Harnessing the Performance of PCIe Gen 5

A Deep Dive into Unprecedented RAID Performance with SupremeRAID™, KIOXIA® NVMe Solid-state Drives, and Supermicro® Servers

Abstract

This whitepaper delineates a profound exploration into the capabilities of SupremeRAID™ in a RAID6 configuration, in stark contrast to RAID6/10 on Linux MD RAID. In synergy with Supermicro servers and KIOXIA CM7 solid-state drives (SSDs), our examinations have unveiled groundbreaking results, demonstrating the proficiency of SupremeRAID™ in maximizing KIOXIA CM7's performance while maintaining judicious CPU resource utilization.

Introduction

In the contemporary era marked by an escalating demand for highly efficient and advanced storage solutions, RAID configurations have emerged as critical components. The amalgamation of SupremeRAID™ with RAID6, coupled with Supermicro AS -2125HS-TNR and 24 KIOXIA CM7 NVMe SSDs, embodies an innovative system, proffering an unparalleled 676 TB of usable capacity under RAID6 protection. This confluence achieves read and write throughputs of 260 GB/s and 134 GB/s, respectively, catering to high-demand applications spanning HPC, AI, and databases.

RAID6: Advancing into Superior Storage Solutions

RAID10 necessitates substantial forfeitures in usable capacity to amalgamate all 24 CM7 into a singular volume, while RAID5 presents with merely a single redundant drive. By adopting RAID6, the SupremeRAID™ system harmoniously amalgamates performance and reliability, presenting itself as a versatile, industry-leading solution addressing a myriad of enterprise requisites.

Challenge and Solution

Traditionally, the deployment of software RAID solutions, especially with RAID6, is synonymous with substantial CPU resource consumption, owing to the imperative for parity calculation and meticulous I/O handling by read-modify-write operations. SupremeRAID™ adroitly circumvents these constraints, guaranteeing optimal performance without an undue strain on CPU resources.

Supermicro Servers

Hyper A+ Servers

Supermicro Hyper A+ Servers are dual-processor servers with AMD EPYC™ 9004 Series processors and up to 6TB of memory using 3DS ECC RDIMM DDR5-4800MHz in 24 DIMMs. They offer configurable PCIe slot options for up to 8 PCIe 5.0 x8 or 4 PCIe 5.0 x16 FH add-in cards, including SupremeRAID™. Key applications include Software-defined Storage, Virtualization, Enterprise Server, Cloud Computing, AI Inference, and Machine Learning.

Hyper A+ Server AS -2125HS-TNR
Certified for SupremeRAID™ and supports up to 24 2.5-inch NVMe/SAS/SATA drives.

Hyper A+ Server AS -2025HS-TNR
Certified for SupremeRAID™ and supports up to 12 3.5-inch NVMe/SAS/SATA drives.

Hyper SuperServers

Supermicro Hyper SuperServers are dual-processor servers with 4th Gen Intel® Xeon® Scalable processors (LGA-4677) and up to 8TB of memory using 3DS ECC RDIMM DDR5-4800MHz in 32 DIMMs. They offer configurable PCIe slot options for up to 4 PCIe 5.0 x16 and up to 4 PCIe 5.0 x8 add-in cards, including SupremeRAID™. Key applications include Software-defined Storage, Virtualization, Enterprise Server, Cloud Computing, AI Inference, and Machine Learning.

Hyper SuperServer SYS-221H-TN24R
Certified for SupremeRAID™ and supports up to 24 2.5-inch NVMe/SAS/SATA drives.

Hyper SuperServer SYS-621H-TN12R
Certified for SupremeRAID™ and supports up to 12 3.5-inch NVMe/SAS/SATA drives.

Storage A+ Server

The Supermicro Storage A+ Server ASG-2115S-NE332R is a single-processor EDSFF storage server with AMD EPYC™ 9004 Series processor and up to 6TB of memory using 3DS ECC RDIMM DDR5-4800MHz in 24 DIMMs. It offers 2 PCIe 5.0 x8 slots for add-in cards, including SupremeRAID™, and 2 5.0 x16 AIOM connectors (OCP 3.0 SFF compliant). Key applications include Software-defined Storage, In-Memory Computing, Data Intensive HPC, Private and Hybrid Cloud, and NVMe Over Fabrics Solutions.

