

应用说明

**Broadcom
9660RAID5在
CSD-3310 3.84TB
NVMe SSDs
上的性能展现**



目录

1 概览.....3

2 测试设置.....3

3 正常模式下运行.....4

4 重建性能.....6

5 写放大.....8

6 结论.....9

1 概览

Broadcom 9600系列的RAID控制器专为高性能NVMe固态硬盘（SSD）而设计，特别强调RAID 5的性能，以提供更好的容量利用率和更高的存储效益。9600系列目前有两种基本配置可供选择：

1. **9660-16i**: 该设备采用PCIe Gen4 x 8主机接口，为存储设备提供了16个PCIe Gen4通道的支持。在RAID 5中，其最大随机4kB写入IOPS为55万，随机4kB读取IOPS为350万。此外，该设备的DDR内存总线宽度为64b+8b。
2. **9670W-16i/24i**: 该设备具备一个PCIe Gen4 x 16主机接口，为存储设备提供了16个或24个PCIe Gen4通道的支持。在RAID 5中，其最大随机4kB写入IOPS为110万，随机4kB读取IOPS为700万。此外，该设备的DDR内存总线宽度为128b+16b。

在RAID 5模式下，为了完成单个主机的写入操作，RAID控制器需要进行两次读取-修改-写入操作，一次用于数据，另一次用于奇偶校验。因此，要充分发挥9600系列RAID 5的性能，需要考虑每个固态硬盘的读写性能，从而确定所需的SSD数量。

透明压缩技术非常适用于同时有读写任务的混合工作负载环境，因为它可以减少固态硬盘由于写入对读取造成影响带来的性能瓶颈。在这份应用笔记中，我们展示了透明压缩技术使得9660-16i能够在，只使用四个CSD-3310 NVMe SSDs（占用所有16个PCIe通道的最低配置）的情况下，以满额的性能运行。此外，研究还表明，透明压缩技术将重建时间减少了近一半。总结而言，CSD-3310的透明压缩功能可以用更少的固态硬盘构建最大性能的RAID 5阵列，实现更短的重建时间，并且在降低RAID解决方案成本和提高可靠性方面不会牺牲性能。

2 测试设置

测试系统采用了Dell R650服务器，搭载了两个Intel Xeon Gold 6342处理器，主频为2.80GHz。系统内存容量为512GB的DDR内存。操作系统为Ubuntu 22.04.2 LTS，内核版本为5.15.0-73-generic。RAID卡型号为9660-16i，固件版本为8.5.1.0-00000-00001，而驱动程序版本为8.5.1.0.0。四个3.84TB的CSD-3310固态硬盘通过U.2分线缆进行连接。这些固态硬盘的固件版本为2.3。

