doi:10.3772/j.issn.1002-0470.2017.05.005

希捷瓦记录磁盘评测①

张 强②*** 刘振军* 董欢庆* 马留英*** 李猛坤***

(*中国科学院计算技术研究所 北京 100190)

(**中国科学院大学 北京 100049)

(***首都师范大学管理学院 北京 100045)

摘 要 分析了可提升磁盘存储密度的瓦记录(SMR)技术的优势及局限性,指出 SMR 采用部分叠加相邻磁道的方式增大磁盘的面密度,但是这也带来了写覆盖(Write Overlay)问题,即更新某一磁道上的数据时会覆盖相邻磁道上的有效数据,导致瓦记录磁盘(SWD)无法支持原地更新,使其非顺序写性能受到影响。为了促进该技术的应用,通过 Fio 和 Filebench 测试工具,对 SMR 技术的代表性产品——希捷瓦记录磁盘(SSWD)做了比较全面的测试,并对多种应用场景下 SSWD 的性能做了详细的测试和分析,以便为 SSWD 在数据中心的应用提供重要的参考。

关键词 瓦记录(SMR),瓦记录磁盘,写覆盖,原地更新

0 引 言

磁盘的存储容量虽然过去以每年 30% ~50% 的速度增长,但磁盘单碟片的面密度将很快到达 1Tb/in^2 的上限,主要限制因素是超顺磁效应(super-paramagnetic effect)[1]。为了保证磁盘容量的增长速度,业内提出了多种新的存储技术,如晶格介质(bit patterned media)[2]、热辅助磁记录(heat assisted magnetic recording)[3]、微波辅助记录(microwave assisted magnetic recording)[4]等,由于这些技术的实现都需要大幅改变现有磁盘的物理结构,实现成本高,因此还处于实验阶段。一种称为瓦记录(shingled magnetic recording,SMR)的技术已经被业内公认为是下一代磁盘最可行的实现方式,它可在最小化现有磁盘物理改进的前提下实现磁盘容量的扩增,可将现在磁盘的面密度提升 2 ~3 倍[5]。同时国际信息技术标准委员会(INCITS)的 T10 和 T13 研究组也正在研究制定瓦记录磁盘(shingled write disk,SWD)的相关接口标准,标准化 SWD 的磁盘接口。

瓦记录(SMR)通过叠加相邻磁道[6]的方式来大幅增加磁盘密度。与其他新的存储技术相比较,其优势在于,SMR 技术是建立在现存的磁盘技术上,无需对磁盘物理结构做太多的改变,实现成本相对较低,且便于集成到现有的存储系统中。虽然 SMR 技术可大幅提升磁盘存储容量,但磁盘的磁道重叠引起了磁道间的干涉,写一条磁道时会覆盖后续相邻的若干条磁道的数据,这就导致了 SWD 不能按位置更新的局限性,即无法原地更新[7]。要想保证数据写入的正确性,就必然要保证已经存在的数据不会被破坏,因此会引入额外的开销,一种简单的

① 中国科学院战略性先导科技专项(XDA06010401),中国科学院重点部署项目(KGZD - EW - 103 - 5(7))和北京市教育委员会科技计划(km201510028019)资助项目。

② 男,1982 年生,博士生;研究方向:瓦记录,磁盘冗余阵列,分布式系统;联系人,E-mail: zhangqiang@ict.ac.cn

(收稿日期:2016-12-16)

方案是引入读修改写技术(read-modify-write, RMW),即写入数据前先将可能被覆盖的数据读出,待写入操作完成后再将读出的数据写回,这极大的影响了 SWD 的非顺序写性能。而对于顺序写,由于磁头是按照一个方向顺序移动写入数据,因此不会产生数据覆盖问题,也就无需引入额外的操作。

希捷作为知名的磁盘厂商之一,在 2014 年率先发布了其首款 SWD,并将其定位在桌面应用和动态归档领域。从其定位可以看出,虽然希捷对其 SWD 内部进行了重新设计,一定程度上减小了读修改写(RMW)技术带来的性能损失,但是依然没有很好地解决 SMR 的非顺序写性能问题。本文介绍了希捷的瓦记录磁盘(Seagate shingled write disk, SSWD)的工作原理并在多种场景下对其进行了测试。测试结果表明,对于顺序和非顺序读,SSWD 的性能略优于普通磁盘;对于以非顺序写为主,但 IO 间隔较大的场景,其性能表现也与普通磁盘相当,甚至要优于普通磁盘,但是在写密集且非顺序的场景下,其写性能会大幅下降,从而也说明其更适合于读请求占比高、IO 顺序化好和 IO 非密集型的应用场景。

1　SSWD

希捷的瓦记录磁盘(SSWD)将磁盘物理空间分成了三部分[8]:持久缓存区、数据区和 Map 区。

持久缓存区位于 SSWD 的外圈,占用磁盘存储空间几十 GB 左右,大小与磁盘容量有关。持久缓存区以 Log 的方式管理,即发往磁盘的所有非顺序写请求会先被顺序地写入磁盘的持久缓存区,待缓存区满了或者磁盘空闲一定时间后,SSWD 会将持久缓存区内的数据合并然后移动到磁盘的数据区。

