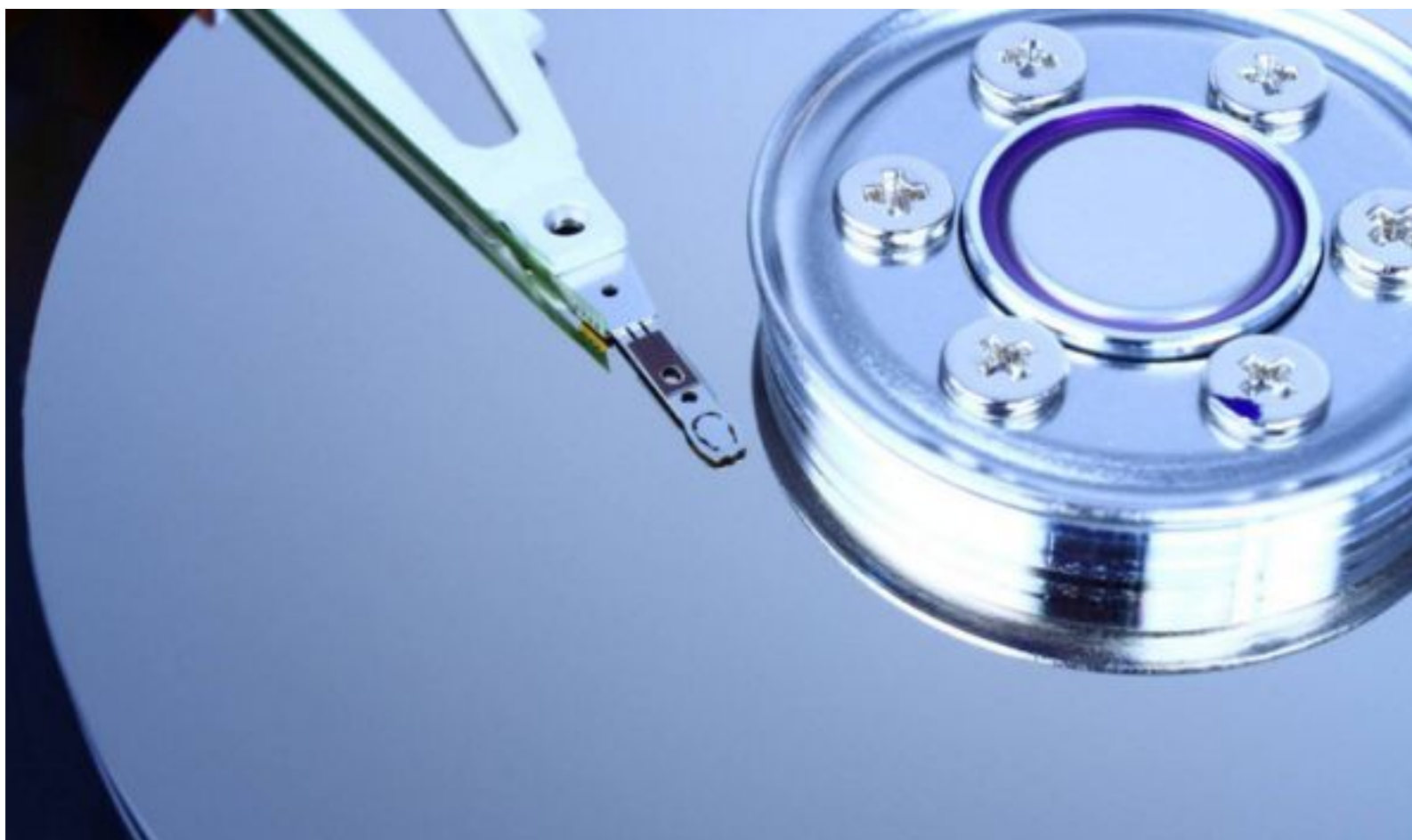


详解叠瓦式磁记录SMR盘基础知识

原创：岁月葱葱 架构师技术联盟 2018-05-25



↑ 点击蓝字，轻松关注

SMR(Shingled Magnetic Recording)叠瓦式磁记录盘是一种采用新型磁存储技术的高容量磁盘。SMR盘将盘片上的数据磁道部分重叠，就像屋顶上的瓦片一样，这种技术被称为叠瓦式磁记录技术。该技术在制造工艺方面的变动非常微小，但却可以大幅提高磁盘存储密度。

在数据量飞速增长的当今世界，SMR技术可以有效降低单位容量的磁盘存储成本，是未来高密度磁盘存储技术的发展潮流。

尽管SMR盘的读行为和普通磁盘相同，但它的写行为有了巨大的变化：不再支持随机写和原地更新写。这是由于SMR盘上新写入的磁道会覆盖与之重叠的所有磁道，从而摧毁其上的数据。换言之，相较传统磁盘而言，SMR盘不再支持随机写，只能进行顺序追加写。写入方式的限制给欲使用SMR盘的存储系统带来了巨大的挑战。

