SOLUTION BRIEF

STREAMLINING EXTREME PERFORMANCE

SUPREMERAID™ HE, SUPERMICRO'S ALL-FLASH SBB SYSTEM, AND BEEGFS FOR UNMATCHED NVME STORAGE

TABLE OF CONTENTS

Executive Summary

In today's data-driven landscape, enterprises and high-performance computing (HPC) environments demand storage solutions that deliver exceptional speed, scalability, and resilience while optimizing cost-efficiency. Graid Technology introduces SupremeRAID™ HE, a cutting-edge GPU-accelerated NVMe RAID solution with array migration and cross-node high-availability (HA), paired with Supermicro's 2U All-Flash Storage Bridge Bay (SBB) SSG-221E-DN2R24R system and the BeeGFS parallel file system. Highlighting its performance, this architecture achieves the world's highest throughput in a 2U system:

- Delivering 132 GB/s reads and 83 GB/s writes post-RAID, (near theoretical maximums) on local storage as measured by StorageBench
- Saturating a 400 Gb/s network with 93 GB/s read and 84 GB/s write from the client side as measured by IOzone
- By eliminating cross-node replication, it reduces NVMe costs and offers scalable adaptability.

Unlike advanced software RAID approaches, SupremeRAID™ HE leverages GPU parallelism to maximize NVMe performance, eliminate CPU bottlenecks, and reduce total cost of ownership (TCO). This whitepaper explores how this innovative design meets evolving demands.

Solution Overview

SupremeRAID™ HE integrates with Supermicro's SSG-221E-DN2R24R and BeeGFS to form a high-performance, highly available NVMe storage platform. Unlike software RAID, which consumes significant CPU resources, SupremeRAID™ HE offloads RAID operations to a GPU, preserving CPU capacity for critical upper-layer applications like BeeGFS. This GPU-accelerated approach, combined with array migration for cross-node high availability (HA), delivers exceptional throughput—up to 28 million IOPS and 260 GB/s per card—while supporting up to 32 drives in a compact 2U footprint. Reducing CPU overhead streamlines system performance, enhances scalability, and lowers total cost of ownership (TCO) for data-intensive enterprise and HPC workloads, offering a cost-effective and efficient alternative to advanced software RAID solutions.

Supermicro Storage Bridge Bay (SBB) - SSG-221E-DN2R24R

The Supermicro SSG-221E-DN2R24R is a 2U all-flash SBB system featuring two hot-pluggable nodes, each powered by a 5th or 4th Gen Intel® Xeon® Scalable processor (up to 350W TDP) and supporting up to 2TB of DDR5-5600 ECC memory across 8 DIMMs. It accommodates 24 hot-swap U.2 NVMe drives with dual-port connectivity, shared between nodes, and offers PCIe 5.0 slots, delivering high-density, energy-efficient storage with redundant 1200W Titanium-level power supplies and a 1GbE node-to-node link.

(Angled View – System)

(Angled Rear View – System)

SupremeRAID™ HE (HPC Edition) by Graid Technology

SupremeRAID™ HE, developed by Graid Technology, is a GPU-accelerated NVMe RAID solution featuring array migration for cross-node high availability (HA). When combined with hardware that enables the distribution of a drive set among various nodes (such as an SBB system with dual-port drives or conventional nodes linked via NVMeoF), applications or storage solutions can depend entirely on SupremeRAID™ HE to ensure drive redundancy and facilitate array migration across nodes. This approach removes the need for data replication between nodes, thus guaranteeing high availability while significantly reducing the total cost of ownership (TCO).

BeeGFS Parallel File System

BeeGFS is a hardware-independent file system developed by ThinkParQ with a strong focus on performance, ease of use, simple installation, and efficient management. Based on an Available Source development model (with publicly accessible source code), ThinkParQ offers BeeGFS as both a self-supported Community Edition and a fully supported Enterprise Hive Edition with additional features and functionalities.

System Design and High Availability

This solution leverages the Supermicro SSG-221E-DN2R24R's dual-node architecture, enhanced by SupremeRAID™ HE's array migration capabilities, to deliver exceptional performance and robust high availability (HA) within a 2U chassis. Integrated into a 400G network, the system connects seamlessly with BeeGFS clients, optimizing resource utilization by eliminating cross-node data replication and reducing costs. The design supports linear scalability, allowing additional SBB units to expand capacity and performance as needed.