数据区为瓦记录磁盘(SWD)内数据最终存放的区域,区域按带顺序划分,带大小在 18MB ~38MB 左右[8],带内依然有写覆盖问题,无法做原地更新,但是相邻的带之间预留了一部分空间,使得带间数据的写入不会互相覆盖。持久缓存区被写满后会以带为单位将数据移动到数据区的带上,即下刷。下刷前会先将目标带内的旧数据读出,与持久缓存区内的新数据合并,最终将新的带内数据整体顺序写回带中,即读修改写(RMW),这种以带为单位的 RMW 过程避免了更新数据区时可能产生的写覆盖问题,但是依然引入了额外的数据迁移开销。

Map 区位于盘片的中间,主要存放持久缓存的映射信息。由于持久缓存区采用 Log 方式写入数据,因此需要维护数据块在持久缓存区和带内的位置信息。并且为了尽可能避免掉电后持久缓存区写入的数据丢失,需要周期性地持久化映射信息到 Map 区域,而这一过程必然会导致磁头的抖动。Map 区位于盘片中间可能是为了避免访问 Map 区域时大范围的磁头抖动。

对于非顺序写,数据先被写入持久缓存区然后再下刷,下刷策略为 RMW,这种设计原理必然导致在持久缓存未满之前 SWD 的非顺序写性能会很好,一旦启动下刷,性能会急剧下降。而对于顺序写,则无需写入持久缓存区,可直接写入数据区,并不会导致写覆盖,因此可保持较高的写性能。

从磁盘架构上来看,SSWD 采用了将原地更新(in-place update)和非原地更新(out-place update)结合的方式,但是依然没有避免或改进两种技术的缺点,从希捷将其定位在桌面应用和动态归档领域也能看出,SSWD 在写压力较大的非顺序场景下的持续 IO 性能应该是较差的。

2　测试环境配置

本测试选择的测试工具是 Fio 和 Filebench。Fio 是一款在 linux 下的对块设备做压力测试的工具,可以通过调整测试参数,对块设备做各种类型的读写测试。Filebench 是一款可以模拟实际应用的块设备测试工具,通过模拟实际应用场景下的文件系统操作来测试块设备的 IO 性能。

本测试采用 Fio 对普通磁盘 HDD(hard disk drive)和 SSWD 分别做压力测试,是为了对比 HDD 和 SSWD 的极限性能。采用 Filebench 是为了模拟实际应用的 IO 特征,从而说明 SWD 的适用场景。测试所用的软硬件环境和配置见表 1。

表 1　测试环境配置

硬件环境：		
服务器	Dell R730	
CUP 类型	Intel(R) Xeon(R)CPU E5-2630 v3@2.40GHz	
CPU 数量	16	
内存容量	16GB	
HDD	4TB\SATA\7200 RPM\128MBcache\ write-caching = 1(on)	
SSWD	8TB\SATA\5900 RPM\128MBcache\ write-caching = 1(on)	
软件环境：		
操作系统	CentOS6.5	
内核版本	3.18.30	
Fio 版本	2.6-31-gde40	
Fio 配置	direct = 1(use non-buffered I/O) Io depth = 待测磁盘数量 ×32 Io engine = libaio (随机读写测试需开启异步引擎)	
Filebench 版本	1.4.9.1	
RAID 配置		
	RAID 级别	RAID0/RAID10/RAID5
	磁盘数量	4
	RAID5 Stripe Cache	16384
	RAID5 Chunk Size	512kB
	Queue Depth	磁盘数量 ×32

3　Fio 测试

3.1　单盘 Fio 顺序读测试

测试对比了 SSWD 和 HDD 在 4kB 和 64kB 粒度下的顺序读带宽(图 1)。

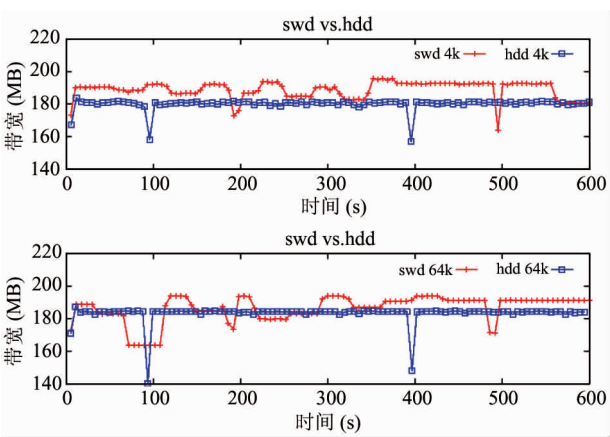

图 1　SSWD 与 HDD 顺序读带宽对比

由于 SSWD 的面密度比普通 HDD 要高,因此对于顺序读来说,在磁盘具有相同转速的前提下,相同时间内 SSWD 可读取更多的数据,从图中测试结果也可以看出,即使 SSWD 的转速比普通 HDD 要低很多,但在相同时间内其带宽依然高于普通 HDD,也说明了面密度的提升对顺序 IO 性能的提升。

从测试结果也能看出,SSWD 的顺序读性能波动较大,这可能是由于对于读操作来说,SSWD 需要先检查是否命中持久缓存,因为有可能持久缓存中的数据与带上的数据是不一致的,因此不能直接从带上读取数据。而检查缓存是否命中就需要先从 SSWD 的 Map 区域读取映射表信息,从而进一步判断读请求是否应该访问持久缓存。读取 Map 区域会引入磁头抖动,这可能是导致 SSWD 顺序读抖动严重的主要原因。另外,由于是顺序读,因此 Map 区域的映射表读取操作可能也是顺序的,因此磁头不需要频繁寻道到 Map 区域读取映射表,所以对顺序读的性能损失不大,不足以抵消面密度增大带来的顺序读性能提升。