SMR盘的技术原理

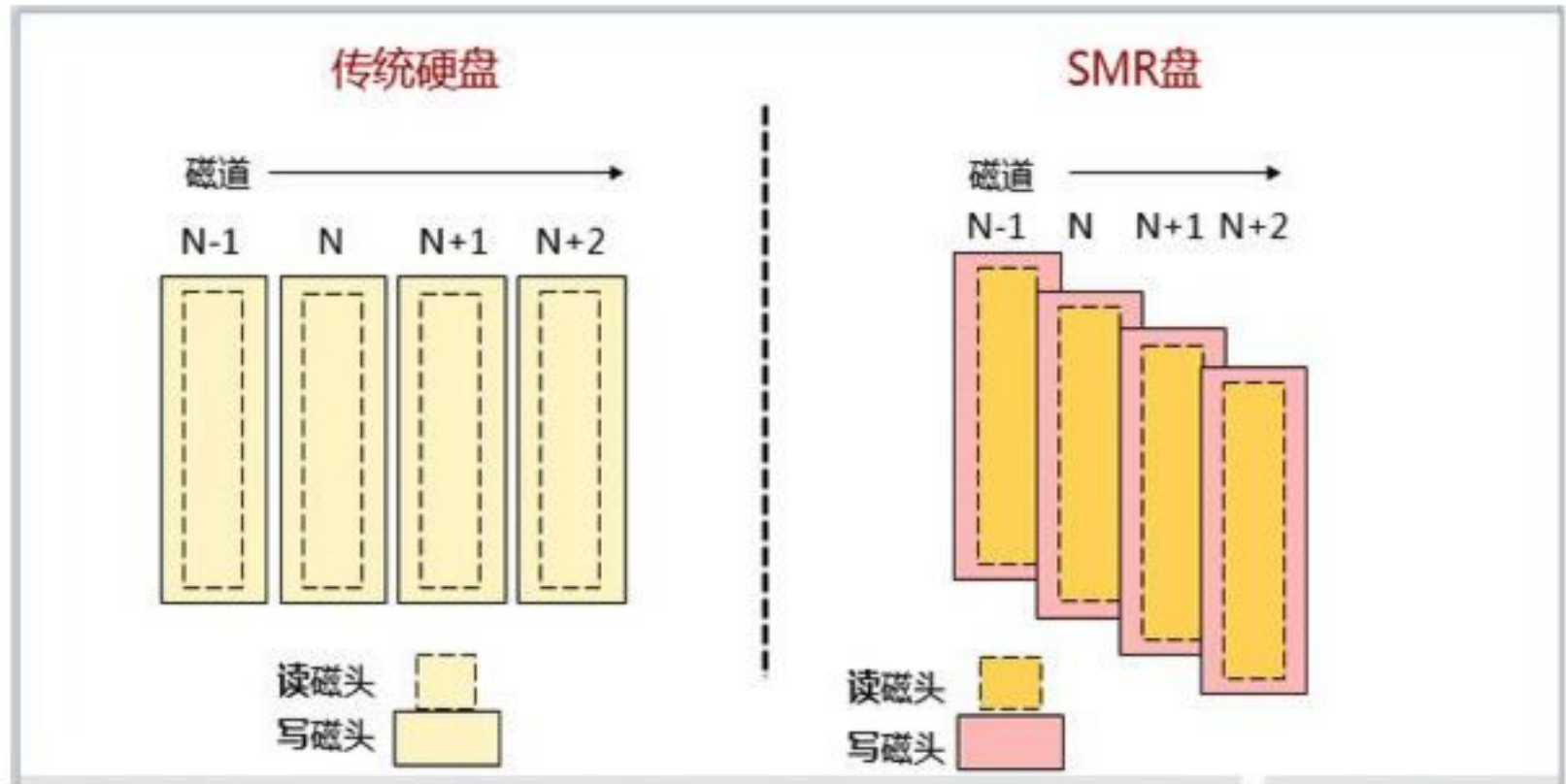
大数据时代的数据量呈指数级增长，预计每两年就会翻一番。硬盘作为一种经济高效的存储介质，在当今的存储数字世界中依然起着中流砥柱的作用。



硬盘驱动器的存储面密度(每单位面积存储的位数)受到物理定律的限制，目前硬盘使用的**PMR垂直磁记录**(Perpendicular Magnetic Recording)技术即将达到其存储密度极限。业界迫切希望引入新的技术来克服限制，继续稳步提升容量。

