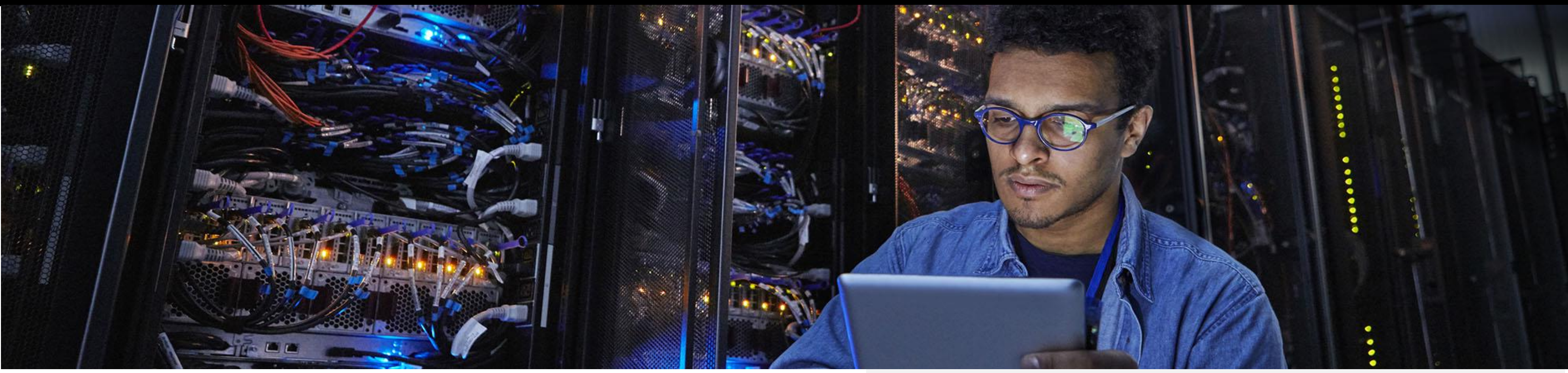

WEKA CLUSTERS WITH MICRON® HIGH CAPACITY, MAINSTREAM, OR PERFORMANCE[1] NVMe™ SSDs DELIVER MAJOR BENEFITS[2]

Cloud-native, scale-out applications help maximize business value. When faced with the broad range of system components and designs for these applications, IT organizations need to simplify and optimize.

This document is designed to help readers select which Micron® NVMe™ SSD and server configuration is the right choice for specific use cases. It shows Micron test results for three types of SSDs (High Capacity, Mainstream, and Performance NVMe) across nine different data center workloads in a multi-node, WekaFS cluster.[3]

Micron 6500 ION NVMe SSD:[4] High Capacity Configuration	
Capacity Range[5]	30.72TB
Form Factors	U.3, E1.L
Design Focus	Best-in-class value vs. competitor's QLC SSDs; better performance, endurance, and QoS — all using 20% less power

Micron 7450 NVMe SSD:[6] Mainstream Configuration	
Capacity Range	400GB to 15.36TB
Form Factors	U.3, M.2, E1.S
Design Focus	Best-in-class latency (sub-2ms QoS); available in a wide range of form factors, capacities, and configurations

Micron 9400 NVMe SSD:[7] Performance Configuration	
Capacity Range	6.4TB to 30.72TB
Form Factor	U.3
Design Focus	Best-in-class performance per TB for the most demanding mixed data center workloads and multi-tenant architectures

Figure 1: SSDs tested

1. The term performance in this document means throughput or IOPS.
2. We used WekaFS in this testing. See https://www.weka.io/ for additional information about all WEKA products.
3. All results measured in Micron test lab using 6-node clusters. Results may vary.
4. Results measured using the Micron 6500 ION SSD (30.72TB). See micron.com/6500 for details.
5. 1GB = 1 billion bytes. Formatted capacity is less.
6. Results measured using the Micron 7450 PRO SSD (7.68TB). See micron.com/7450 for details.
7. Results measured using the Micron 9400 PRO SSD (7.68TB). See micron.com/9400 for details.