Storage A+ Server ASG-2115S-NE332R
Certified for SupremeRAID™ and supports up to 32 E3.S (7.5mm) NVMe drives.

Storage A+ Server

The Supermicro Storage A+ Server SSG-121E-NE24RS is a dual-processor EDSFF storage server with 4th Gen Intel® Xeon® Scalable (LGA-4677) and up to 8TB of memory using 3DS ECC RDIMM DDR5-4800MHz in 32 DIMMs. They offer 2 PCIe 5.0 x16 slots for add-in cards, including SupremeRAID™, and 2 5.0 x16 AIOM connectors (OCP 3.0 SFF compliant). Key applications include Software-defined Storage, In-Memory Computing, Data Intensive HPC, Private and Hybrid Cloud, and NVMe Over Fabrics Solutions.

Storage SuperServer SSG-121E-NES24RS
Certified for SupremeRAID™ and supports up to 24 E1.S (9.5mm or 15mm) NVMe drives.

CloudDC SuperServers

The Supermicro CloudDC SuperServers are dual-processor servers with Intel® Xeon® Scalable processors supporting SupremeRAID™. The SYS-620C-TN12R server supports 3rd Gen Intel® Xeon® Scalable processors (LGA-4189), up to 6TB of memory using 3DS ECC RDIMM DDR4-3200MHz in 16 DIMMs and offer configurable PCIe slot options for up to 4 PCIe 4.0 x16 and 2 PCIe 4.0 x8 add-in cards. The SYS-121C-TN10 server supports 4th Gen Intel® Xeon® Scalable processors (LGA-4677), up to 4TB of memory using 3DS ECC RDIMM DDR5-4800MHz in 16 DIMMs and offer configurable PCIe slot options for up to 2 PCIe 5.0 x16 add-in cards. Key applications include Web Server, Firewall Applications, Data Center Optimized, Value infrastructure as a service (IaaS), Cloud Computing, Compact Servers, Domain Name Servers (DNS) and Gateway Servers, Content Delivery Network (CDN) Servers, and Edge Nodes.

CloudDC SuperServer SYS-620C-TN12R
Certified for SupremeRAID™ and supports up to 12 3.5-inch NVMe/SAS/SATA drives.

CloudDC SuperServer SYS-121C-TN10R
Certified for SupremeRAID™ and supports up to 10 2.5-inch NVMe/SAS/SATA drives.

KIOXIA CM7-R Series NVMe SSDs

The KIOXIA CM7-R Series NVMe SSDs, tailored for read-intensive applications, is engineered to support a diverse range of enterprise applications and workloads. It features 112-layer BiCS FLASH™ 3D TLC flash memory, delivering up to 2,700K IOPS (random read) and 310K IOPS (random write) and offering up to 30.72 TB of storage capacities, thereby establishing itself as a robust backbone for read-intensive enterprise applications.

KIOXIA CM7-R Series (2.5-inch)
Certified for SupremeRAID™ and supports up to 30.7GB of capacity per drive.

KIOXIA CM7-R Series (E3.S)
Certified for SupremeRAID™ and supports up to 15.3GB of capacity per drive.

SupremeRAID™

Heralded as the world's fastest solution for PCIe Gen 3, 4, and 5, SupremeRAID™ stands as a paragon of unprecedented NVMe/NVMeoF performance. It augments scalability, server lifespan, and cost-effectiveness while realizing up to 26M IOPS and 220GB/s with a single card.

SupremeRAID™ SR-1010
Delivers up to 28M IOPS and 260GB/s
and supports up to 32 NVMe drives.

Testing Environment

Hardware

- Server: Supermicro AS -2125HS-TNR x 1
- CPU: AMD EPYC 9654 96-Core Processor x 2
- Memory: Samsung M321R2GA3BB6-CQKVS DDR5 4800 MT/s 16GB x 24
- NVMe Drive: KIOXIA CM7-R 3.84T KCMY1RUG3T84 x 24
- RAID Controller: SupremeRAID™ SR-1010 x 1

Software:

- Linux OS: Ubuntu 22.04.1 LTS
- Linux kernel: 5.15.0-83-generic
- Linux mdadm (Multiple Disk and Device Management): 4.2 - 2021-12-30
- Linux fio (Flexible IO Tester): 3.28
- SupremeRAID™ Driver: 1.5.0-659.g10e76f72.010

Performance Benchmark Analysis

Examination of Benchmark Specifics

This evaluation focuses on the four predominant metrics: 4K random read, 4K random write, 1M sequential read, and 1M sequential write. The assessment employs fio with 192 jobs to simulate authentic workloads on a system harboring a total of 192 physical cores. Regarding the RAID configuration, both SupremeRAID™ RAID6 and MD RAID6 are formulated with 4K striping, while MD RAID10 is constituted with a default 512K strip, with all 24 SSDs uniformly allocated across both sockets. The rationale behind establishing RAID6 with a 4K strip is the minimization of write amplification, triggered by read-modify-write operations, especially considering RAID6 incorporates two parity chunks.

