队列中的一个，并根据闪存读队列、磁盘读队列、磁盘写队列、磁闪存写队列从高到低的优先级进行进行页面的置换。相较传统的LRU算法提高了I/O性能。

但是，在HDD上的随机写和连续写开销并没有被考虑到。且文件内部的并行性[25]、芯片的擦除损耗[26]等并没有考虑在内。

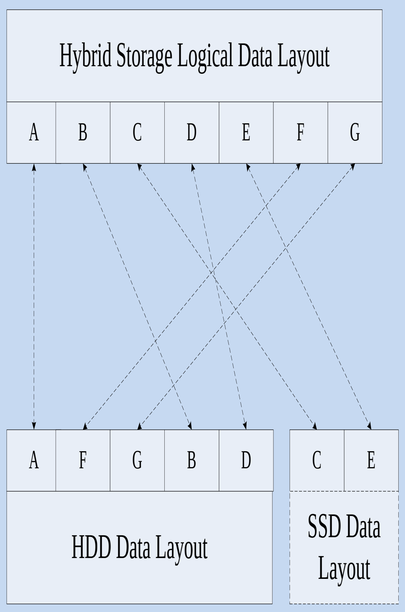

图 4 SSD与HDD设备同层架构

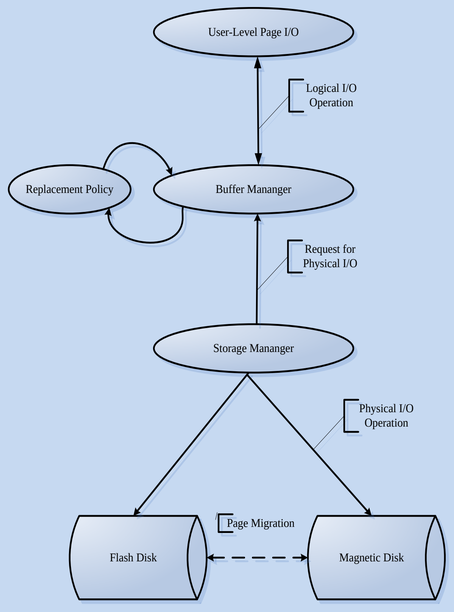

图 5系统整体架构图

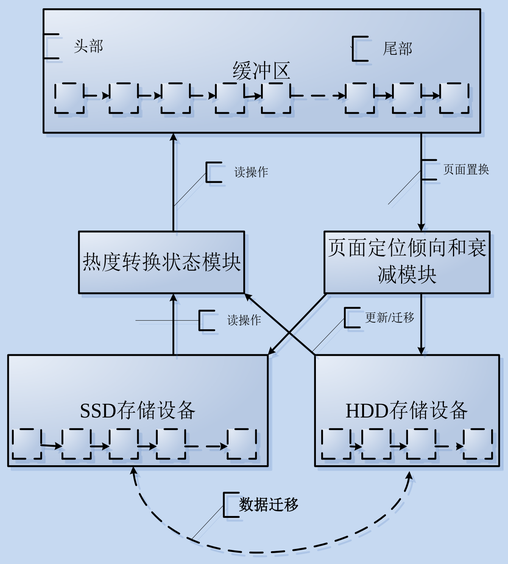

图 6 系统架构图

文献[27]在文献[24]基础上提出了一种改进的混合存储模型，其结构图如图6所示，在文献[24]提到的模型中I/O统计涵盖了该页面自生成以来的所有访问请求，这必然会带来累积效应，这种累积效应会使得访问负载变化时不能快速响应。因此文献引入了一个时间衰减因子，当一个页面的热度在冷热状态之间转换或者物理访问之间的时间间隔过长，那么将会根据衰减因子降低I/O代价计算的值对页面分类的影响并且重置I/O统计值。这样就能够避免统计的累积效应。能够对页面进行更加准确的分类。做出更加准确的判断。其次，为了能够更加准确的分类页面，在前文的基础上为页面引入了一个处于hot和cold中间的warm状态。这样能够避免那些偶尔变热的cold页面进行迁移，减少了不必要的迁移操作。文献还将SSD与HDD容量的不同比例组合进行了研究，并分析性价比。实验的结果表明系统的性能获得明显的提升。