Fast Facts

Micron chose WekaFS for this comparison because it is a high-performance, scalable software storage solution. According to WEKA, their software-only, high-performance, file-based storage solution is highly scalable and easy to deploy, configure, manage, and expand.

High-capacity SSDs like the Micron 6500 ION in AMD EPYC™ 7003 series servers can help address massive storage requirements, delivering workload results without the compromises that are typically found on the competitor's QLC SSDs.

Mainstream SSDs like the Micron 7450 PRO in AMD EPYC 7003 series servers can help address the performance demands of the many data center workloads that need a balance of density and form factor flexibility.

Performance SSDs like the Micron 9400 PRO in AMD EPYC 9004 series servers can further improve results for performance-critical mixed workloads.

High Capacity NVMe **Configuration Results**

- 245.76TB storage per node (8x 30.72TB SSDs)
- Up to 112 GB/s bandwidth (0.1 read GB/s per TB)
- Up to 5.2 million IOPS (3,526 read IOPS per TB)

Mainstream NVMe **Configuration Results**

- 76.8TB storage per node (10x 7.68TB SSDs)
- Up to 104 GB/s bandwidth (0.2 read GB/s per TB)
- Up to 5.1 million IOPS (11,068 read IOPS per TB)

Performance NVMe **Configuration Results**

- 76.8TB storage per node (10x 7.68TB SSDs)
- Up to 227 GB/s bandwidth (0.5 read GB/s per TB)
- Up to 6.6 million IOPS (14,323 read IOPS per TB)

Testing with standard benchmarks and workloads

This document shows test results for five small-block (4KB) random workloads and four large-block (128KB and 1MB) sequential workloads to illustrate performance differences between High Capacity, Mainstream, and Performance configurations using WekaFS. The highest performance value for each configuration in each test is highlighted in the workload analysis figures.

Twelve workload generation clients each running 32 FIO jobs create workload IO using the Flexible-IO tester (FIO) benchmark.[8] The IO depth for each FIO job on all 12 clients is shown in each performance figure along the horizontal axis (the job count was set to 32). Adjusting the IO depth characterizes low, medium, and high activity within the cluster. Average latency values (in milliseconds, abbreviated ms) are shown for small-block workloads (latency is often less informative for large-block workloads due to their large IO size).[9]

Small-block results for random workloads

Workload performance is summarized in Table 1. IOPS values shown are the maximum measured for that workload across all IO depths. Mixed-workload IOPS values are the sum of read IOPS and write IOPS.

Type of SSD	4KB Random 100% Read IOPS	4KB Random 100% Write IOPS	4K Random 70% Read / 30% Write (Read + Write) IOPS	4K random 50% Read / 50% Write (Read + Write) IOPS	4KB Random 30% Read 70% Write (Read + Write) IOPS
High Capacity	5.2 million	1.1 million	2.3 million	1.8 million	1.4 million
Mainstream	5.1 million	1.1 million	2.4 million	1.8 million	1.4 million
Performance	6.6 million	2.4 million	4.1 million	4.2 million	3.1 million

Table 1: Random workload maximum IOPS comparisons, entire cluster

Large-block results for sequential workloads

Workload performance is summarized in Table 2. GB/s values shown are the maximum measured for that workload across all IO depths.

Type of SSD	128KB Seq. 100% Read	128KB Seq. 100% Write	1MB Seq. 100% Read	1MB Seq. 100% Write
High Capacity	77 GB/s	24 GB/s	112 GB/s	26 GB/s
Mainstream	73 GB/s	24 GB/s	104 GB/s	26 GB/s
Performance	185 GB/s	84 GB/s	227 GB/s	109 GB/s

Table 2: Sequential workload maximum GB/s comparisons, entire cluster

Server configurations should match their intended use

Deployment and configuration optimization requires a balance of capacity, flexibility, performance, and cost. This is especially true when budgets are a concern. There may be specific use cases that require a higher-density configuration, others that require form factor flexibility, and still others where high-performance storage and the latest generation servers are ideal.