在所有的测试案例中，RAID 5阵列都以相同的方式构建，并采用以下选项：

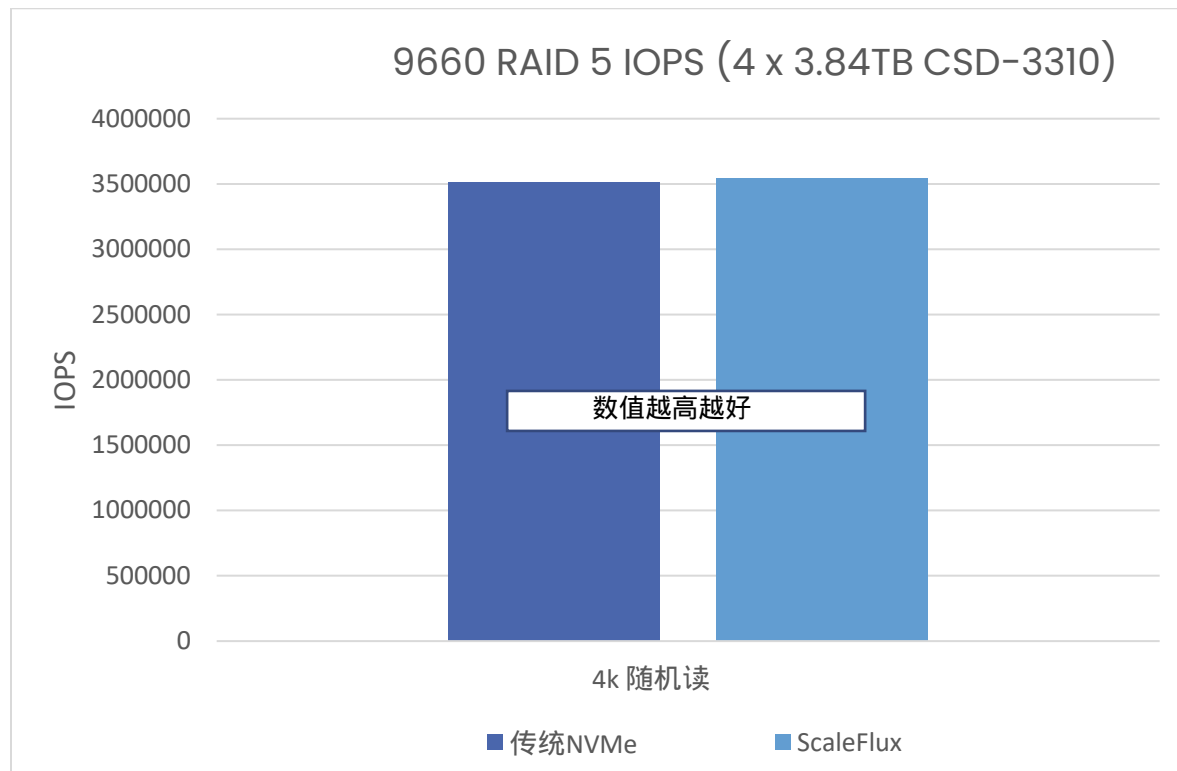
- 设置自适应写缓存（AWB）
- 选择64k的条带大小（只适用于固态硬盘）
- 完全初始化，在创建或重建RAID 5阵列时对所有存储设备进行完整的初始化和校验。

遵循Broadcom的建议，我们对生成的块设备设置进行了优化。所有的基准测试都使用了FIO版本3.35-2-g954b8。在所有的测试案例中，我们使用了io_uring引擎，以直接IO的方式对原始块设备进行操作。通过将"buffer_compress_percentage"选项设置为54，我们的透明压缩技术可以将数据压缩到原始大小的一半，即实现2:1的压缩比。

