TOSHIBA

Whitepapper

关于在 RAID 系统中混合使用 512n 和 512e 硬盘的探讨

https://toshiba-semicon-storage.com/cn/storage.html

前言

用于 NAS 和企业存储的硬盘驱动器 (HDD) 有两种不同类型的扇区大小。传统的 HDD 内部扇区大小为 512 字节，外部接口也采用 512 字节，这项技术称为 512 字节原生，简称 512n。较新的 HDD 型号采用 4k 字节的内部扇区大小来提高存储效率，通过模拟仿真处理仍然可以在外部接口上提供 512 字节格式，该技术称为 512 字节模拟 (512e)。此外，内部 4k 字节扇区格式的 HDD 配合 4k 字节原生接口也是可行的，但这类 HDD 不太常用，一般仅用于一些特定的企业级应用。

除了较旧的 HDD 型号采用 512n，现代 HDD 出于向下兼容的考量，4TB 以下容量的型号仍旧提供 512n 格式，超过 4TB 的现代 HDD 则均为 512e。

如 smartmontool (https://www.smartmontools.org/) 等诊断工具显示的如下配置：

```
>smartctl -a /dev/sdx

Device Model:    TOSHIBA HDWG440
Sector Size:     512 bytes logical/physical                ( = 512 native )

Device Model:    TOSHIBA HDWG480
Sector Size:     512 bytes logical, 4096 bytes physical    ( = 512 emulated )

Device Model:    TOSHIBA MG06ACA10TA
Sector Size:     4096 bytes logical/physical               ( = 4k native )
```

在同样的硬件代次中，512e 的 HDD 通常比 512n 的稍稍快一些，因为 512e 中的辅助控制位数较少，因此相同转速下的数据传输率更高（512n - 东芝 MG08ADA400N：233 MB/s vs 512e - 东芝 MG08ADA400E：255 MB/s）。

我们的评测

在东芝 HDD 实验室中，我们评估了在 RAID 系统中 512n 与 512e 硬盘互换的可行性。RAID(Redundant Array of Independent Disks)将一批物理硬盘组合成一个更大的阵列，通常具有一定的冗余性（奇偶校验或镜像）。如果其中一个硬盘发生故障，RAID 控制器可以利用奇偶校验或镜像信息将阵列信息重建到替换上的新硬盘中。

随着 512e 硬盘的出现，我们很想知道：512n 的硬盘阵列是否可以接受用 512e 的硬盘来替换故障硬盘？为了确认两种格式是否能够自由混用，我们还进行了反向测试（用 512n 来替换故障的 512e 格式硬盘）。

HDD 设置：512n 和 512e

在评测中，我们使用流行的东芝 NAS HDD N300 作为 512n 扇区格式的 4TB 版本，使用 MG 系列企业级型号 MG08ADA400E 作为 512e 扇区格式版本。这两个都是常见型号，且 4TB 这个容量点常被用于配置 2 盘位或 4 盘位的小型 RAID 组。

我们通过一个简易的方式模拟故障硬盘，即在系统全功率正常运行的状态下拔下其中一块硬盘（热插拔），并等待 10 分钟，同时检查存在一块故障硬盘模式下 RAID 阵列的性能和功能。然后同样在全功率状态下重新插入替换硬盘，并等待 RAID 重建，整个过程不做任何电源重启动作。

图 1：
512n HDD (N300 4TB)

图 2：
512e HDD (企业级 MG08ADA400E)

RAID 系统设置：4 盘位桌面 NAS 和内部 RAID 控制器

RAID 配置有两种最常见的实际应用，一种是NAS(Network Attached Storage) 设备，另一种是通过 RAID 控制器 结合PCIe 扩展卡驱动的企业级服务器中的存储子系统。
我们的评测基于两种常见的 4 盘位桌面 NAS 型号：QNAP 4 盘位桌面 NAS TS-464-4G 和 Synology 4 盘位桌面 NAS DS-420+。

图 3：
QNAP 4 盘位桌面 NAS TS-464-4G

图 4：
Synology 4 盘位桌面 NAS DS-420+

至于服务器 RAID 子系统，配合硬盘前置服务器机箱，我们选取了几种流行的 RAID 控制器进行评测。第一个是 Microchip Adaptec® 的 SmartRAID 3204。