Hardware Configuration

The solution features two SBB nodes, each equipped with a PCIe switch for internal data routing and connected via a dual-port backplane that enables both nodes to access 24 NVMe SSDs, organized into two groups of 12. Each node includes a Network Interface Card (NIC) and a GPU, with SupremeRAID™ HE utilizing GPU acceleration to offload RAID operations, enhancing throughput and preserving CPU resources. The setup is tested over a 400G network infrastructure.

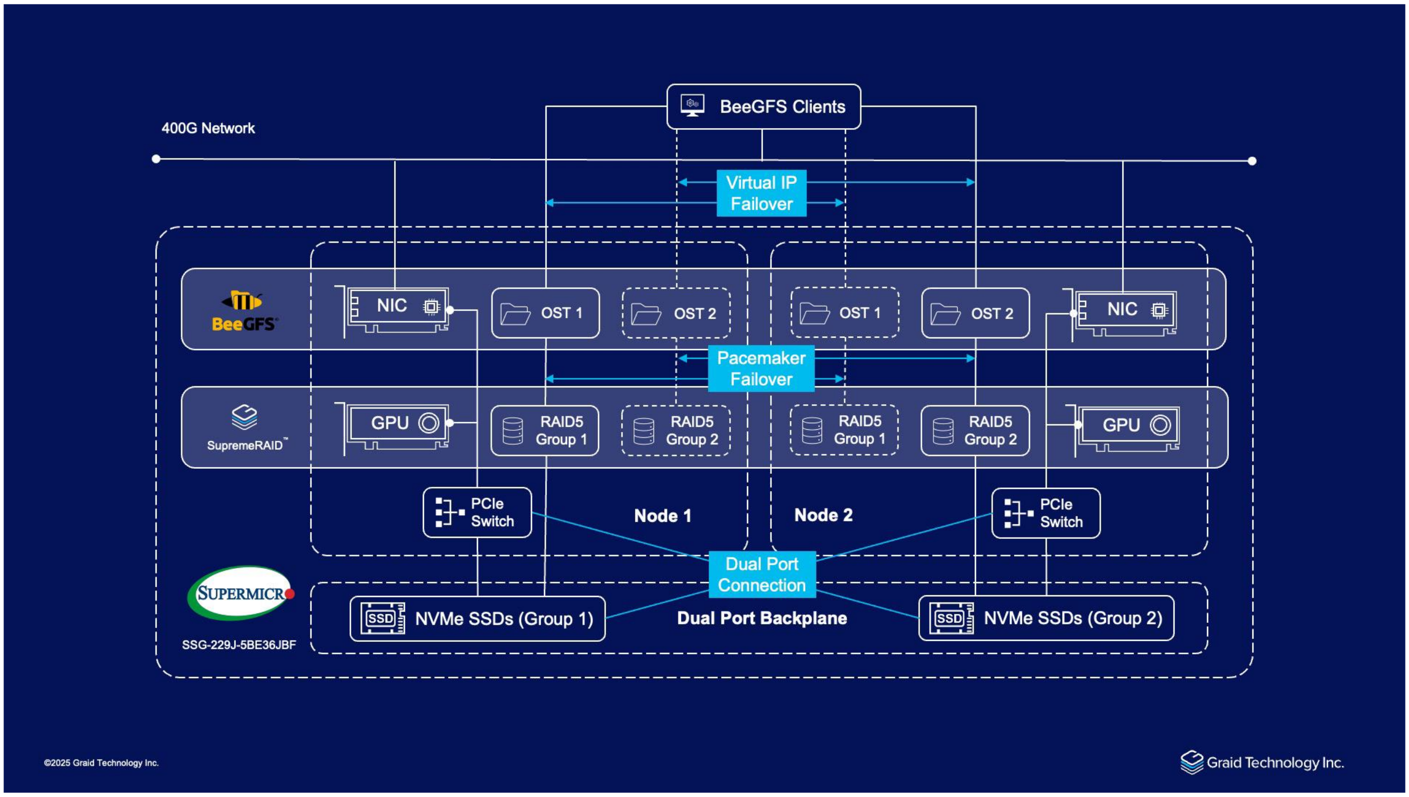
RAID and Storage Setup

SupremeRAID™ HE manages the 24 dual-port NVMe SSDs, with each node handling 12 drives. This configuration supports the system's high-performance capabilities.