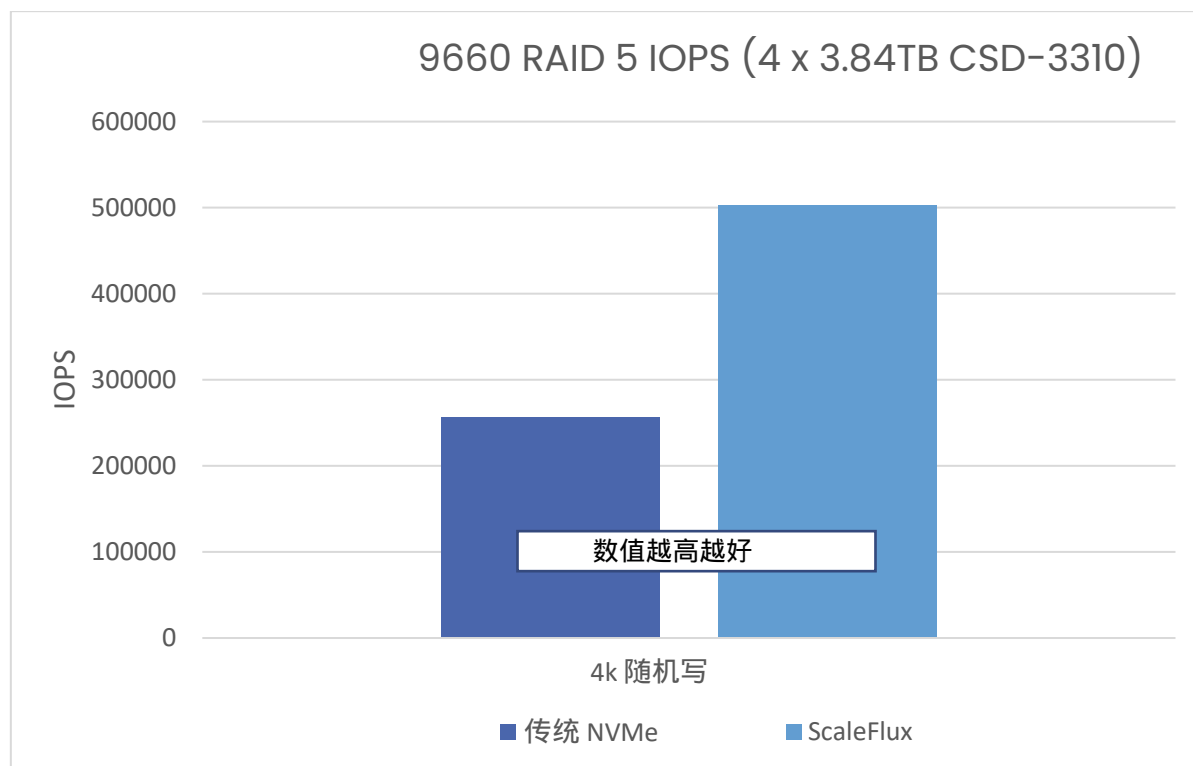
3 正常模式下运行

在正常模式下运行时，阵列不会出现硬盘出现故障或重建的情况。阵列经过完全的预处理，进行了4KB随机写入，并进行了三个测试案例：100%的随机4KB读取，100%的随机4KB写入以及混合的70%随机4KB读取和30%随机4KB写入。我们将阵列的性能与使用ScaleFlux CSD-3310和传统NVMe固态硬盘但不使用透明压缩的情况进行了比较。