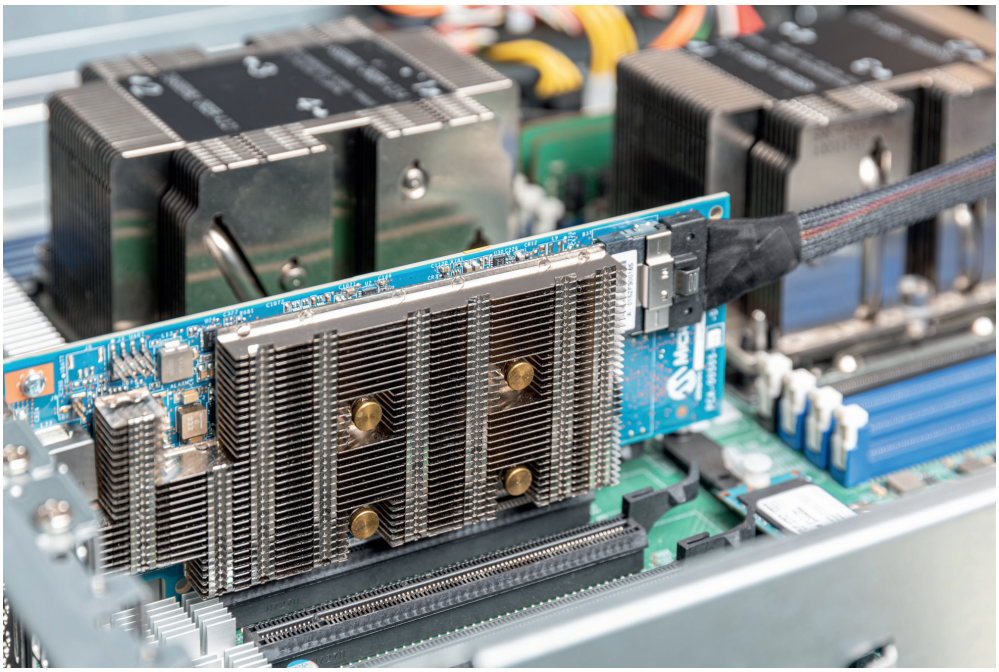

图 5：安装在服务器中的 Microchip Adaptec® SmartRAID 3204

由于此 RAID 控制器被优化为仅针对较少盘位系统，因此我们还评测了上一个系列中的多用途控制器：Microchip Adaptec® SmartRAID 3154。

此外，我们还评测了另一个来自 Broadcom 的被广泛使用的 RAID 控制器：MegaRAID SAS 9460。

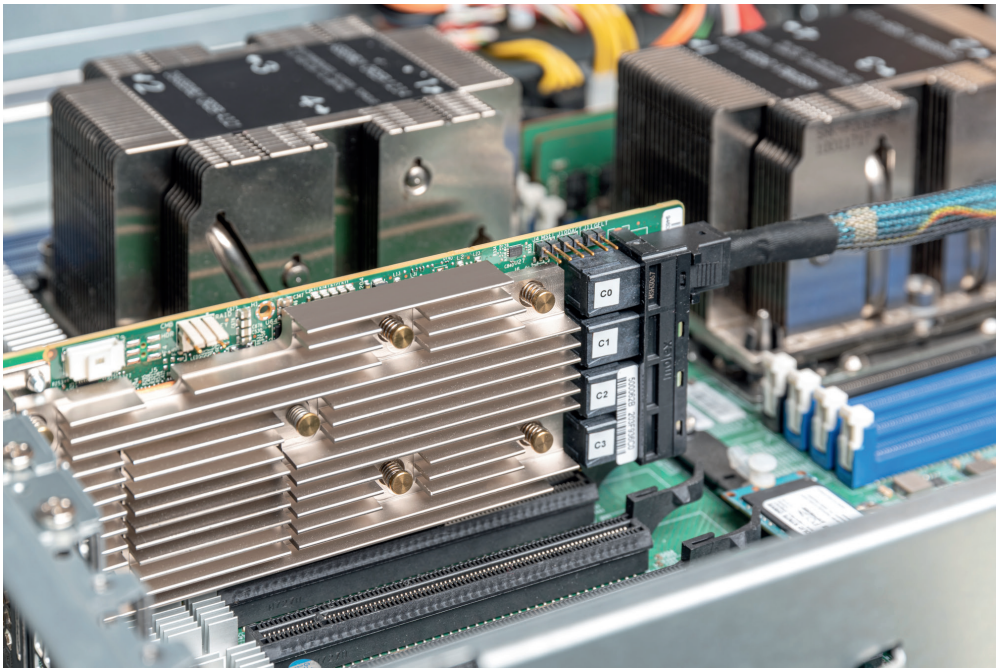

图 6：安装在服务器中的 Broadcom MegaRAID SAS 9460

然而，由于该型号已经停产，我们还测试了它的后继型号 MegaRAID SAS 9560（带 PCI-Express Gen4 主机接口），结果几乎一致（见表 1）。