Articulation of Benchmark Outcomes

The analysis of the benchmark outcomes paints a comprehensive picture of the varied capabilities and efficiencies of different RAID configurations under distinct operational conditions.

The results from the fio testing demonstrate that SupremeRAID™ RAID6's random read IOPS significantly eclipses its counterparts, exhibiting agility that is at least quintuple that of MD RAID6 and RAID10 while maintaining a CPU utilization that is less than half. This is especially pertinent under heightened stress workloads generated by multiple processes, where instances of locking and CPU contention with the application are elevated.

Moreover, as increments in I/O are positively correlated with rises in interrupts—given the dependency of software RAIDs predominantly on native NVMe drivers and interrupts—these results underscore the challenges that software RAID faces in striving to surpass 10M IOPS in kernel space.

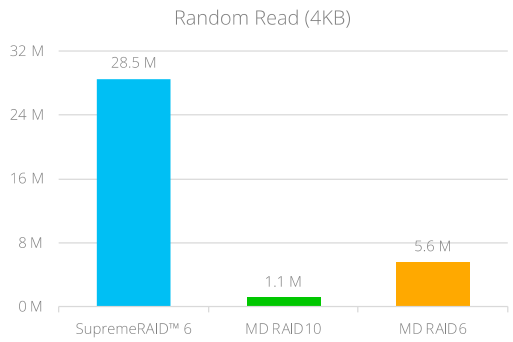
Assessing random write reveals that MD RAID10 excels significantly over RAID6, primarily due to the absence of parity calculations. However, SupremeRAID™ RAID6 not only achieves a twofold enhancement in efficiency but also maintains minimal CPU utilization. This is attributed to the innovative approach of SupremeRAID™ in offloading the parity calculation to the GPU.

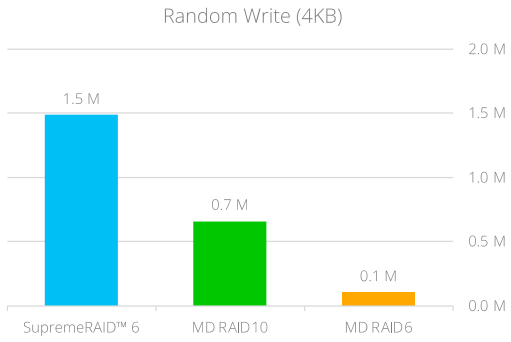
In the realm of 1M sequential read, while MD RAID10 displays commendable efficiency due to the 512k strip size, SupremeRAID™ RAID6 manages to secure a throughput of 279GB/s, even under a 4k strip, and with lower CPU utilization in comparison to the 21.1GB/s by MD RAID6. This discrepancy in performance, with the smaller strip size causing large read I/O to fragment into multiple smaller I/Os, underscores the limitations imposed by the I/O handling capability on throughput.

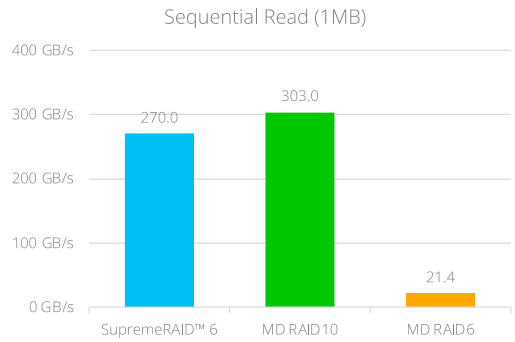
A close examination of 1M sequential write unveils a notable degradation in throughput for MD RAID6, shedding light on the pervasive reluctance to incorporate RAID6 with NVMe drives and the consequent propensity to forgo usable capacity in favor of RAID10. SupremeRAID™ reshapes this paradigm by delivering a staggering 134 GB/s, outclassing RAID10 by over 30% and minimizing redundant write I/O. This pivotal advancement is a testament to the transformative potential in SupremeRAID™ storage solution technologies.