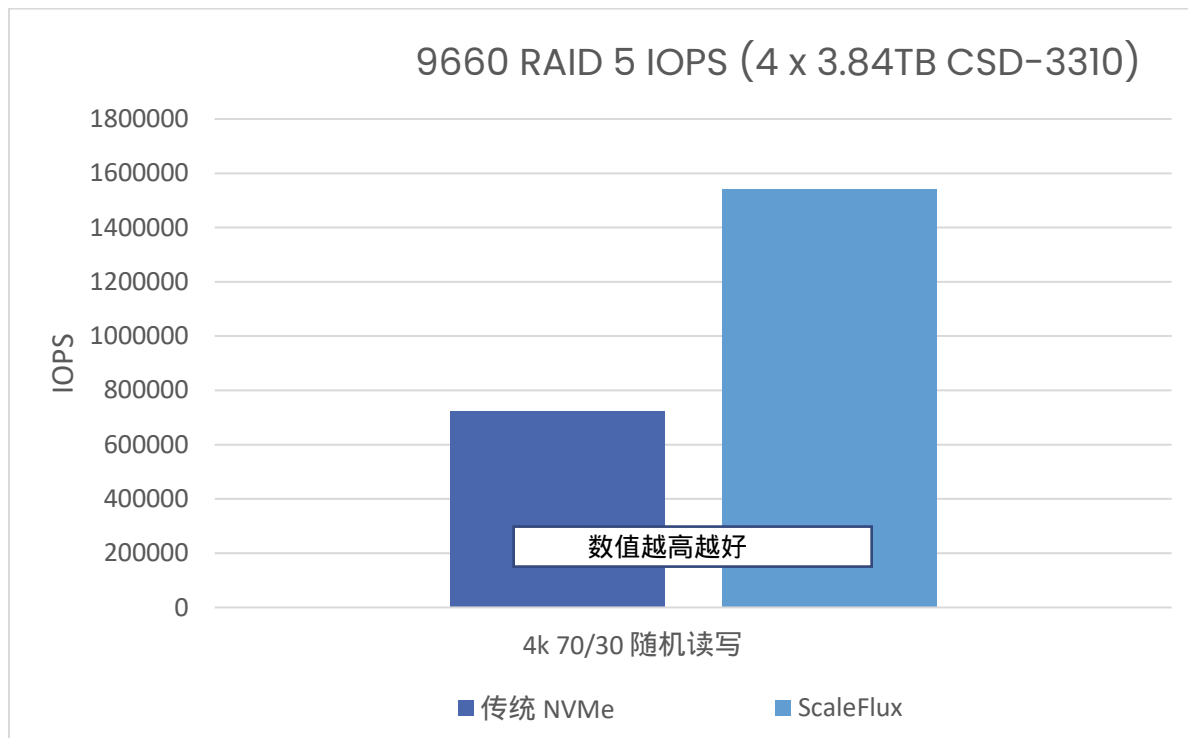
如预期的那样，纯随机读取性能相同，并达到了9660-16i的最大性能：



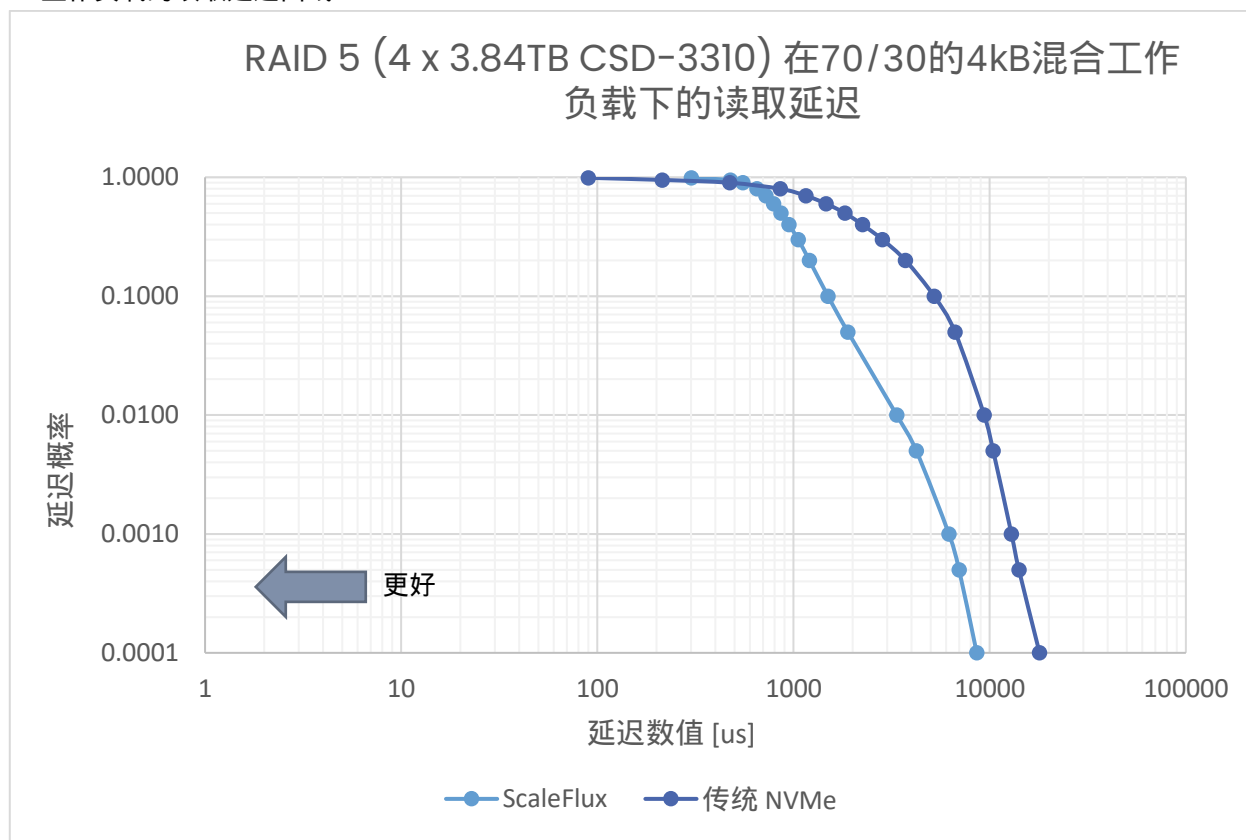
在随机写入的情况下，只有CSD-3310接近9660-16i的最大性能。而使用传统的NVMe设备，性能减少了一半：