3.2　单盘 Fio 顺序写测试

测试对比了 SSWD 和 HDD 在 4kB 和 64kB 粒度下的顺序写带宽(图 2)。

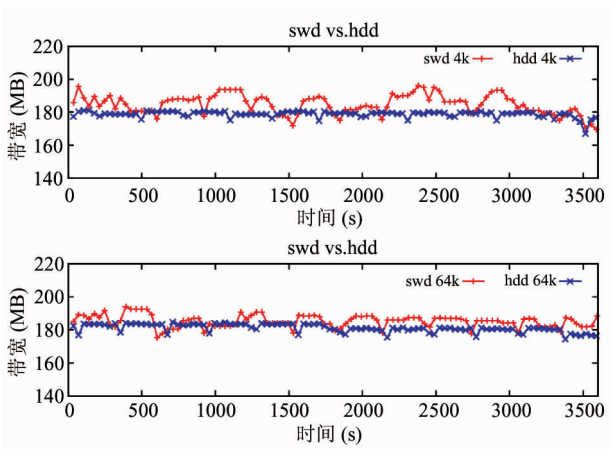

图 2　SSWD 与 HDD 顺序写带宽对比

从图中测试结果可以看出,在顺序写方面,SSWD 带宽略高于 HDD。同顺序读类似,说明了磁道间距离变窄以后,即使 SSWD 的磁盘转速比 HDD 低,相同时间内 SSWD 依然可以写入更多的数据。并且,由于顺序写可以绕过持久缓存区直接写入带

内,因此不需要更新 Map,也避免了磁头的大范围抖动。

3.3　单盘 Fio 随机读测试

测试对比了 SSWD 和 HDD 在 4kB 和 64kB 粒度下的随机读带宽(图 3)。

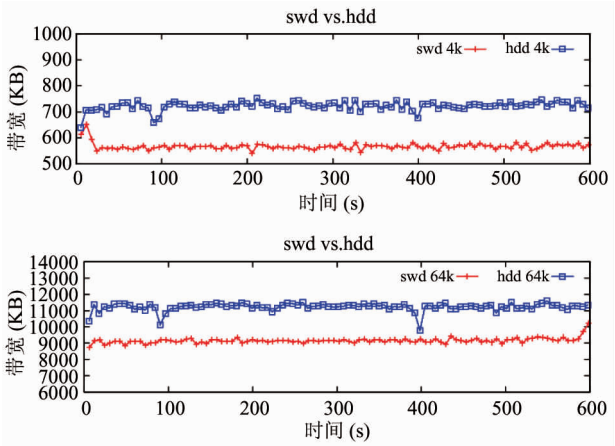

图 3　SSWD 与 HDD 随机读带宽对比

从测试结果可以看出,SSWD 比 HDD 的随机读性能要略微低一些。这是由于 SSWD 对于随机读操作来说需要先检查是否命中持久缓存。同顺序读一样,检查缓存是否命中就需要先从 SSWD 的 Map 区域读取映射表信息,从而进一步判断读请求是否应该访问持久缓存。由于测试的是全盘随机读,因此 Map 区域的读取映射表操作很可能也是随机读操作,并且磁头需要频繁寻道到 Map 区域读取映射表,这可能是 SSWD 随机读性能比 HDD 差的主要原因。

3.4　单盘 Fio 随机写测试

测试对比了 SSWD 与 HDD 分别在 4kB 和 64kB 粒度下 IO 的平均延迟。对于 SSWD,分别将发送 IO 的时间间隔设置为 0μs、100μs、1ms、10ms,对于 HDD,IO 发送的时间间隔为 0μs(图 4)。

从测试结果可以看出,在 IO 发送时间间隔为 0 的情况下,SSWD 在 4KB 和 64KB 两种 IO 粒度下的请求平均响应延迟差不多,只是 IO 粒度变大后延迟急剧变大的时间点有所延后,说明 IO 粒度变大后,SSWD 的持久缓存的有效缓存容量更大[8]。另外,从图中可以看出,随着 IO 时间间隔的增大,SSWD 启动回收的时间也不断延后。由于 SSWD 回收一个带的时间在0.6s ~ 1.6s 之间[8],可以预测,当 IO 间隔足够大以后,SSWD 可有充分的时间做回收,因此持久缓存区可吸收所有非顺序写,从而大幅降低 IO 延迟。因此,对于 IO 压力不大的应用场景,SSWD 可以表现出持续的低 IO 延迟。