- **Metadata Target (MDT)**:
  - Configured as RAID 10 with 2 drives per node (4 total), forming 4 virtual drives for low-latency metadata operations.
- **Object Storage Target (OST)**:
  - Configured as RAID 5 or RAID 6 with 10 drives per node (20 total across both nodes), creating multiple virtual drives for high-capacity, fault-tolerant data storage.

High Availability Implementation

High availability is ensured through Pacemaker, which manages failover using a virtual IP, maintaining service continuity during node failures. SupremeRAID™ HE's array migration enables seamless RAID array transfers between nodes, eliminating replication needs and optimizing NVMe usage. The dual-port backplane and 400G network connectivity enhance fault tolerance and data throughput.



Testing Environment

Hardware Configuration

Storage Server

Component	Specification	Per Node	Total
Server	SSG-221E-DN2R24R (2U2NSP)	1	1
CPU	Intel Xeon Gold 6442Y (24 cores)	1	2
Memory	SK Hynix DDR5 32GB 5600MHz	8	16
SSD	Samsung PM1743 3.84TB NVMe	24 (shared)	24
RAID Controller	SupremeRAID™ HE with SR-1010	1	2
Network	Mellanox ConnectX-7 NDR (400Gb/s)	1	2

Client Server

Client Server Component	Specification	Per Node	Total
Server	SYS-621BT-HNTR (2U4NDP)	1	1
CPU	Intel Xeon Gold 6442Y (24 cores)	2	4

Networking

The system uses an Nvidia QM9700 Quantum-2 InfiniBand switch with 400Gb/s aggregate bandwidth.

Hardware Tuning

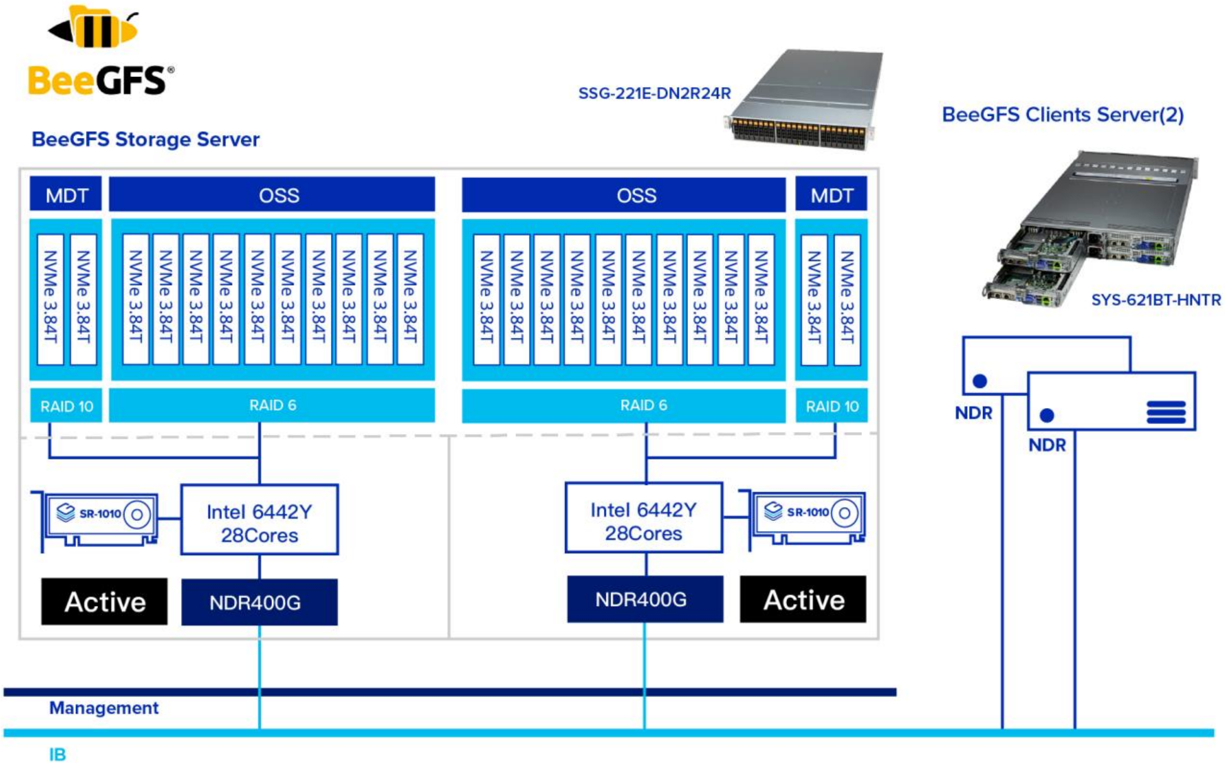
BIOS settings disable C-states and enable maximum performance mode, SupremeRAID™ HE is pinned to GPU cores, and network MTU is set to 9000, with interrupt coalescing disabled to achieve full 400Gb/s saturation.