Table 3 highlights hardware components in the High Capacity, Mainstream, and Performance NVMe SSD configurations noted below. All are single CPU, AMD EPYC designs with 200 Gb/s network adapters (one NIC used in the High Capacity and Mainstream NVMe server and two NICS in each Performance NVMe test platform), and the specific SSDs listed below. The CPU core count and DRAM quantity, SSD type and count, networks, and WekaFS software configurations intentionally differ to reflect each configuration's typical use case.

Configuration	Server (Qty.)	CPU and DRAM	SSD (Qty.)	Networking (Qty.)	WekaFS Configuration
High Capacity	1 socket AMD (6 servers)	AMD EPYC 74F3 (24 cores) 8x 32GB Micron DDR4 (256GB total)	Micron 6500 ION NVMe SSD 30.72TB (8 per node)	NVIDIA Mellanox® ConnectX®-6 (200 Gb/s x1)	19 CPU cores
Mainstream	1 socket AMD (6 servers)	AMD EPYC 74F3 (24 cores) 8x 32GB Micron DDR4 (256GB total)	Micron 7450 NVMe SSD 7.68TB (10 per node)	NVIDIA Mellanox ConnectX-6 (200 Gb/s x1)	19 CPU cores
Performance	1 socket AMD (6 servers)	AMD EPYC 9654 (96 cores) 12x 64GB Micron DDR5 (768GB total)	Micron 9400 NVMe SSD 7.68TB (10 per node)	NVIDIA Mellanox ConnectX-6 (200Gb/s x2)	38 CPU cores

Table 3: Server configuration details

8. See https://fio.readthedocs.io/en/latest/#
9. Different FIO configurations may produce different results

Cluster capacities and costs differ[10]

The total procurement cost for cluster hardware depends on several factors. Figure 2 compares each configuration's relative procurement cost versus the terabytes of total raw capacity. Cluster cost is shown as the sum of server acquisition cost in orange, networking acquisition cost in light blue, and storage acquisition cost in green. Note that Figure 2 is not an all-inclusive configuration cost, as it does not include many typical costs, such as network switches, cabling, or software licensing costs.[10]

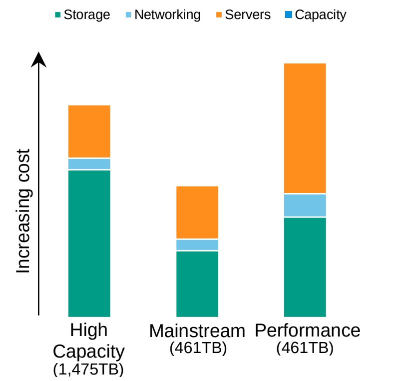

Figure 2: Relative cluster hardware cost and total raw capacity

Servers: The cost of the AMD EPYC 7003 series servers in the High Capacity and Mainstream NVMe WekaFS configurations are the same, while the AMD EPYC 9004 series servers used in the Performance configuration are much higher. Higher AMD EPYC 9004 series server costs are the primary contributor to the higher total cost of the Performance NVMe configuration (server costs reflect the platform, CPU, and DRAM installed).

Network: The Performance NVMe configuration's network cost is double that of the High Capacity and Mainstream NVMe configurations, as each server in the Performance configuration uses two NICs, while the High Capacity and Mainstream configurations each use only one NIC.

Storage: The High Capacity NVMe configuration's storage cost is the highest of the three, which is expected since it delivers 3X the capacity of the other two configurations. The Micron 6500 ION SSDs in the High Capacity configuration offer 1,475TB of raw cluster storage (973TB usable), while the SSDs in the Mainstream and Performance configurations each offer 461TB of raw cluster capacity (304TB usable).

Conclusion

Data center workloads place a wide variety of performance, configuration, and capacity demands on storage. Some deployments demand massive capacity with good workload performance, while others thrive on capacity and form factor flexibility, and still others are IOPS and latency-sensitive. Some organizations have the budget to deploy the fastest, latest technology while others are more budget constrained and must strike a delicate balance between capacity, economy, and performance.

The right solution depends on workload, performance, budget, and capacity requirements.