同样地，使用CSD-3310的混合70/30性能提升了一倍。



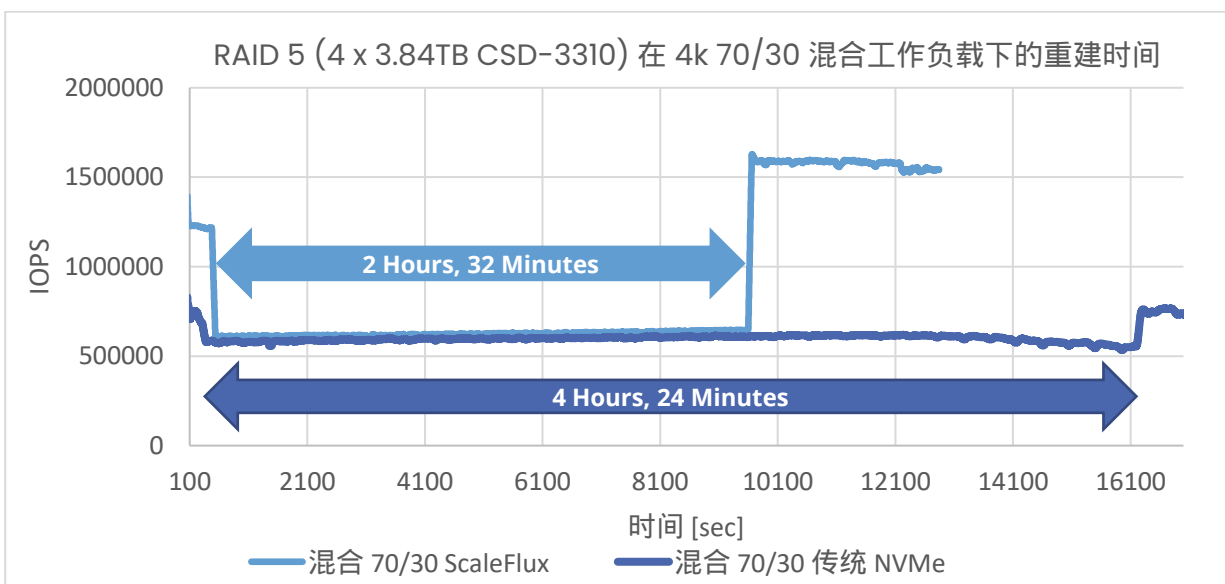
与传统NVMe相比，CSD-3310在提供两倍的读取操作的同时，显著改善了长尾延迟。下图显示了混合70/30工作负载的读取延迟曲线：