Software Configuration

The testbed runs RHEL 9.1, with BeeGFS version 7.4.4, SupremeRAID™ driver 1.6.1, OFED5.8.5.1.1.2, and Pacemaker for array migration management.

Testing Environment Overview

The testing environment features a configuration with RAID setups on two storage nodes, where each node uses RAID 6 with 10 drives for Object Storage Service (OSS) and RAID 10 with 2 drives for Metadata Target (MDT). The diagram below illustrates the network and hardware layout used for testing:



Performance Benchmarking

Methodology

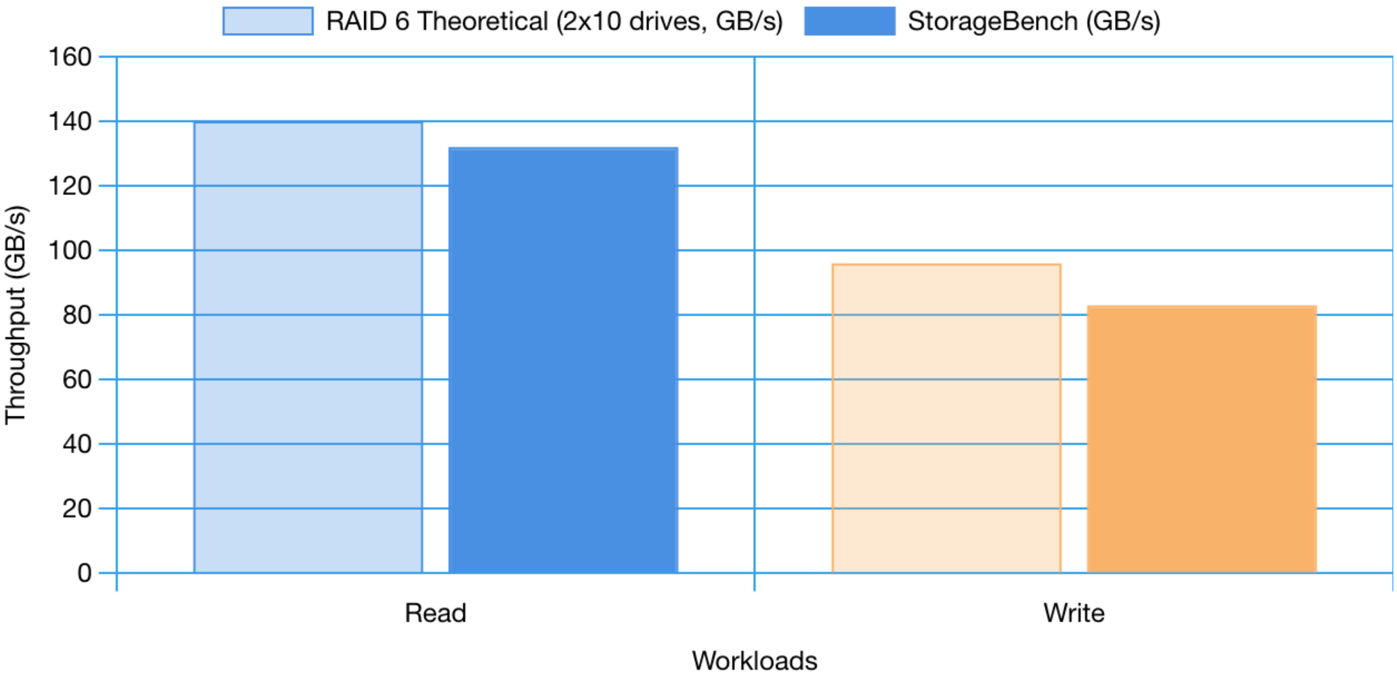
- **Local:** BeeGFS StorageBench tests measured raw storage throughput on the SBB nodes.
  - StorageBench:Targets: All targets.
    - **File Size:** 100GB per job.
    - **Direct I/O:** Enabled.
    - **Block Size:** 1 MB.
    - **Threads:** 64 (32 per node across two nodes).
  - **Purpose:** Assess BeeGFS-specific performance metrics, focusing on sequential read/write throughput.
  - **Execution:**
    - Start BeeGFS services: mgmtd, meta, storage, helperd, and client.
    - Execute StorageBench on a client node:
      - **Write:** [sbb-a]# beegfs-ctl --storagebench --alltargets --write --blocksize=1m --size=100g --threads=64 --odirect
      - **Read:** [sbb-a]# beegfs-ctl --storagebench --alltargets --read --blocksize=1m--size=100g --threads=64 --odirect