512n RAID 组中替换为 512e HDD 的结果

System	Model	Initial Array	Replace Drive	Function	Rebuild Time	Post-processing
Broadcom	9460	512n	512e	YES	6h 5min	None
Broadcom	9560	512n	512e	YES	5h 45min	None
Adaptec	3154	512n	512e	YES	4h 15min	8h BPI
Adaptec	3204	512n	512e	YES	4h 5min	8h BPI
Synology	DS-420+	512n	512e	YES (1/2)	4h 0min	6h Scrub
QNAP	TS-464-4G	512n	512e	YES	6h 0min	None

(1) automatic rebuild does not start, need manual interaction
(2) System issues a warning that MG08ADA is not in the compatibility list, but rebuild still works

表 1：512n RAID 组中替换为 512e HDD

结论：对于以上所有配置，都可以用 512e 型号来替换故障的 512n 硬盘，没有限制。

关于重建时间的观测：

测试结果证实了不同 RAID 引擎对于重建策略的不同。对于 Broadcom 和 QNAP，整个硬盘都将被重建：不管它是否有写满数据。不过重建完成后，不再需要执行进一步的一致性检查。

Adaptec 和 Synology 则采用不同方法。它们只重建硬盘上的实际数据，因此对于未写满数据的阵列，单纯的重建时间明显更短。RAID 重建后会执行后续处理任务（后台奇偶校验初始化 (BPI) 或数据清理）。此时 RAID 组不再脆弱或被定义为降级，但后续处理任务需要占用一些性能资源（但它会以低优先级运行）。

作为参考，表 2 列出了一些条件和配置的详细信息：

Details	
512e Drive	Toshiba MG08ADA400E 4TB FW 0102
512n Drive	Toshiba N300 4TB HDWG440 FW 0601
Broadcom	MegaRAID SAS 9460-16i FW 5.010.00-0671
Broadcom	MegaRAID SAS 9560-8i FW 5.120.02-2904
Adaptec	SmartRAID 3154-16i FW 1.60
Adaptec	SmartRAID 3204-8i FW 03.01.17.056
Synology	DS-420+
QNAP	TS-464-4G
RAID Conf.	RAID5, Cache On, Fully initialized, partially filled (6TB Data)
Rebuild Time	Without Workload

表 2：条件和配置

我们使用 RAID5（奇偶校验 RAID），因为这是比基于简单镜像（RAID1、RAID10）的一类 RAID 更复杂的配置。重建时间基于部分写入数据（未写满）的阵列评估，并且在重建期间没有实际业务工作负载。需要注意的是，如果重建期间存在工作负载，则重建的时间可能显著加长。

512e RAID 组中替换为 512n HDD 的结果

我们还确认了用 512n HDD 替换 512e RAID 组中出现的故障盘在表 3 的所有配置下均可奏效。结合之前的结果，证实了列出的所有 RAID 配置都可以混合使用 512n 和 512e。

System	Model	Initial Array	Replace Drive	Function	Rebuild Time	Post-processing
Broadcom	9460	512e	512n	YES	6h 0min	None
Broadcom	9560	512e	512n	YES	6h 0min	None
Adaptec	3154	512e	512n	YES	4h 40min	8h BPI
Adaptec	3204	512e	512n	YES	4h 10min	8h BPI
Synology	DS-420+	512e	512n	YES (1/2)	3h 30min	6h Scrub
QNAP	TS-464-4G	512e	512n	YES	6h 30min	None