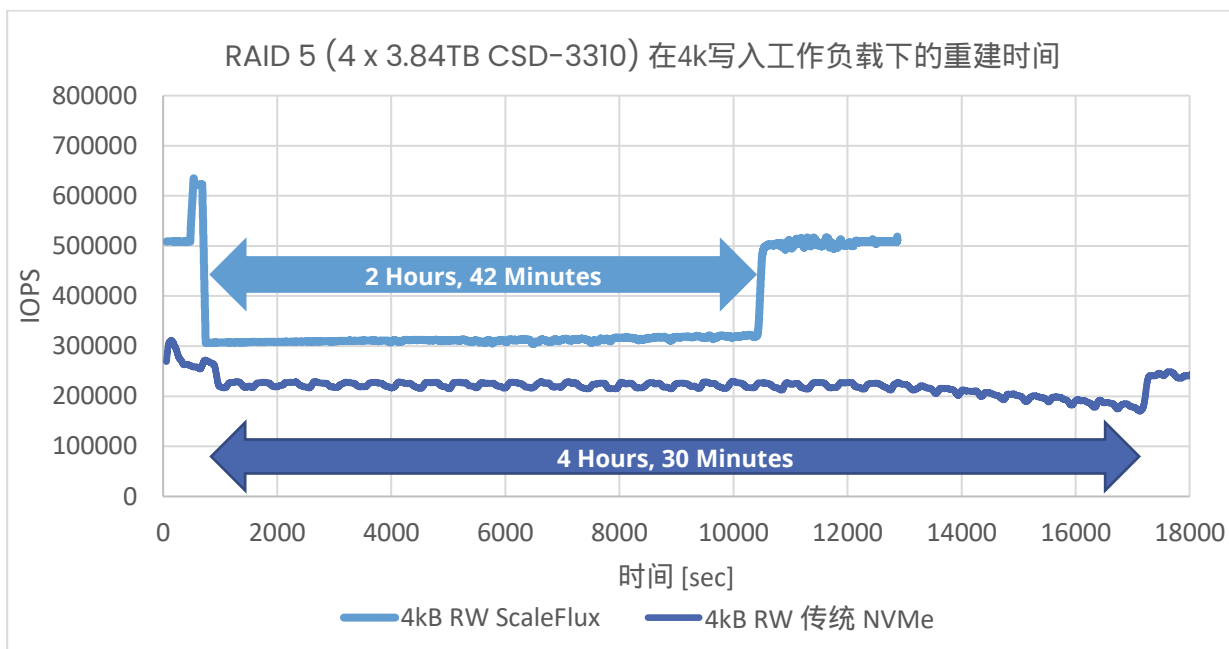
4 重建性能

重建过程中，性能会下降，因为主机IO必须与重建过程中的阵内RAID IO竞争。重建操作完成得越快，主机暴露在性能下降状态的时间就越短。在此测试中，我们使用30%的重建优先级，对100%随机4kB写入和混合（70%随机4kB读取/30%随机4kB写入）的主机工作负载下进行性能和重建时间的测量。使用传统的NVMe，在混合（70%随机4kB读取/30%随机4kB写入）主机工作负载下，重建时间为4小时24分钟。