High Capacity

Some deployments may demand extreme capacity without sacrificing performance. When capacity is paramount, the Micron 6500 ION offers 3X the capacity of the other configurations with performance similar to the Mainstream NVMe configuration, at a cost that lands between the Performance and Mainstream configurations. If your budget is constrained and your storage needs are growing, a High Capacity NVMe configuration is likely a good fit.

Mainstream

Not every WekaFS cluster requires the need for either extra capacity or high performance, and Mainstream NVMe WekaFS configurations are ideal for more balanced configurations that have lower capacity and/or performance per TB requirements, while also offering the broadest range of SSD capacity points and form factor options.

Performance

The Performance configuration using Micron 9400 NVMe SSDs in AMD EPYC 9004 series servers shows the highest performance results across all tested workloads (as expected). This configuration is best when performance is a differentiator and a dominant requirement, particularly for more mixed workloads and/or when IOPS/TB or bandwidth/TB is a relevant metric.

The Bottom Line: Micron NVMe SSDs and AMD EPYC servers are available to meet the needs of all three configurations below. Running WEKA software can help simplify and scale the deployment of the all-flash, all-NVMe infrastructure that is needed in data centers today.

To learn more about each of the Micron SSDs used in this testing, visit:
micron.com/6500ION
micron.com/7450
micron.com/9400

10. All cost information is approximate and based on publicly available information as of the date of this document's publication. Actual costs may vary.

Performance Test Results

High Capacity NVMe configuration results are shown in teal, Mainstream results are shown in purple, and Performance results are shown in deep pink. Higher IOPS (or GB/s) and lower latency is better.

4KB Random 100% Read

The Mainstream and Performance configuration results are remarkably similar at a job IO (queue depth) of 1 (average read latency values are also similar).

As the job IO depth increases, the Performance configuration's IOPS and latency improvements become clear.

The High Capacity configuration reaches a maximum performance of more than 5.2 million IOPS at a job IO depth of 8. The Mainstream configuration reached a maximum performance of 5.1 million IOPS at the same job IO depth. The High Capacity and Mainstream configurations also reflect similar latency.

The Performance configuration reached a maximum of 6.6 million IOPS at a job IO depth of 16. The Performance configuration reflects consistently higher IOPS and lower average read latency for all job IO depths of 4 or more.

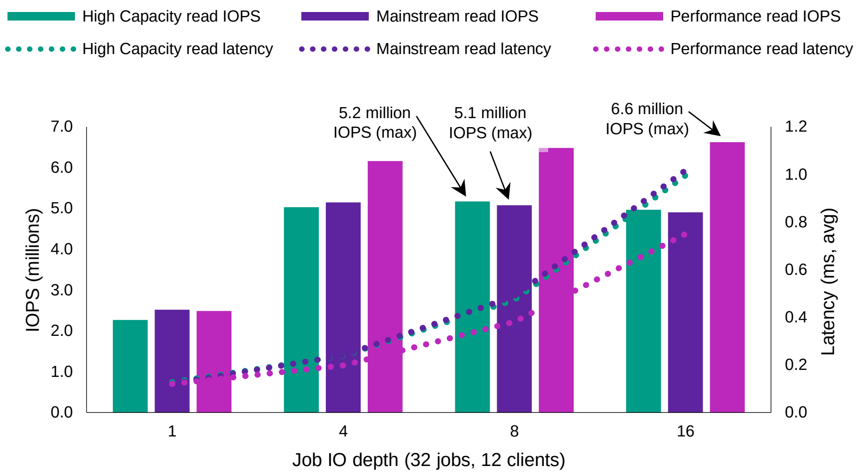

Figure 3: 4KB random 100% read

4KB Random 100% Write

The performance and latency differences between the High Capacity and Mainstream configurations are small at all tested job IO depths. When they are compared to the Performance configuration results, the differences are clear across all job IO depths as seen in Figure 4.

The High Capacity and Mainstream configurations each deliver 1.1 million 4KB random write IOPS starting at a job IO depth of 4 and exhibit similar performance through job IO depths of 8 and 16.

The Performance configuration results are much better than the High Capacity and Mainstream configuration's performance at all tested job IO depths, reaching its maximum of 2.4 million IOPS at job IO depth of 16.