表 3：512e RAID 组中替换为 512n HDD

更灵活：用更大容量或其他品牌硬盘更换故障硬盘

接下来，我们继续提出进一步的问题。

如果没有现成的初始容量的备件在手，可以用更大容量的硬盘代替吗？结果如下表 4 所示。

System	Model	Initial Array	Replace Drive	Function	Rebuild Time	Post-processing
Broadcom	9460	512n	6TB	YES	6h 10min	None
Adaptec	3154	512e	6TB	YES	4h 0min	8h BPI
Synology	DS-420+	512n	6TB	YES (1/2)	4h 0min	6h Scrub
QNAP	TS-464-4G	512e	6TB	YES	6h 30min	None

Details	
Replace Drive	N300 6TB HDWG460 FW0601

表 4：更大容量的替换硬盘

结论：故障硬盘可以用容量更大的型号替换，不过替换硬盘的额外容量不会被使用。

如果没有可用备件硬盘或原始硬盘厂商不再生产，那怎么办？其他 HDD 制造商提供的类似硬盘可以用作替换吗？请参见表 5。

System	Model	Initial Array	Replace Drive	Function	Rebuild Time	Post-processing
Broadcom	9460	512n	6TB	YES	6h 10min	None
Adaptec	3154	512e	6TB	YES	4h 0min	8h BPI
Synology	DS-420+	512n	6TB	YES (1/2)	4h 0min	6h Scrub
QNAP	TS-464-4G	512e	6TB	YES	6h 30min	None

表 5：其他硬盘厂商的产品

结论：同样可以。无论来自哪个硬盘厂商，只要接口技术 (SATA)相同、扇区大小（512 字节，无论是模拟仿真还是原生）相同，同等或更大容量的 HDD 都可以混合使用。

不过，东芝以前的一份测评[1]也显示混合存在一定的局限性。扇区大小为 512 字节与扇区大小为 4k 字节的HDD不能混合使用或作为备件替换，采用不同接口技术（SATA 和 SAS）的 HDD 也不能混合使用。尽管一些 RAID 控制器也接受 SSD 作为故障 HDD 的替换品，反之亦然，但此类用例的实现更多的是出于学术兴趣，而不是实际解决方案。显然，用较小容量的设备替换故障设备也不可行。

其他 RAID 系统

RAID 技术不仅用于 NAS，也用于 DAS (Direct Attached Storage)，DAS 通过专用 USB 或 Firewire 接口将内建 RAID 的RAID box（硬盘阵列盒）连接到主机。为了确认在 DAS 系统中的替换可行性，我们测试了来自 ICY-Box 的 4 盘位 RAID box 和来自 RAIDON 的 SafeTANK GR5640-SBA31+。

图 7：
ICY-Box 的 4 盘位 RAID IB-RD3640SU3

图 8：
RAIDON 的 4 盘位 RAID GR5640-SBA31+

在这些 USB 接口的 RAID box上完全替换兼容性也得到了验证，如表 6 所示：

System	Model	Initial Array	Replace Drive	Function	Rebuild Time	Post-processing
ICY Box	IB-RD3640SU3	512e	512n	YES	6h 50min	None
		512n	512e	YES	6h 30min	None
RAIDON	GR5640-SBA31+	512e	512n	YES	6h 45min	None
		512n	512e	YES	6h 30min	None

表 6：在 DAS 用的 USB 接口 RAID box 上的替换可行性

东芝意识到，最终用户（数据中心、大型商用、中小型商用和个人用户）可能会长时间仍然使用较旧的传统系统，甚至超出原厂的保修和支持时间期限。尤其是 RAID 控制器，古老的 3ware 型号仍然有很庞大的安装基础。东芝支持基于旧款控制器來测试新型号 HDD[2]，因此我们测试了用 512e HDD 替换早期 RAID 组中故障的 512n HDD，使用到的是已有 10 多年历史的 3ware 9750-4i RAID控制器（参见表 7）。

System	Model	Initial Array	Replace Drive	Function	Rebuild Time	Post-processing
3ware	9750-8i	512n	512e	YES	6h 5min	None

表 7：HDD 在旧控制器上的替换

有趣的是，这个旧控制器的 RAID 重建时间与其他 RAID 系统所需时间在同一范围内。这表明 RAID 重建性能更多的是取决于硬盘的速度，而不是 RAID 控制器的性能。

总结与结论

无论是基于 RAID 的存储、通过 RAID 控制卡连接的服务器，还是 NAS 亦或是经USB 接口的 RAID box，SATA 512 字节的故障HDD都可以用相同或更大容量的任何其它 SATA 512字节的HDD 型号替换，无论是 512 n还是 512e。512n和 512e的HDD可以在任何配置中混合使用。