- **Client-Side:**
  - **IOzone:** Tests evaluated client-side file system performance over the network.
    - **Tool Version:**
      - IOZone: 3.506
      - **mpirun:** 5.0.5
    - **Configuration:**
      - **Jobs:** 2 to 256 total jobs (1 to 128 jobs per node across 2 SYS-621BT-HNTR nodes).
      - **File Size:** 100GB per job.
      - **Block Size:** 1 MB.
      - **Purpose:** Evaluate file system performance across the network, focusing on scalability and bandwidth utilization.
    - **Execution:**
      - Run IOzone with varying thread counts: [sys621bt-a]# /opt/iozone/bin/iozone -i<type> -MCcew -I -r 1m -s 100g -t $threads -+n -+u -+m <hostfile>
  - **IOR:** Tests assessed parallel I/O performance in multi-client scenarios.
    - **Tool Version:**
      - **IOR:** 4.0.0
      - **mpirun:** 5.0.5
    - **Configuration:**
      - **Jobs:** 2 to 256 total jobs (1 to 128 jobs per node across 2 SYS-621BT-HNTR nodes).
      - **File Size:** 100GB per job.
      - **Block Size:** 1 MB.
      - **Purpose:** Assess parallel I/O performance, simulating HPC workloads with distributed clients.
    - **Execution:**
      - Execute IOR with MPI: [sys621bt-a]# mpirun --allow-run-as-root -np $threads --hostfile <hostfile> --bynode /usr/local/bin/ior -a POSIX -C -F -E -k -e -g-b $bs -t 1m -o /mnt/beegfs/testfile -w -r --posix.odirect