SMR盘是领先的下一代磁盘技术。SMR盘在制造工艺方面的变动非常微小，因为它保持现有磁头和介质技术不变，它通过将磁道像屋顶上的瓦片那样重叠在一起来实现其存储面密度的增长。

具体来说，硬盘盘片被划分为同心圆状的磁道，磁道与磁道之间具有一定间隔，起到保护作用。硬盘的磁头位于盘片记录介质的上方，包含用于写入的写磁头和用于读取的读磁头，二者是独立的部件。在磁盘上想要写入数据，需要写磁头施加较强的磁场来改变盘片上的记录，这就进一步要求写磁头足够大以产生所需的写入磁场强度。因此，如果磁道没有任何重叠，磁盘上每英寸能容纳的磁道数量由写磁头的宽度所决定。



传统硬盘与SMR盘磁记录方式的对比示意图

叠瓦式写入则利用了磁盘读取所需要的磁场强度低于写入所需要的磁场强度的事实。换言之，磁盘读取所需的磁道宽度可以小于写入所需的轨道宽度，读磁头的宽度可以小于写磁头宽度。SMR盘上数据的写入是通过部分重叠前一条磁道写入的，同时留出了足够的空间给较窄的读磁头读取前一条磁道的数据。叠瓦式写入的结果就是留给读用的窄磁道，以及最后用于写入的较宽的写磁道，写入将会覆盖若干个读磁道的宽度。

和传统磁盘相比，SMR盘通过叠瓦式写入在相同的面积内容纳了更多的磁道，从而提高了存储面密度，进而降低了磁盘单位存储容量的成本。

SMR磁盘将磁道划分为若干个带(Band)，即由连续磁道所构成的可连续写入区域，每个区域构成一个需要顺序写入的基本单元。Band是SMR盘上的物理概念，其对应的逻辑概念是“区域”(zone)，是上层软件所能看到的SMR盘的顺序写区域。一个zone的大小通常在256 MB量级。由于SMR盘仍然确保可以从磁道的未重叠部分读取数据，zone内的数据读取依旧可以随机进行。

SMR盘在生产制造时就划分好了分区，通过在zone之间留出较大的间隔来标识zone的起始位置。此外，硬盘厂商在SMR盘上一小块区域内，通过采用传统硬盘非重叠的磁道组织方式，牺牲了部分存储密度以换取在这些zone内随机写入的能力。这部分zone被称为“传统区”(Conventional Zone)，为上层应用的随机写提供了一个落脚点。

在传统硬盘上，存储在磁盘上的用户数据由文件系统来管理，文件系统假定磁盘

提供的块接口中，块是独立可更新的单元。但是，由于叠瓦式写入导致随机写和原地更新写不再受支持，SMR盘需要全新的数据管理解决方案。寻找一种方法来高效地存储、检索和管理SMR盘中这些大块顺序写区域中的数据，而不会导致太多空间浪费是非常具有挑战性的。

SMR盘的分类

SMR盘采用叠瓦式记录通过将zone内的磁道按序部分重叠，获得了更高的存储面密度、更大的磁盘容量、更低的单位容量价格。在获得这一好处的同时，SMR盘也对上层应用提出了一些限制，包括SMR盘在zone内只能顺序写，不支持随机写，SMR盘的空间回收以一个Zone的大小(例如256MB)为粒度。

SMR盘对写入的约束可在硬件侧由硬盘本身来管理，亦可在主机侧由软件来进行。前者被称为“驱动器自管理的SMR盘”(Drive Managed SMR disk)，以下简称DM SMR盘。后者则将SMR内部的组织结构通过新增接口向上层软件暴露。

制定SCSI协议的T10国际组织和制定ATA协议的T13国际组织都针对SMR盘对SCSI和ATA协议进行了相应的拓展。

- **T10**组织引入了**ZBC**(Zoned Block Commands)，定义了SMR盘相关的术语和行为，成为SAS接口SMR盘所必须遵守的规范。
- **T13**组织引入了**ZAC**(Zoned-device ATA Commands)，与ZBC类似，是SATA接口SMR盘所必须遵守的标准。

ZBC和**ZAC**标准将SMR盘抽象成为一组Zone的集合，其中的**Zone**由一段连续的**LBA**构成。ZBC与ZAC标准中均定义了SMR盘中可以存在的三种类型的Zone：

- 传统**Zone**(Conventional Zone简写为C-Zone)
- 首选顺序写入**Zone**(Sequential-Preferred Zone简写为SP-Zone)
- 强制顺序写入**Zone**(Sequential-Required Zone简写为SR-Zone)

Zone类型	缩写	写入限制	设计用途
传统Zone (Conventional Zone)	C-Zone	无	随机写 原地更新写
首选顺序写入Zone (Sequential-Write-Preferred Zone)	SP-Zone	无	大体上顺序写
强制顺序写入Zone (Sequential-Write-Required Zone)	SR-Zone	必须顺序追加写	全部顺序写