4.磁盘作为SSD的缓存架构

文献[33]在前文的基础上同样提出了一个将磁盘作为固态硬盘的写缓存架构，如图7所示，确保所有的写操作第一时间发生在磁盘上，相较前文文献提出了一个以页和块为数据粒度的迁移算法，将频繁读的页面迁移到SSD上以利用SSD良好的读性能，将写频繁的页存放在HDD上以减少对SSD的磨损。实验的结果表明系统降低了整个系统的I/O开销。

文献[34]专注存储系统的可靠性和高性能，提出了一个崭新的框架，其系统架构图如图8所示，其将两个磁盘组成一个RAID1，并将这个RAID1作为固态硬盘磁盘阵列RAID4的写缓存，其中一个HDD作为RAID4的校验盘，而另外一个HDD则作为RAID4的写缓存，负责几乎所有的写请求，并且还充当了另一个磁盘失效时作为恢复的磁盘。由于HDD作为检验盘因此避免了SSD的校验更新操作。另外，RAID1组成写缓存又降低了针对SSD的写操作，同时组成的利用RAID技术提高了存储系统的可靠性。

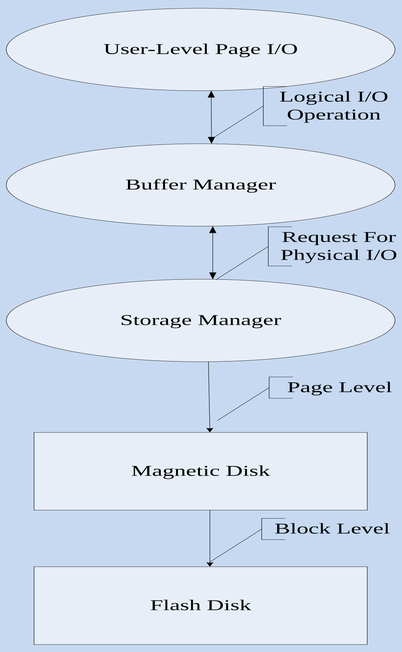

图 7系统架构图

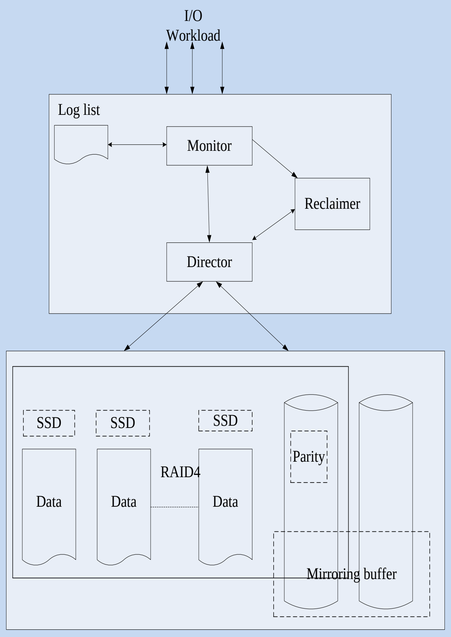

图 8 HPDA架构图

5. 总结与展望

本文分析了目前这类流行混合存储系统中涉及的关键技术。总结如下：

- 目前SSD还不能完全取代传统的磁盘，因此结合传统的磁盘利用各自性能的优势组合成的高效混合存储系统是一个很好的选择。
- SSD固有的写前擦除的特性以及受限制的写寿命仍然是影响混合存储系统性能的一个非常关键的因素，虽然对此国内外学者提出一些相应的解决方法，但是并没有获得很好的突破。
- 混合存储系统中针对数据的热度识别技术也不够成熟，虽然在不同的粒度下提出了一系列相应的热点数据识别方法，但对于热点数据的识别仍然不够精确。这些问题将依旧是今后研究的重点。
- 随着大量新型存储介质如：新型非易失型存储诸如铁电存储器（ferroelectric RAM，FeRAM）、磁性存储器（magnetic RAM，MRAM）、自旋转移力矩存储器（spin-transfer torque RAM，STT-RAM）、相变存储器（phase change memory，PCM）、阻变存储器（resistIve RAM，RRAM）、赛道存储器（domain-wall memory，DWM）等的出现给未来的混合存储系统带来了新的机遇。利用新型存储器良好的写性能、固态硬盘良好的读性能和传统磁盘大容量低成本的特性组合成的混合存储系统可能是未来发展的方向。