Results

Local Storage Performance

Each Samsung PM1743 NVMe drive, connected via Gen5 x2, delivers sequential read performance of 7 GB/s and sequential write performance of 6 GB/s. For two RAID 6 groups with 10 drives each (8 effective data drives per group after two parity drives), the theoretical maximum throughput is 7 × 10 × 2 = 140 GB/s for read and 6 × (10 - 2) × 2 = 96 GB/s for write across both nodes. The StorageBench results with SupremeRAID™ HE managing 20 drives in RAID 6 show 132 GB/s read and 83 GB/s write.

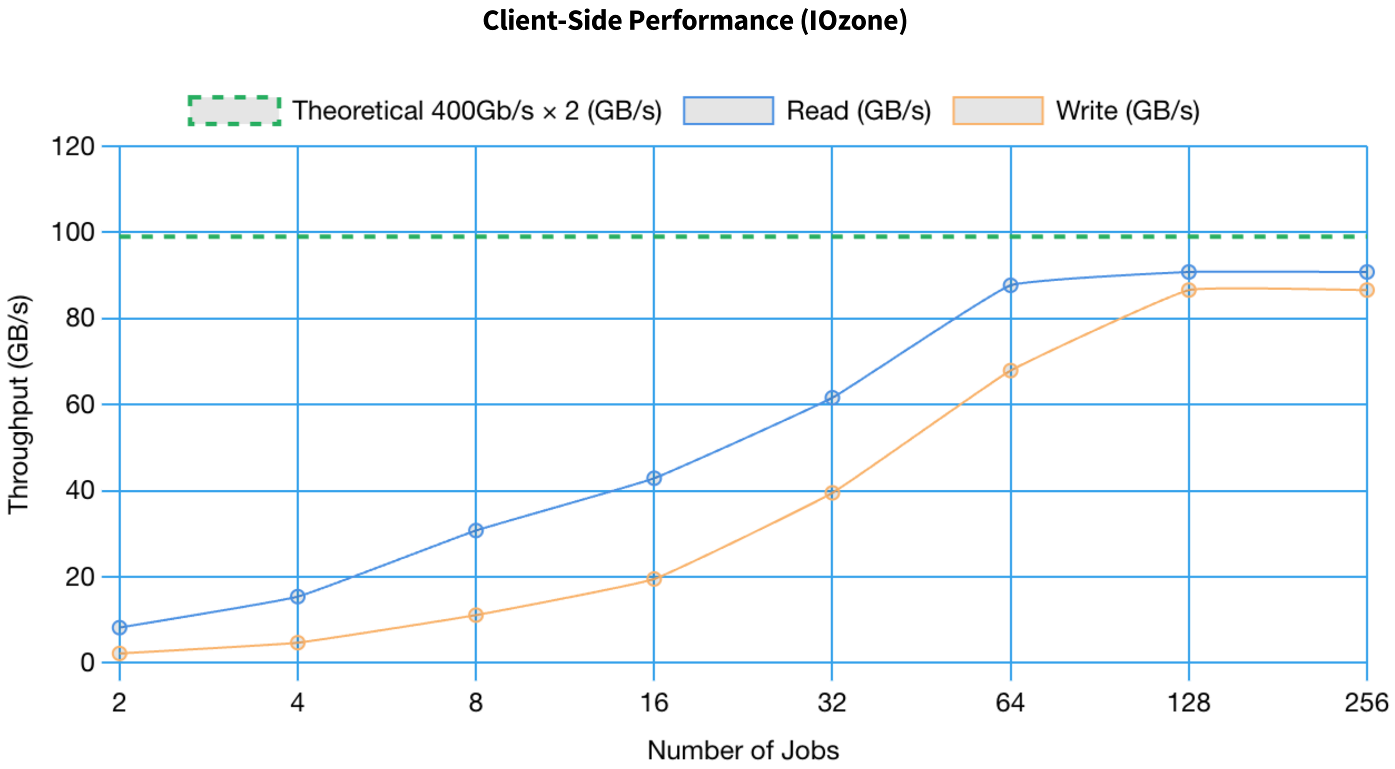
Local Storage Performance (StorageBench)



Metric	Read	Write
RAID 6 Theoretical (2x10 drives, GB/s)	140	96
StorageBench (GB/s)	132	83

The chart compares the theoretical maximum throughput of two RAID 6 groups (140 GB/s read,96 GB/s write) to the StorageBench results (132 GB/s read, 83 GB/s write), showing read performance at 94% and write at 86% of theoretical maximums, highlighting SupremeRAID™ HE's efficiency in delivering near-theoretical performance after RAID protection.