对C-Zone而言，和传统硬盘一样，它具有完全的随机写能力，对其中写入IO的逻辑块地址(LBA)的顺序不做任何限制。C-Zone所占的大小一般不超过SMR盘容量的1%。

SP-Zone也不对写入顺序做强制性限制，但是，和C-Zone不同的是，SP-Zone位于SMR盘的顺序写入区域，最适合顺序写。SP-Zone引入了写指针(Write Pointer，以下简称WP)的概念，该指针等于该Zone内的最大LBA，应用程序应该首选在SP-Zone内的写指针处写入数据。尽管该硬盘本身也能够处理非写指针处的写入(随机写)，但这是有代价的，会引起后续的数据迁移、垃圾回收等。

对于每个Zone，SR-Zone也暴露了写指针WP的位置，但硬盘设备将拒绝写入除写指针指向的LBA以外的任何位置，也就是SR-Zone内必须以写指针当前所在位置为起点，进行顺序追加写。

当SR-Zone写满之后，则需通过Reset Write Pointer操作将该WP重置到该Zone开头，即清空该Zone以回收空间，从头开始写入。SP-Zone和SR-Zone被统称为写指针区(Write Pointer Zone)。

如前所述，如果一块硬盘采用了叠瓦式写入技术，但是对外不暴露Zone和WP，由硬盘固件本身来管理写入的顺序性约束，我们称之为Drive Managed SMR。

如果一块硬盘采用了叠瓦式写入技术，对外暴露了Zone和WP，并且所包含的Zone类型仅仅含有C-Zone和SR-Zone，那么我们称之为Host Managed SMR(主机管理的SMR盘)。

如果一块硬盘采用了叠瓦式写入技术，对外也暴露了Zone和WP，并且所包含的Zone类型由C-Zone和SP-Zone组成，那么我们称之为Host Aware SMR(主机感知的SMR盘)。

HA SMR盘可以视为**DM SMR盘**和**HM SMR盘**的折中类型。

SMR盘类型	描述
DM (Drive Managed)	DM SMR盘采用和传统硬盘相同的接口，对外隐藏SMR盘的细节。软件无需改动，但是性能可预知性不好，受IO负载影响大。
HA (Host Aware)	HA SMR盘对主机软件提供了SMR盘相关的新接口，使得软件可以根据SMR盘的特点进行针对性的IO优化。接口向前兼容。若软件进行了非顺序写，SMR盘的性能会受到影响。性能可预知性一般。
HM (Host Managed)	HM SMR盘提供了SMR盘相关的新接口，要求上层软件必须根据SMR盘的特点进行顺序写IO操作。接口不向前兼容，不符合SMR盘写入规则的IO将会被拒绝。性能可预知性好。

上表总结了三种SMR盘的特点。虽然DM SMR盘采用和传统硬盘相同的接口，在内部实现一个叠瓦转换层STL(Shingle Translation Layer)对外隐藏了顺序写入的限制，但它也有一个主要缺点：

在某些带有随机写的工作负载下，DM SMR盘内顺序写入区域需要进行数据迁移、垃圾回收等背景操作，DM SMR盘的性能可能受到不可预测的影响，性能波动巨大。

相比之下，HM SMR盘不将自身伪装为传统磁盘，而是根据ZBC/ZAC标准对外暴露了zone和WP的抽象结构，允许上层软件来完全管理SMR盘的顺序写入约束，对不符合顺序写规则的IO请求则直接返错。通过向上暴露新接口，HM SMR盘的IO行为完全受软件控制，因而其性能是可控的，能最大化发挥SMR盘的优势。

但劣势是上层的软件必须进行修改才能使用HM SMR盘。HA SMR盘提供了和HM SMR盘一样的新接口，但是当其收到不符合顺序写原则的IO是，HA SMR盘将允许其写入而非报错，具有一定的灵活性。

无论采用哪一种类型的SMR盘，我们都需要将上层应用的IO行为特点与SMR盘顺序写的内在特点相匹配，否则就无法使SMR盘工作在最佳状态。

点击[原文链接](#)获取已整理成文的<[传统企业存储知识完全解析](#)>电子书。



热文阅读

- [埃森哲全球“技术展望2018”报告解析\(附下载链接\)](#)
- [容器原生存储: 红帽Gluster容器化技术解析\(附实战视频\)](#)

温馨提示：

请搜索“[ICT_Architect](#)”或“[扫一扫](#)”二维码关注公众号，点击[原文链接](#)获取更多技术资料。



Stay hungry Stay foolish



发现更多精彩 点击箭头下方
[阅读原文](#)

[阅读原文](#)