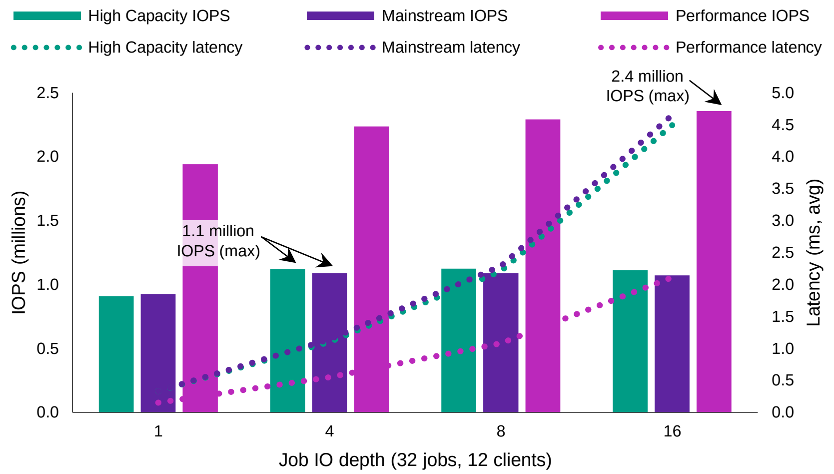

Figure 4: 4KB random 100% write

Figure 4 also shows average write latency results for all three configurations, with a nearly identical value at job IO depth of 1. The High Capacity and Mainstream configurations show higher latency at job IO depths of 4, 8, and 16. The Performance configuration's average latency at higher job IO depths is less than half of the average latency of the High Capacity and Mainstream configurations at the same job IO depths.

4KB Random 70% Read 30% Write

The IOPS values reflected in Figure 5 are the sum of read and write IOPS for each configuration at each job IO depth.

The High Capacity and Mainstream configurations deliver impressive performance ranging up to 2.3 million IOPS for the High Capacity configuration and 2.4 million IOPS for the Mainstream configuration, both at job IO depth of 8.

The Performance configuration delivers a maximum of 4.1 million IOPS at job IO depth of 16.

Read latency values are relatively close at job IO depths of 1 and 4, diverging at 8, and substantially diverging at 16 where the Performance configuration shows a much lower average read latency compared to the High Capacity and Mainstream configurations.

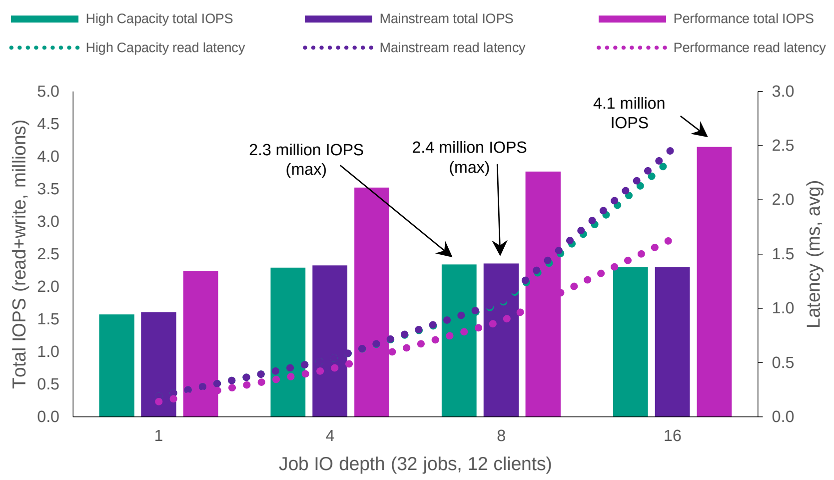

Figure 5: 4KB mixed 100% random 70% read, 30% write

4KB Random 50% Read 50% Write

The IOPS values reflected in Figure 6 are the sum of read and write IOPS for each configuration at each job IO depth.

Again, the High Capacity and Mainstream configurations deliver impressive performance of up to 1.8 million IOPS at a job IO depth of 8.