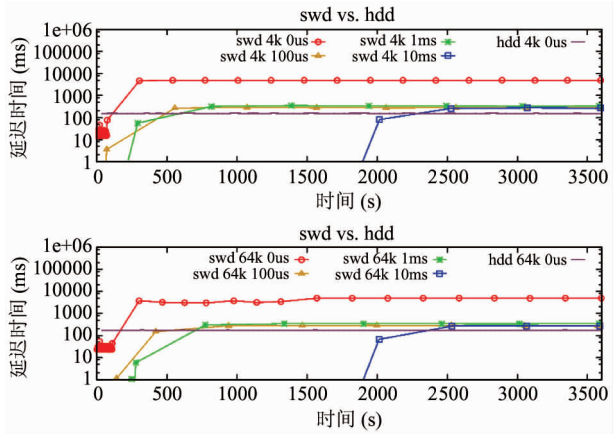

图 4　SSWD 与 HDD 随机写延迟对比

3.5　独立冗余磁盘阵列(RAID)0\10\5 Fio 顺序读测试

测试分别对比了 RAID0\10\5 在 IO 粒度为 4kB 和 64kB 时的顺序读带宽(图 5 ~ 图 7)。

从测试结果可以看出,SSWD RAID0 的顺序读性能要略微高于 HDD RAID0,与单盘顺序读测试结果类似。对于顺序读来说,RAID0\10\5 都是对多块磁盘的并发读操作,对于每个磁盘来说也都是顺序读,因此性能高的原因主要是磁盘面密度增大了,单位时间内可以读取更多的数据。