虽然这是个好消息，但还是有一些要点值得我们注意。

如不备份，数据仍有风险。值得反复强调的是，RAID 技术并非备份技术，它只是帮助您在部分存储介质发生故障时维持系统继续工作。如果 RAID 控制器/阵列盒出现故障或发生其他存储错误，您的数据都将丢失。因此，请务必备份 RAID/NAS 上的数据，最好备份到便携式组件，如东芝 Canvio 2.5 英寸外置 USB 接口 HDD，并将其单独存储在系统之外。

监测您的 RAID。RAID 技术使得系统能够在 一个或多个 HDD 发生故障时自动继续工作，因此您仅通过观察系统的功能和性能是无法注意到故障的，这就是为何 RAID 阵列盒和 NAS 系统配备有 LED 指示灯的原因。对于 RAID 阵列盒，由于点对点连接较短，RAID 盒通常被放置在靠近主机的位置，因此可以轻松观察到。但具有网络连接功能的 NAS 设备可以被安置于网络内的任何位置，通常位于机柜、机架或储藏室中，因此红色闪烁 LED 指示灯很容易被忽视。故建议您定期登录系统的控制面板去检查硬盘状态，更好的做法是对系统配置进行设置，以便故障信息可以通过电子邮件等方式通知到您。

时刻为硬盘故障做好准备。请提前准备好备件硬盘，以防有硬盘出现故障。理想情况下，备件型号应该与原始 RAID 组中的硬盘型号一致；但正如本白皮书所示，备件的选择上可以有很大的灵活性。只需确保它是 SATA 型号，具有相同或更高的容量，并且不是 4kn 原生扇区大小即可。

保持较短的重建时间。有了可用的替换 HDD 后，请更换故障 HDD 并检查系统，以确认自动重建是否正常开始。如果没有，请在相应的菜单中手动启动重建。您可以继续处理数据，但最好保持 RAID 系统空闲。这样可以缩短重建时间。切记：数据在重建过程中相当脆弱，因为此时没有冗余机制。重建时若再发生硬盘故障会导致数据丢失，因此强烈建议缩短重建时间。

机会只有一次。切勿将故障硬盘重新插回 RAID 系统。一旦它发生故障，故障会再次发生，即使在重新插回时它可能看起来正常工作。更重要的是，许多 RAID 系统可能通过识别硬盘上存在已知元数据而判断该硬盘曾经是 RAID 配置的一部分，这样会导致 RAID 系统混乱。系统可能会尝试将其重新组回 RAID 组，或者假设它是另一个 RAID 的一部分，这样会真正导致严重的后果。

认真配置。确认您的需求。比如，是否用 4 块硬盘组成 RAID5（单奇偶校验）或 RAID10（条带和镜像）来使用，或者是否只购置两块硬盘以双倍容量在 RAID1（简单镜像）下运行，这个主要取决于您对系统容量和速度的需求。

更多评测正在进行中。东芝正在编写一份包含更多数据、背景和约束条件的白皮书，敬请期待。同时，通过典型的 RAID 配置，我们已经证实，在 USB3 接口或通过千兆以太网连接到主机的环境下，一个 4 盘位 HDD 配置的 RAID5 在网速上限范围内能提供最大的可用数据。

参考资料

1. Enterprise HDD Legacy Systems: https://toshiba.semicon-storage.com/content/dam/toshiba-ss-v3/emea/en_gb/storage/technologycentre/2018-05-toshiba-whitepaper-enterprise-hdd-legacy-systems-en-screen.pdf
2. How New HDDs Interact with Old RAID Controllers: https://toshiba.semicon-storage.com/content/dam/toshiba-ss-v3/master/en/semiconductor/design-development/innovationcentre/whitepapers/How_new_hard_disk_drives_interact_with_old_RAID_controllers.pdf

TOSHIBA

Toshiba Electronic Devices & Storage Corporation
东芝电子元件及存储装置株式会社
https://toshiba.semicon-storage.com/cn/storage.html

Copyright © 2024 Toshiba Electronic Components Taiwan Corporation. All rights reserved. Product specifications, configurations, prices and component / options availability are all subject to change without notice. Product design, specifications and colours are subject to change without notice and may vary from those shown. Errors and omissions excepted.