在重建期间，读取性能约为420,000 IOPS，而写入性能约为200,000 IOPS。使用CSD-3310，重建时间缩短为2小时32分钟：

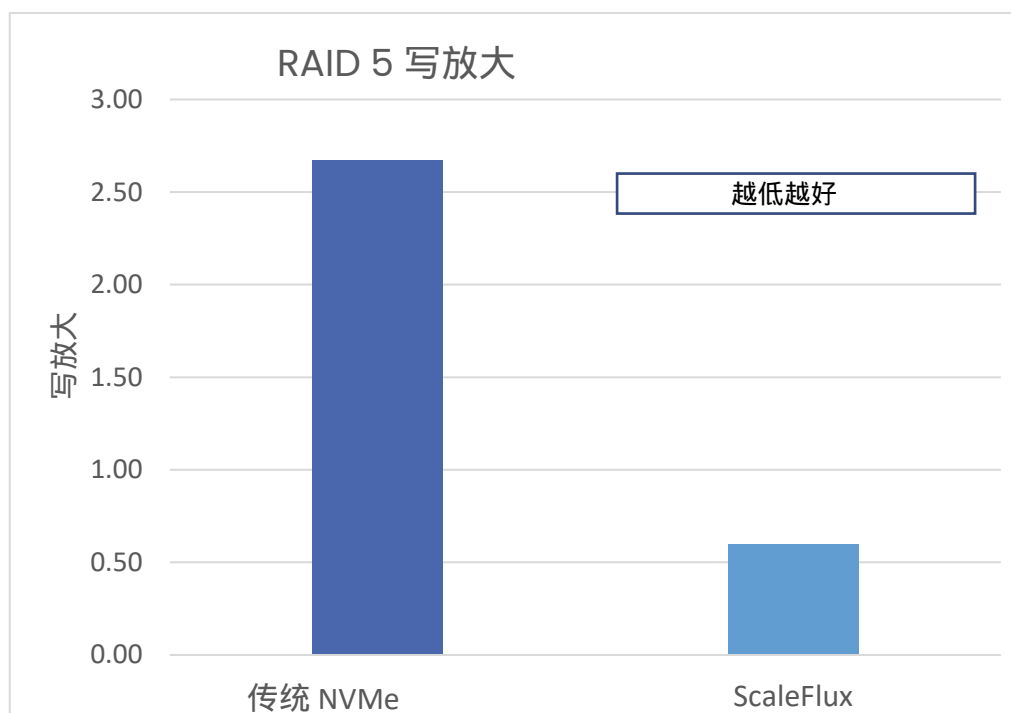


在纯随机写入工作负载下进行重建时，使用传统的NVMe固态硬盘的重建时间为4小时30分钟，持续IOPS水平约为225k。使用CSD-3310时，重建时间缩短为2小时42分钟，并且持续IOPS更高，约为300k。



5 写放大

由于RAID 5阵列的工作方式，减少写放大对于延长硬盘的使用寿命非常重要。CSD-3310硬盘的持续写放大因子仅为0.596，而没有透明压缩的传统硬盘的理论最低值为1。在这个工作负载下，传统硬盘的写放大因子为2.671。



6 结论

透明压缩是一款强大的工具，可以最大限度地提升基于Broadcom 9660-16i的RAID 5阵列的性能和可靠性。以下表格总结了本应用说明中描述的测试的关键结果（所有数值均基于4个3.84TB的CSD-3310 NVMe固态硬盘）：

指标	传统NVMe	ScaleFlux
稳态下的4KB随机写IOPS	257K	503K (+195%)
稳态下的4KB随机读IOPS	3.51M	3.54M
稳态下的混合工作负载（70%读取 / 30%写入）下的4KB随机IOPS	724K	1.54M (+212%)
50%稳态下4KB随机读取延迟 (us)	536	528
99.9%稳态下4KB随机读取延迟 (us)	1826	1417 (-22%)
4KB随机写入工作负载下的重建时间	4h:30m	2h:42m (-40%)
混合70%读取 / 30%写入的4KB随机工作负载下的重建时间	4h:24m	2h:32m (-43%)
写放大因子	2.671	0.596 (-4.4X)

使用CSD-3310 NVMe固态硬盘和透明压缩构建的RAID 5阵列将获得以下好处：

- 使用更少的硬盘实现最佳的RAID性能：
 - 成本更低
 - 可靠性更高（阵列平均故障间隔时间更长）
 - 功耗更低
- 重建时间更短
 - 更高的可靠性（降低了额外的硬盘故障风险）
 - 替换故障盘的时间缩短了
- 更长的寿命
 - 稳态下的写放大系数低于1
- 持久且强劲的性能表现
 - 减少写入对读取操作的干扰，以获得更好的混合IO延迟表现。
 - 最小的内部垃圾回收，从而对RAID本身的IO造成最小的竞争。

关于 Broadcom

Broadcom Inc. 是一家总部位于加州圣何塞的技术公司，在全球范围内拥有领先地位。

专门设计、开发和供应各类半导体和基础设施

软件解决方案。Broadcom的产品广泛应用于数据中心、网络、软件、宽带、无线、存储和工业等重要市场。其解决方案包括数据中心网络和存储、企业和大型机软件，注重自动化、监控和安全，同时涵盖智能手机组件、电信和工厂自动化。



www.broadcom.com

“

全球科技
领导者

”



关于 ScaleFlux

ScaleFlux是大规模部署计算存储的领导者，旨在帮助其客户利用数据增长作为竞争优势，提供企业级计算存储芯片解决方案，其硬件计算加速引擎极大优化了NVMe SSD，提升了存储的能力。有效加速应用程序并优化数据中心、企业和边缘网络的基础设施资源。让客户在处理数据库、分析、物联网和5G等工作负载时获得更大的竞争优势。

“

更好的SSD，
更快捷的服务

”



ScaleFlux



sales@scaleflux.com



www.scaleflux.cn