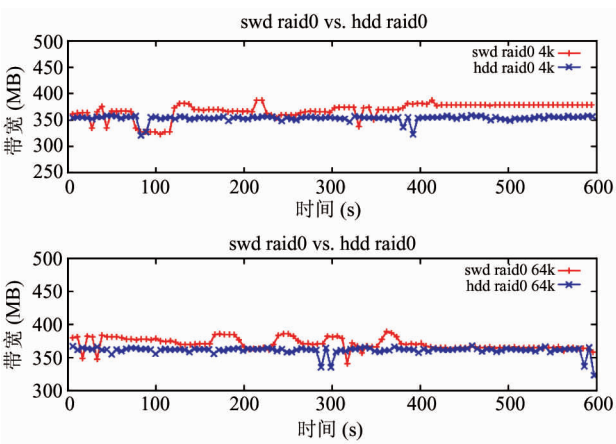

图 5　RAID0 顺序读带宽对比

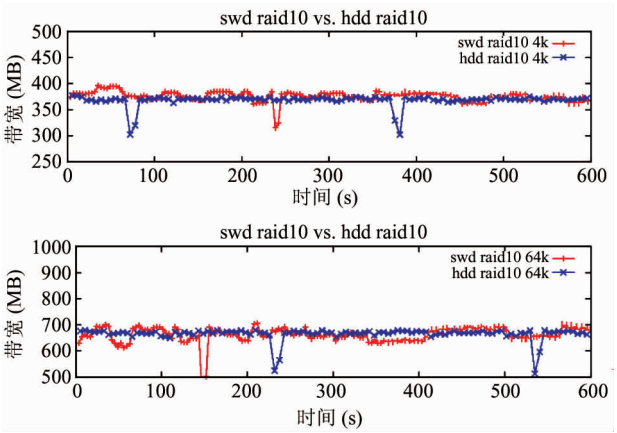

图 6　RAID10 顺序读带宽对比

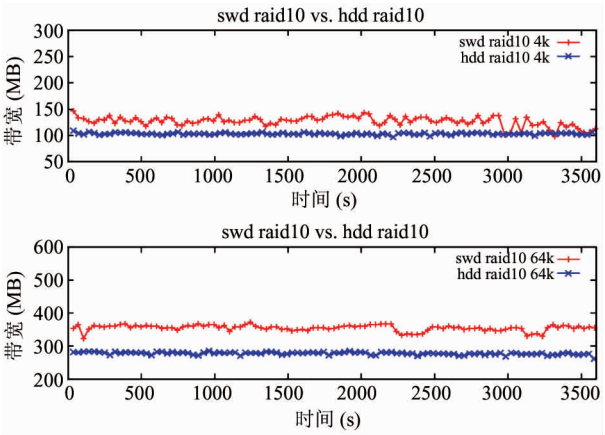

图 9　RAID10 顺序写带宽对比

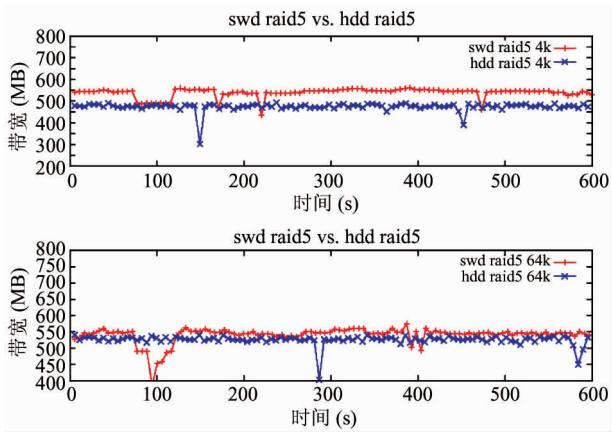

图 7　RAID5 顺序读带宽对比

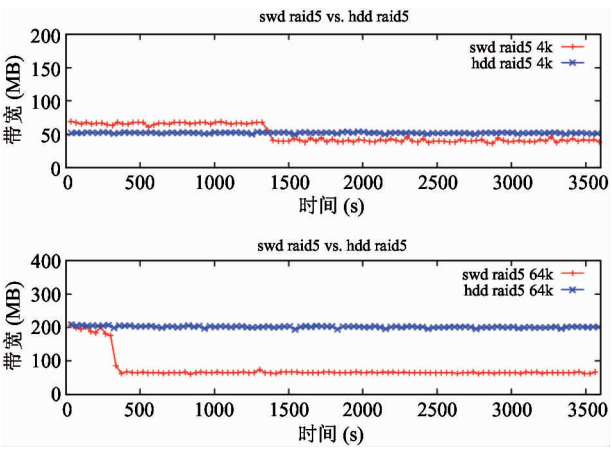

图 10　RAID5 顺序写带宽对比

3.6　RAID0\10\5 Fio 顺序写测试

测试对比了 RAID0\10\5 在 IO 粒度为 4kB 和 64kB 时的顺序写带宽(图 8 ~ 图 10)。

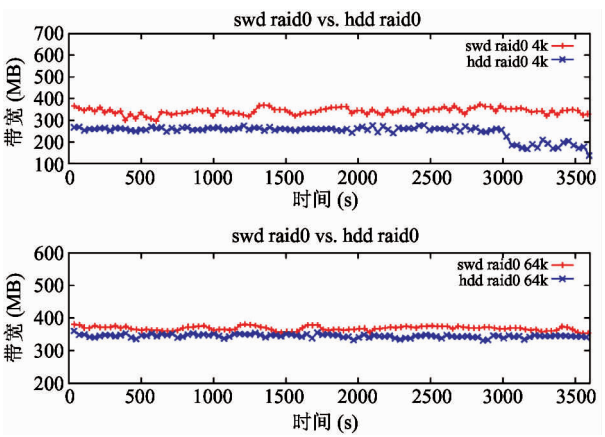

图 8　RAID0 顺序写带宽对比

从测试结果可以看出,SSWD RAID0\10 性能要高于 HDD RAID0\10。对于 RAID0\10,性能高的主要原因依然是磁盘面密度的增大提升了顺序写盘的带宽。

对于 RAID5 来说,测试过程中发现,虽然是顺序 IO,磁盘上依然会有很多读操作,也就是说,对于 RAID5,条带深度为 512kB 来说,4kB 和 64kB 粒度的顺序 IO 不能完全组成全条带写,因此会频繁触发 RIAD5 的读改写或者重构写过程,这是导致 SSWD RAID5 和 HDD RAID5 性能都较低的主要原因。

对于 SSWD RAID5,会有一个带宽突然降低的转折点,这是磁盘启动了回收,从而大幅降低了系统的带宽。对于 RAID5 系统,有一个延迟机制,会等待一段时间凑全条带写,以尽可能避免读改写或者重构写的发生,最大化写带宽。而同时并发执行的

条带数量由测试环境中设置的 Stripe Cache 决定,由于会有多个条带在并发执行,落在单盘上的写请求不一定是完全顺序化的,因此持久缓存会介入到 RAID5 的顺序写过程中,所以当 SSWD 启动回收以后导致整体系统的带宽大幅降低。

3.7　RAID0\10\5 Fio 随机读测试

测试对比了 RAID0\10\5 在 IO 粒度为 4kB 和 64kB 时的随机读带宽(图 11 ~ 图 13)。

从测试结果可以看出,SSWD RAID0 4KB 的随机读性能与 HDD RAID0 差不多,SSWD RAID0 64KB 的随机读性能比 HDD RAID0 差一些,SSWD RAID5 的随机读性能也比 HDD RAID5 要差一些。原因可能跟单盘随机读类似,主要是检查持久缓存区带来的性能开销。