As expected, the Performance configuration delivers substantially higher results of 4.2 million IOPS at job IO depth 16.

Read latency values are relatively close at job IO depth of 1. They begin to diverge at job IO depth 4 and continues diverging through job IO depths of 8 and 16 where the Performance configuration latency is much lower than the High Capacity or Mainstream values.

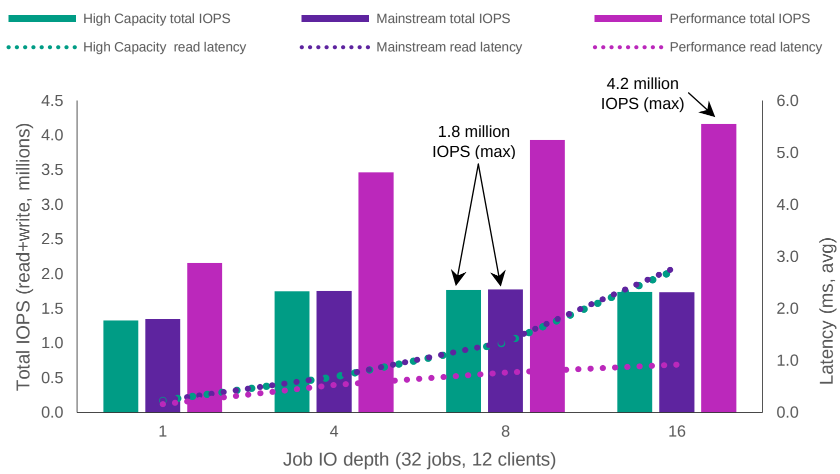

Figure 6: 4KB mixed 100% random 50% read, 50% write

4KB Random 30% Read 70% Write

The IOPS values reflected in Figure 7 are the sum of read and write IOPS for each configuration at each job IO depth.

The High Capacity and Mainstream configurations deliver impressive performance of 1.4 million IOPS at job IO depth of 8, while the Performance configuration delivers 3.1 million IOPS at job IO depth 16.

Read latency values diverge at job IO depth 4 and beyond. The High Capacity and Mainstream configurations show a steady average read latency increase from job IO depth of 1 through 16 while the Performance configuration's average read latency is much more consistent through all job IO depths.

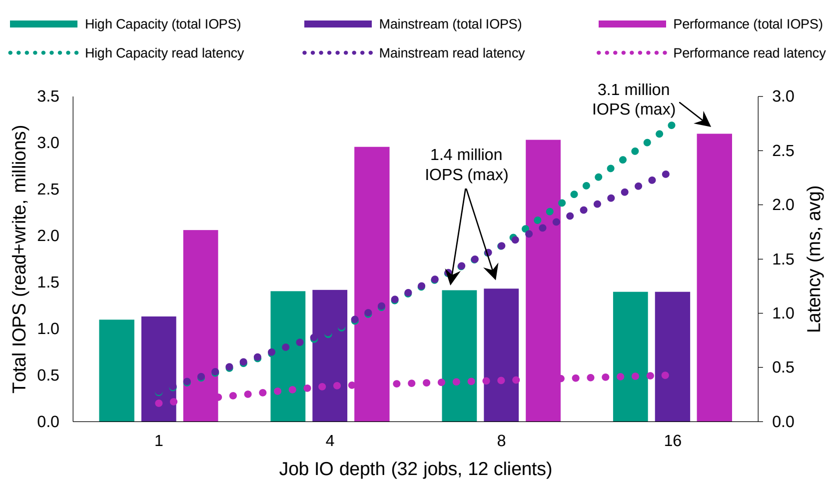

Figure 7: 4KB mixed 100% random 30% read, 70% write

128KB and 1MB Sequential 100% Read

The 128KB and 1MB sequential 100% read values are reflected in Figures 8 and 9. Latency is not shown in these figures because the transfer sizes are large. IO job depths greater than 8 are not shown as there is no performance improvement with these larger IO job depths.

Figure 8 reflects the performance differences between the High Capacity and Mainstream configurations compared to the Performance configuration. The Performance configuration shows higher results across all tested job IO depths.