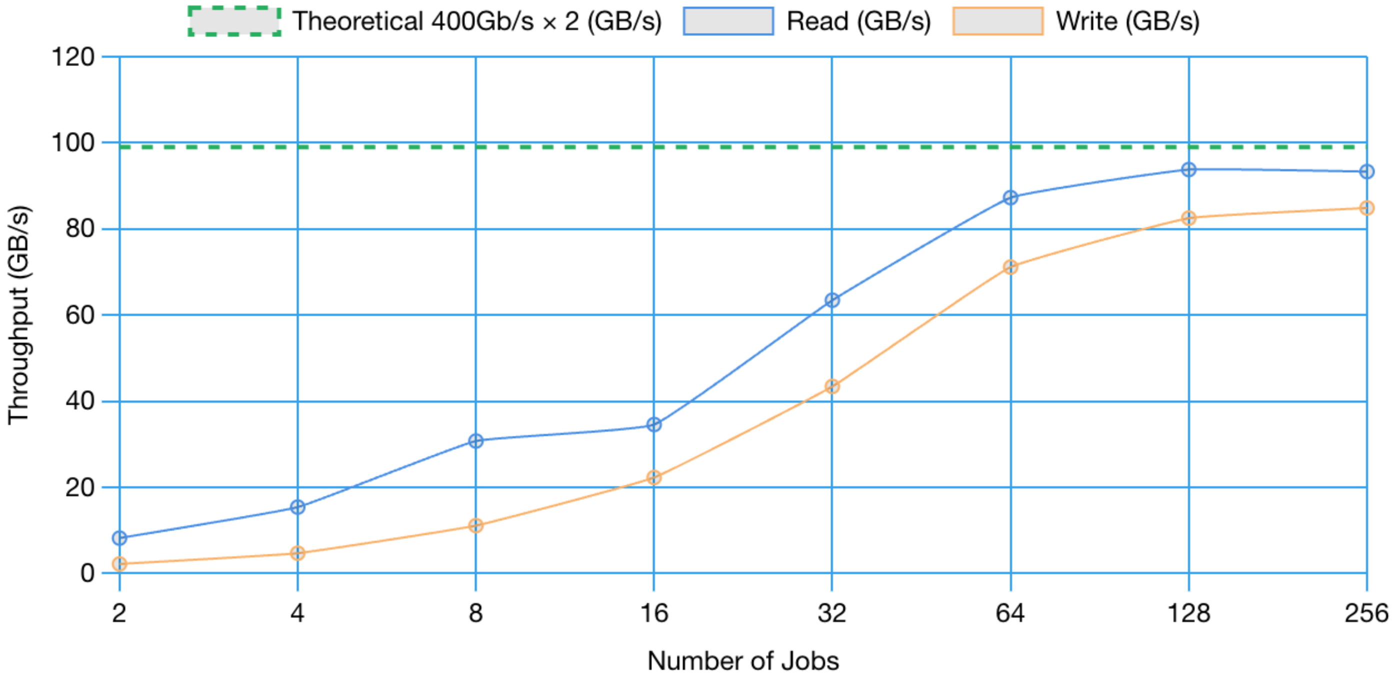
Client-Side Performance



Throughput /Number of Jobs	2	4	8	16	32	64	128	256
Read (GB/s)	8.2	15.4	30.75	42.9	61.6	87.7	90.8	90.8
Write (GB/s)	2.2	4.7	11.11	19.5	39.5	67.9	86.6	86.6

IOzone peaks at 90.8 GB/s read and 86.6 GB/s write with 128 jobs, approaching the theoretical 400Gb/s × 2 (800 GB/s) network bandwidth, demonstrating SupremeRAID™ HE’s ability to saturate the network nearly.

Client-Side Performance (IOR)



Throughput /Number of Jobs	2	4	8	16	32	64	128	256
Read (GB/s)	8.2	15.4	30.75	34.58	63.45	87.27	93.79	93.32
Write (GB/s)	2.2	4.7	11.11	22.29	43.43	71.14	82.51	84.87

IOR reaches 93.32 GB/s read and 84.87 GB/s write with 256 jobs, closely approaching the theoretical 400Gb/s × 2 (800 GB/s) network bandwidth, underscoring SupremeRAID™ HE's capability to saturate the network nearly.

Conclusion

SupremeRAID™ HE, integrated with Supermicro's SSG-221E-DN2R24R and BeeGFS, redefines NVMe storage standards. Leveraging array migration for cross-node high availability (HA), it delivers peak performance in a 2U system with two 24-core CPUs, saturating two 400Gb/s networks. By eliminating cross-node replication, it reduces NVMe costs and offers scalable adaptability. Achieving 132 GB/s read and 83 GB/s write locally—near theoretical limits post-RAID—and up to 93 GB/s read and 84 GB/s write from clients, this solution is excellent for your high performance storage needs, including HPC, analytics, and enterprise applications and is validated by rigorous benchmarks.

For More Information

Supermicro Storage Bridge Bay (SBB) - SSG-221E-DN2R24R:
https://www.supermicro.com/en/products/system/storage/2u/ssg-221e-dn2r24r

Supermicro and GRAID Solutions: https://www.supermicro.com/en/solutions/graidtechnology

SUPERMICRO

As a global leader in high performance, high efficiency server technology and innovation, we develop and provide end-to-end green computing solutions to the data center, cloud computing, enterprise IT, big data, HPC, and embedded markets. Our Building Block Solutions® approach allows us to provide a broad range of SKUs, and enables us to build and deliver application-optimized solutions based upon your requirements.

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 LinkedIn.