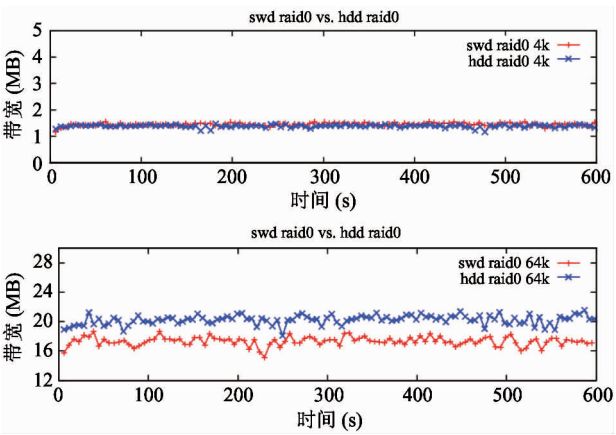

图 11　RAID0 随机读带宽对比

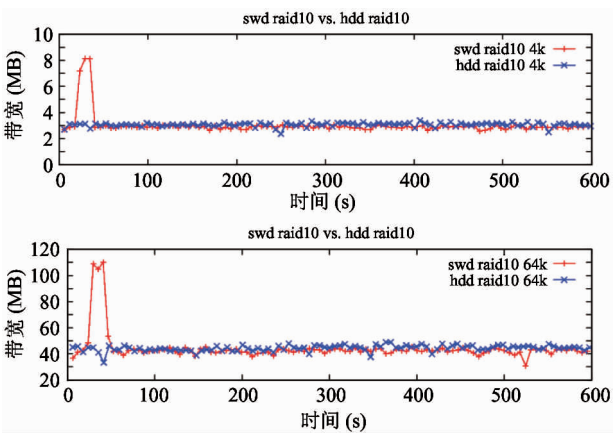

图 12　RAID10 随机读带宽对比

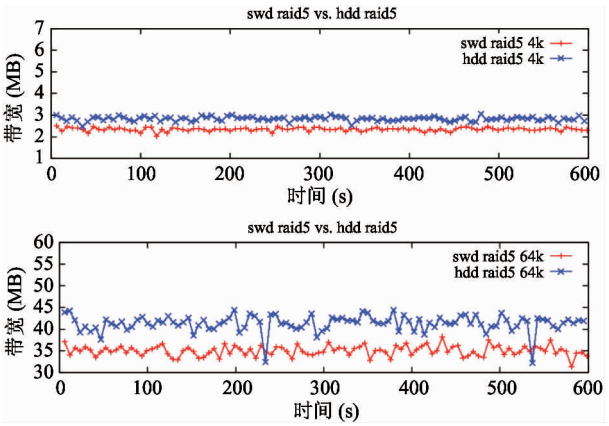

图 13　RAID5 随机读带宽对比

3.8　RAID0\10\5 Fio 随机写测试

测试对比了 RAID0\10\5 在 IO 粒度为 4kB 和 64kB 时的随机写带宽(图 14 ~ 图 16)。

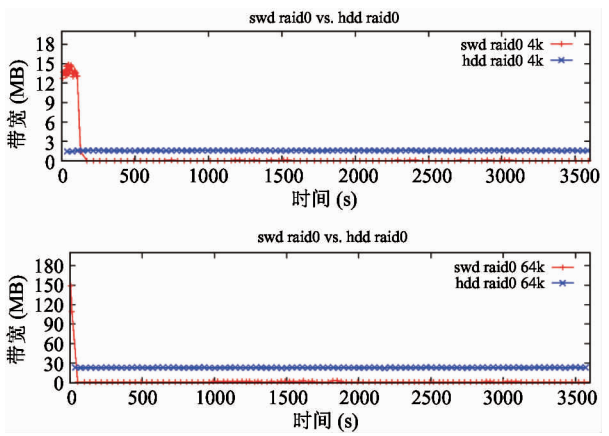

图 14　RAID0 随机写带宽对比

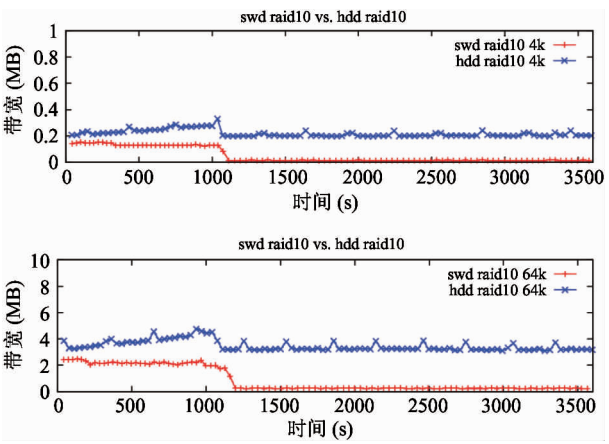

图 15　RAID10 随机写带宽对比

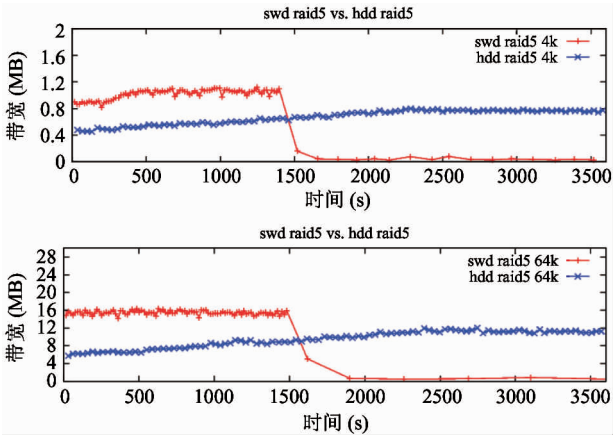

图 16　RAID5 随机写带宽对比

从测试结果可以看出，对于 RAID0\10\5，起初一段时间落盘的随机 IO 都被 SSWD 的持久缓存吸收，随后带宽会大幅下降，也说明了 SSWD 启动回收以后，对系统整体随机写性能产生了非常大的影响。对于 SSWD RAID5 来说，虽然起初随机写会落入持久缓存区，但是由于小写会引入随机读操作，因此对于单盘来说，磁头依然会在全盘和持久缓存区之间大范围抖动，因此即使启动回收前的随机写都可以写入持久缓存区，但是总体性能依然不高。

4　Filebench 测试

4.1　Trace 选取

测试选取了默认的 4 种读写混合的应用场景下的 IO。文件系统为 Ext4，测试运行时间为 1 小时：

- Webserver

模拟了简单的 web-server 的 IO 活动，IO 为目录树下多个文件的打开-读取-关闭操作，还包括日志文件的追加写操作。

- Fileserver

模拟了简单的 file-server 的 IO 活动，工作负载为对目录树的一系列创建、删除、追加、读、写和属性操作。

- Varmail

模拟了简单的 mail-server 的 IO 活动，每封邮件存储在一个单独的文件中，工作负载采用多线程发起创建-追加-同步、读-追加-同步、读和删除操作。

- Webproxy

模拟了简单的 web-proxy-server 的 IO 活动，其中包含对目录树种文件的一系列创建-写-关闭、打开-读-关闭和删除操作，以及文件追加操作模拟 proxy log。