The High Capacity configuration reaches a maximum 128KB sequential read throughput of 77 GB/s while the Mainstream configuration reaches a maximum of 73 GB/s.

The Performance configuration offers substantially higher performance with a maximum of 185GB/s at job IO depth 8.

1MB sequential 100% read throughput reaches maximum values of 112GB/s and 104GB/s for the High Capacity and Mainstream configurations at job IO depth 1. The Performance configuration reached a maximum of 227GB/s at the same job IO depth.

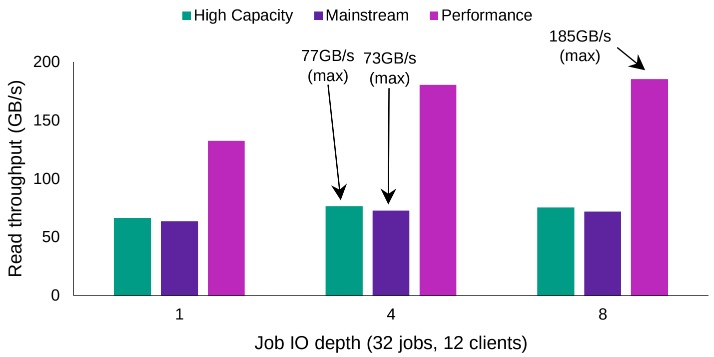

Figure 8: 128KB 100% sequential 100% read

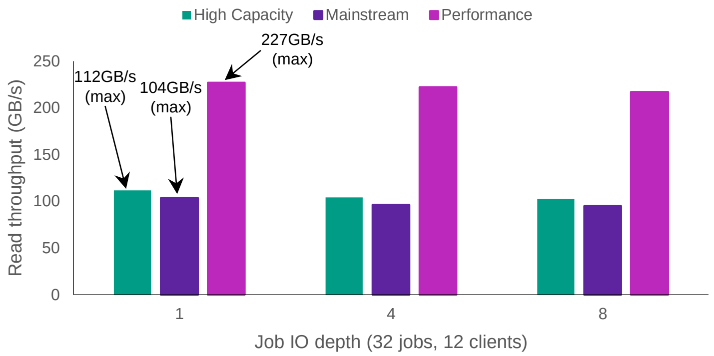

Figure 9: 1MB 100% sequential 100% read

128KB and 1MB Sequential 100% Write

The 128KB and 1MB sequential 100% write values are shown in Figures 10 and 11. Latency is not shown in these figures because the transfer sizes are large.

Both the High Capacity and Mainstream configurations reach maximum 128KB write throughput of 24GB/s at each tested job IO depth. The Performance configuration reached a maximum of 84 GB/s at job IO depth 8.

1MB sequential write performance is again consistent for both the High Capacity and Mainstream configurations across all job IO depths and delivers 26 GB/s of throughput. The Performance configuration reaches a maximum of 109 GB/s.