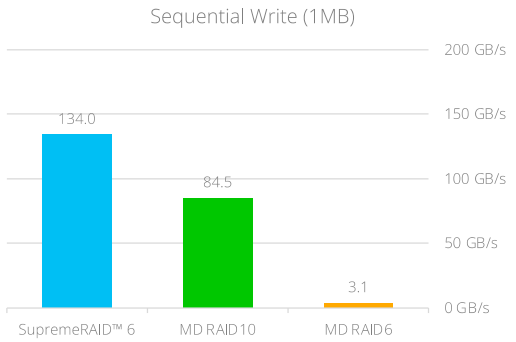
Workloads	SupremeRAID™ RAID6	Linux MD RAID6 4K Strip Size	Linux MD RAID6 512K Strip Size
Random Read 4K	28523K IOPS	5526K IOPS	1119K IOPS
CPU Utilization	22.09%	50.88%	50.07%
Random Write 4KB	1487K IOPS	100K IOPS	658K IOPS
CPU Utilization	3.18%	49.94%	50.42%
Sequential Read 1M	260GiB/s (279GB/s)	19.7GiB/s (21.1GB/s)	283GiB/s (303GB/s)
CPU Utilization	2.81%	51.75%	4.08%
Sequential Write 1M	125GiB/s (134GB/s)	3.21GiB/s (3.37GB/s)	78.7GiB/s (84.5GB/s)
CPU Utilization	10.46%	30.04%	2.52%

NOTE: Higher IOPS and lower CPU utilization are better.









NOTE: Higher IOPS are better.

Conclusion

The synthesis of SupremeRAID™, KIOXIA CM7, and Supermicro signifies a pivotal advancement in storage solution technologies, as demonstrated by the system used for comparison performance testing. This integrated configuration, particularly with RAID6, promises unprecedented performance, reliability, and scalability, catering to applications necessitating extensive CPU and data processing, thus standing as a paradigm of technological evolution in storage solutions.

Appendix

fio Profile for Testing

```
[global]
filename=/dev/gdg0n1
ioengine=libaio
randrepeat=0
direct=1
random_generator=tausworthe64
cpus_allowed_policy=split
group_reporting=1
norandommap=1
numjobs=192
cpus_allowed=0-191
[randread-b4k-j192-d32]
runtime=300
time_based=1
rw=randread
bs=4k
iodepth=32
size=9600G
stonewall
[write-b1m-j192-d16]
rw=write
bs=1m
iodepth=16
size=50G
offset_increment=50G
stonewall
[read-b1m-j192-d16]
runtime=300
time_based=1
rw=read
bs=1m
iodepth=16
size=50G
offset_increment=50G
stonewall
```

```
[randrw73-b16k-j192-d16]
runtime=300
time_based=1
rw=randrw
rwmixread=70
bs=16k
iodepth=16
size=9600G
stonewall
[randwrite-b4k-j192-d16]
runtime=300
time_based=1
rw=randwrite
bs=4k
iodepth=16
size=9600G
stonewall
```

About Supermicro

Supermicro (SMCI) is a global technology leader committed to delivering first to market innovation for Enterprise, Cloud, AI, and 5G Telco/Edge IT Infrastructure. We are transforming into a Total IT Solutions provider with environmentally friendly and energy-saving server, AI, storage, IoT, and switch systems, software, and services while delivering advanced high-volume motherboard, power, and chassis products. Learn more at supermicro.com.

About Graid Technology

Graid Technology, creator of SupremeRAID™ next-generation GPU-based RAID, is led by a team of experts in the storage industry and is headquartered in Silicon Valley, California with an R&D center in Taipei, Taiwan. Designed for performance-demanding workloads, SupremeRAID™ is the world's fastest NVMe and NVMeoF RAID solution for PCIe Gen 3, 4, and 5. A single SupremeRAID™ card delivers up to 28M IOPS and 260GB/s and supports up to 32 native NVMe drives, delivering superior NVMe/NVMeoF performance while increasing scalability, improving flexibility, and lowering TCO. For more information on Graid Technology, visit graidtech.com or connect with us on Twitter or LinkedIn.