4.2　测试

选取的 4 个 trace 都是读写混合的随机 IO，从图 17 所示的测试结果可以看出，整体上 SSWD 的 IO 延迟会比普通 HDD 高很多。对于 webserver，从 1000s 开始读写 IO 延迟非常低，可能是因为 trace 读请求占比高，写请求基本落入了 SWD 的持久缓存区，因此写延迟较低。而对于大多数读请求，会命中持久缓存区，因此相对于普通 HDD 来说，一方面持久缓存区吸收了大部分随机写请求，另一方面，持久缓存区也提高了读请求的命中率，从而将随机读的范围控制在了持久缓存区内，减小了磁头抖动程度，因此降低了 IO 的响应延迟。

对于 fileserver、varmail，在 500s 之前 SSWD 的 IO 延迟明显低于 HDD，原因是这段时间的随机写请求都写入了 SSWD 的持久缓存区，并且对于读请求来说，持久缓存区提升了读命中率，从而将随机读请求控制在了持久缓存区的范围内，减小了磁头的抖动，而 500s 以后 IO 延迟急剧攀升，说明 SSWD 的持久缓存区已经被写满，磁盘做回收的开销加大了应用的 IO 延迟。而对于 HDD 来说，IO 延迟一直较稳定。测试结果一方面说明了回收对 SSWD 性能的影响程度，另一方面也说明了 fileserver 和 varmail 是以随机写 IO 为主且 IO 密度较高的 trace。

对于 webproxy 来说，2800s 之前 SSWD 的延迟较稳定，但是比 HDD 要高，可能是因为 Trace 为读写请求混合，虽然写请求可以直接写入持久缓存，但是部分读请求需要访问磁盘上的带，因此增大了平均 IO 响应延迟。从 2800s 开始 IO 延迟急剧攀升，说明 SSWD 的持久缓存区已经被写满，磁盘启动了回收。

从 Filebench 的测试结果来看，对于读写混合且 IO 压力较大的场景，SSWD 的性能表现不佳，且非常不稳定。

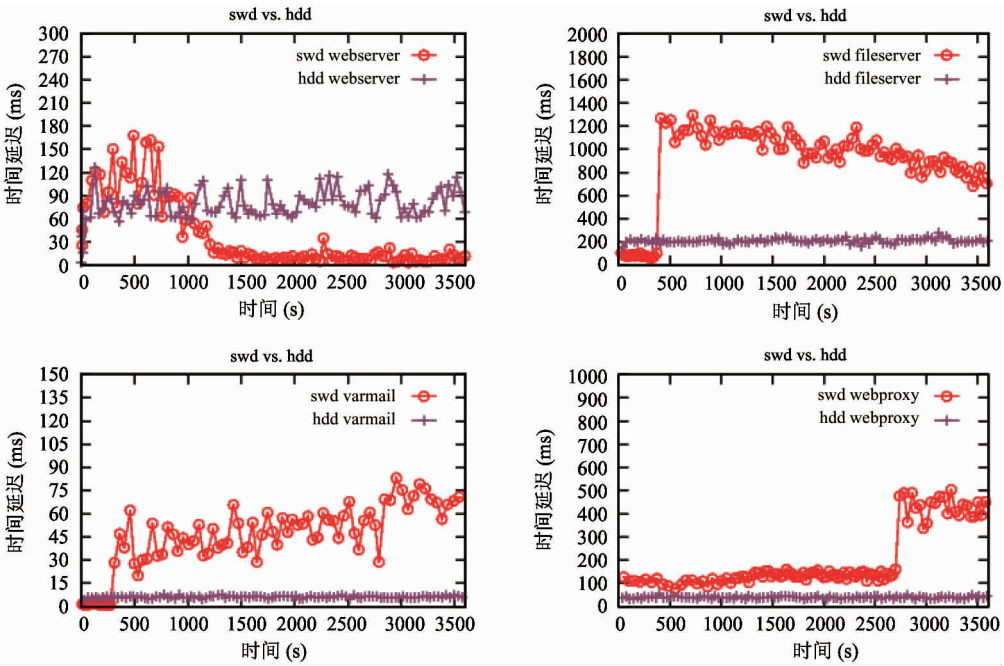

图 17　Filebench 请求延迟对比

5　Trace 测试

本文选取了 4 种真实场景下的应用作为测试用例(如表 2 所示),选择的用例来自于微软剑桥研究院的企业服务器。表 2 中,Trace 为应用名称,Tot. Reqs 表示应用总的请求数,W. percent 表示写比例,W. Size 表示总的写入数据量, W. Update 表示写更新比例,Avg. Len 表示平均请求长度。

表 2　应用特征

Trace	Tot. Reqs	W. percent	W. Size	W. Update	Avg. Len
stg0	2030915	84.8%	15.1GB	97.4%	11853Byte
hm0	3993316	64.4%	20.5GB	92.0%	8185Byte
prn0	5585886	89.2%	46.0GB	73.1%	11358Byte
prxy0	12518968	96.9%	53.8GB	98.7%	4875Byte

选取的 4 条 Trace 均为写比例较高且 IO 粒度较小的应用,但更新比列较高。对比对象依然是 HDD 和 SSWD。测试结果见图 18。

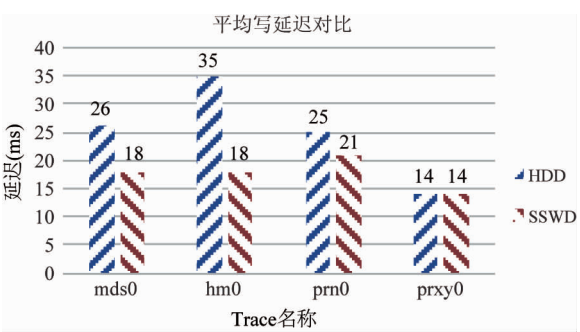

图 18　Trace 延迟对比

从测试结果看,SSWD 的平均写延迟要低于或等于 HDD。也就是说,SSWD 的回收操作并没有对性能带来大幅下降,这与 fio 的随机写测试差别很大,分析其原因有两种可能:

(1) 从表 2 中可知测试的 Trace 的访问空间相对于总数据量来说非常小,有可能 Trace 播放完成后都没有写满 SSWD 的外层持久缓存,也就是回收操作还没有启动,因此写性能较高。

(2) 从表 2 中可知,应用的写更新比例很高,因此对于 SSWD 的外层持久缓存来说,由于其采用的是写顺序化技术,较高的写更新比例会使其自动释放出可用空间,因此持久缓存自身可以做到高效的

回收,从而减小或者延缓下刷操作的发生频率,从而提升了写性能。

6 结 论

通过压力测试和应用模拟测试可以看到,希捷 SWD 在顺序读写方面性能较好,而随机读写方面由于需要检查请求是否命中持久缓存,引入了从 Map 区域读取地址映射表的开销,因此导致磁盘整体性能有所下降。对于 SSWD RAID,顺序读性能依然较好,但是在顺序写场景下,性能不佳,特别是由于 SSWD RAID5 会产生读改写或重构写,导致单盘上的 IO 顺序性降低,因此性能会大幅降低。从对盘阵的随机写测试可以看出,单盘回收带来了巨大的性能影响,因此如果想提升 SSWD RAID 的非顺序写性能,需要对现有 RAID 系统重新做设计,以尽量利用 SSWD 的磁盘特征,尽可能避免回收带来的性能影响。

通过 Filebench 的模拟测试可以看到在非顺序写压力不大的情况下,如 webserver,SSWD 可以表现出较好的性能,而对于写压力较大的场景,如 fileserver、varmail 和 webproxy,当 SSWD 持久缓存被写满启动回收以后,性能会大幅降低,因此也能说明希捷 SSWD 只适合于 IO 非密集或者写入时间间隔较大的应用场景。

通过选取的 4 条 Trace 测试结果可以看出,对于更新比例很高的应用,SSWD 的写性能并不会像压力测试那样大幅下降,持久缓存的利用率较高,性能较好。

希捷将其 SWD 定位为适用于桌面应用和动态归档场景。从本文的一系列测试也可以验证这一点。对于顺序性较好的场景,例如动态归档、视频存储等,IO 会绕过 SSWD 的持久缓存区,顺序地写入磁盘的带中,且不会产生写覆盖,因此可以最大化 SSWD 的写性能;对于桌面应用等 IO 密度较小的场景以及企业级写更新比例较高的场景, SSWD 依然可以保持较好的写性能。

参考文献

[1] Richter H J, Dobin A Y, Heinonen O, et al. Recording on bit-patterned media at densities of 1 Tb/in^2 and beyond. *IEEE Transactions on Magnetics*, 2006, 42(10): 2255-2260

[2] White R L, Newt R M H, Pease R F W. Patterned media: a viable route to 50 Gbit/in^2 and up for magnetic recording? *IEEE Transactions on Magnetics*, 1997, 33(1): 990-995

[3] Kryder M H, Gage E C, McDaniel T W, et al. Heat assisted magnetic recording. *Proceedings of the IEEE*, 2008, 96(11): 1810-1835

[4] Zhu J G, Zhu X, Tang Y. Microwave assisted magnetic recording. *IEEE Transactions on Magnetics*, 2008, 44(1): 125-131

[5] Greaves S, Kanai Y, Muraoka H. Shingled Recording for 2-3 Tbit/in^2. *IEEE Transactions on Magnetics*, 2009, 45(10): 3823-3829

[6] Amer A, Holliday J A, Long D D E, et al. Data management and layout for shingled magnetic recording. *IEEE Transactions on Magnetics*, 2011, 47(10): 3691-3697

[7] Amer A, Long D D E, Miller E L, et al. Design issues for a shingled write disk system. In: Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, USA, 2010. 1-12

[8] Aghayev A, Shafaei M, Desnoyers P. Skylight—a window on shingled disk operation. *ACM Transactions on Storage (TOS)*, 2015, 11(4): 1-28

Test and evaluation of Seagate SWD

Zhang Qiang[*,**], Liu Zhenjun[*], Dong Huanqing[*], Ma Liuying[*,**], Li Mengkun[***]

([*]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

([**]University of Chinese Academy of Sciences, Beijing 100049)

([***]School of Management, Capital Normal University, Beijing 100045)

Abstract

The advantages and limitations of the shingled magnetic recording (SMR) technique which can enhance the disk storage density were analyzed, and it was pointed that SMR adopts the way of partially overlapping adjacent tracks to increase the areal density of the disk, but this brings the write overlay problem, that is when updating a track's data, adjacent tracks' effect data can be covered, thus leading that the shingled write disk (SWD) can not support in-place update and the non-sequential writing performance can be affected. To promote the application of this SMR technique, the testing tools of Fio and Filebench were used to more comprehensively test the Seagate's shingled write disk (SSWD), a representative product of the SMR technique, and its performance in variety of application scopes was analyzed to privide useful references for SSWD's applications.

Key words: shingled magnetic recording (SMR), shingled write disk (SWD), write overlay, in-place update