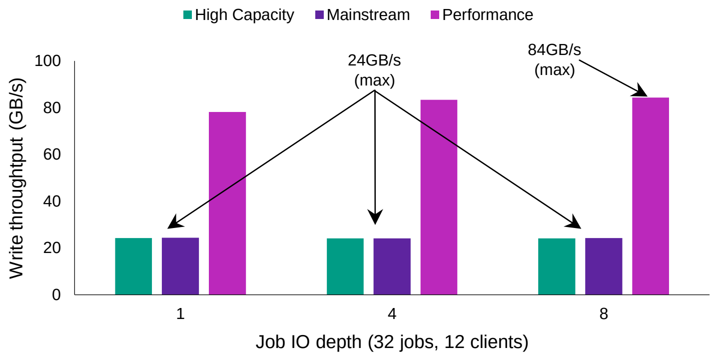

Figure 10: 128KB 100% sequential 100% write

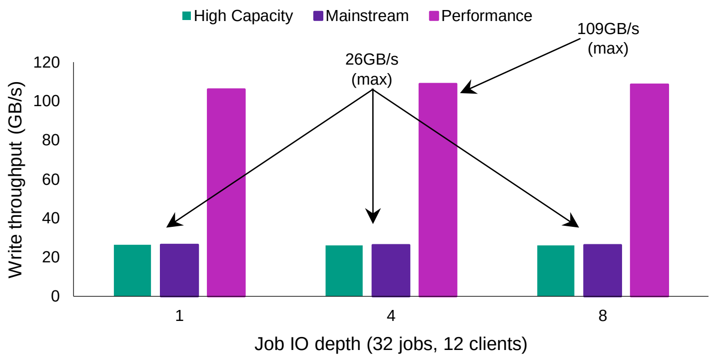

Figure 11: 1MB 100% sequential 100% write

Test Configurations

A synthetic, mixed workload generated these test results. All tests started from the same fixed cluster and SSD state.

Twelve client nodes provide enough demand to ensure that the clients do not limit the performance of the cluster. These client nodes are added to the WEKA cluster as dedicated client nodes with 200 Gb/s NVIDIA Mellanox Connect X-6 NICs. The FIO execution jobs target separate client folders on the shared file system on all the 12 client nodes. Job IO depth increased while the number of FIO jobs was fixed at 32 for each test.

Hardware and Software Configuration

WekaFS High-Capacity and Mainstream Storage Nodes		WEKA Clients		Switch
SSD	**High Capacity:** Micron 6500 ION, 30.72TB, 8 per node **Mainstream:** Micron 7450 PRO, 7.68TB, 10 per node	Micron 7450 PRO 480GB (M.2)		NVIDIA Mellanox SN4700 Ethernet Switch (32 ports @ 400 Gb/s Ethernet)
Server	1U, Single CPU socket	Server	1U, Single CPU socket	
CPU	1x AMD EPYC 74F3	CPU	1x AMD EPYC 74F3	
BIOS	Version: 2.3	BIOS	Version: 2.3	
DRAM (256GB)	8x 32GB Micron DDR4 RDIMMs, 3200 MT/s	DRAM (256GB)	8x 32GB Micron DDR4 RDIMMs, 3200 MT/s	
Network	1x NVIDIA Mellanox ConnectX-6 (200 Gb/s) (MCX653105A-HDA_Ax)	Network	1x NVIDIA Mellanox ConnectX-6 (200Gb/s) (MCX653105A-HDA_Ax)	

Table 4: High Capacity and Mainstream hardware configuration

WekaFS Performance Storage Nodes		WEKA Clients		Switch
SSD	**Performance:** Micron 9400 PRO 7.68TB, 10 per node	Micron 7450 PRO 480GB (M.2)		NVIDIA Mellanox SN4700 Ethernet Switch (32 ports @ 400 Gb/s Ethernet)
Server	1U, Single CPU socket	Server	1U, Single CPU socket	
CPU	1x AMD EPYC 9654	CPU	1x AMD EPYC 74F3	
BIOS	Version 1.0	BIOS	Version: 2.3	
DRAM (256GB)	12x 64GB DDR5 RDIMMs, 4800 MT/s	DRAM (256GB)	8x 32GB DDR4 RDIMMs, 3200 MT/s	
Network	2x NVIDIA Mellanox ConnectX-6 200 Gb/s (MCX653105A-HDA_Ax)	Network	1x NVIDIA Mellanox ConnectX-6 200 Gb/s (MCX653105A-HDA_Ax)	

Table 5: Performance hardware configuration

Software Details		Additional Hardware Details	
OS	Rocky Linux 8.6 Kernel: 4.18.0-372.9.1	BIOS Config	High performance mode enabled. AMD SVM disabled (virtualization). Processor hyper-threading enabled.
Filesystem	WekaFS (version: 4.0.2)	MLNX Settings	Additional (high performance) PCI_WR_ORDERING=1 ADVANCED_PCI_SETTINGS=1
OFED Driver	Nvidia Mellanox OFED version: 5.6-2.0.9.0		
WEKA Software	4.0.2		
FIO Software	3.29		

Table 6: Software and additional WEKA storage node hardware details

micron.com/6500ION
micron.com/7450
micron.com/9400