Micron X100 SSD Transforms Performance on the Aerospike NoSQL Data Platform

The emergence of Micron's X100 NVMe™ SSD, based on 3D XPoint™ technology, marks a new era in storage-class memory (SCM) that enables demanding applications to reach new levels of performance unattainable with conventional NAND storage.

As fast decisions performed closer to data become a critical need for more enterprise applications, architectures with high throughput and extremely low latency over large volumes of data are required. **Aerospike**, a NoSQL database optimized for flash storage, has become one of the fastest NoSQL databases underpinning time-critical web applications like fraud detection, recommendation engines, ad exchange and stock trading.

The high-performance Micron 9300 SSD with NVMe is already among the fastest NAND-based SSDs on the Aerospike certifications list and a great match for performance-focused workloads[1]. But **our testing of the Micron X100 in the Aerospike environment revealed revolutionary performance.** And the Micron X100 refused to slow down as system loads increased.

With a flash-optimized workload, the Micron X100 achieved 8 GB/s of storage throughput at 19 µs of latency. We also saw that the X100 outperforms the Micron 9300 SSD in database operations per second (ops/s) at dramatically lower latency. Across all workloads, we saw tenfold improvements in 99.9% tail latency along with more than two times higher database operations per second.

"The difference in latency is the game changer with the Micron X100. Its *worst* latency is lower than the *best* latency of the top NAND-based SSDs. In the field, this 3D XPoint product delivers fast, stable performance and high QoS even for terabyte-scale databases with dynamically changing data,"

— Raj Hazra, Senior VP Emerging Products, Micron

Micron X100 3D XPoint NVMe SSD vs. fastest NAND-based SSD*:

- 10x improvements in **99.9%** latency, across **all workloads**
- 8 GB/s of **storage throughput**
- At 19 µs of **latency**, an average of 5x to 10x lower
- For more than 2x higher **database ops/s**
- 2x performance advantage in database ops/watt for higher **power efficiency**
- **CPU resource utilization** up to only 40% vs. up to 67% of the NAND-based SSD

**On Aerospike, vs the Micron 9300 MAX 3.2TB NAND-based SSD*

About This White Paper

These test results compare the NoSQL database performance[2] of the Micron X100, our PCIe Gen3 x16 3D XPoint storage device, with the performance of the Micron 9300 MAX 3.2TB NAND-based SSD with NVMe using Aerospike Database Enterprise Edition v 5.2.0.4 and the Yahoo Cloud Serving Benchmark (YCSB). Test results shown are organized by YCSB workload. (A brief workload description and example use case accompany the test results. These descriptions are based on the YCSB descriptions from GitHub.)

1. Data retrieved from aerospike.com on Dec. 2, 2020.
2. The terms "performance" and "database operations per second" are used interchangeably in this paper.

The **X100 NVMe SSD** is Micron's first product in a new family of high-performance memory solutions based on 3D XPoint technology. It delivers game-changing performance with industry-leading high bandwidth, low latency, high quality of service (QoS) and high endurance.

The **Micron 9300 SSD** with NVMe delivers industry-leading sequential write performance with some of the lowest average write latency of any NAND-based SSD on the market. It accelerates a broad range of applications to help meet today's growing data demands.

Industry Background

The most compelling NoSQL database deployments are now built with SSDs. Where performance is the primary driver, high-performance NAND-based SSDs with NVMe have become the preferred storage option, especially when the entire dataset does not fit into memory (DRAM).

The success of NVMe SSD-based NoSQL deployments has increased demand for those databases. As more sophisticated applications drive greater demand, storage performance is again in focus. In fact, as software has matured to take advantage of fast storage, NAND-based SSDs may not provide enough capabilities due to constrained throughput or latency inherent in NAND when compared to memory architectures.

Micron chose Aerospike for this comparison because of its flash-optimized design, strong performance and ultra-low latency. These deployments thrive with the Micron X100 in the proper configurations.

About Aerospike NoSQL Data Platform

According to aerospike.com, "Aerospike enterprises overcome seemingly impossible data bottlenecks to compete and win with a fraction of the infrastructure complexity and cost of legacy NoSQL databases. Aerospike's patented Hybrid Memory Architecture™ delivers an unbreakable competitive advantage by unlocking the full potential of modern hardware, delivering previously unimaginable value from vast amounts of data at the edge, to the core and in the cloud. Aerospike empowers customers to instantly fight fraud; dramatically increase shopping cart size; deploy global digital payment networks; and deliver instant, one-to-one personalization for millions of customers."

About YCSB

Micron chose the Yahoo! Cloud Serving Benchmark (YCSB) because it is an open-source benchmark. It is supplied with several standard workloads that simulate common use cases for NoSQL databases, enabling us to test a broad range of workloads at multiple worker counts (threads). We used 4KB record size and 97 million records (about 400GB) for the YCSB database. We used uniform distribution for workloads A, B, C and F.

About Reporting Latencies

Before we dive into the results, we have this note about latency measurement. We compare the average, 99% and 99.9% read latencies for all workloads to focus on how customers measure their workload targets and meet service level agreements (SLAs). This methodology highlights the performance differences between the Micron X100 and Micron 9300, contrasting latency differences of the devices and their media types.

In all figures, higher database operations per second (DB ops/s) and lower latency are better.

Workload: 50% Read, 50% Update

This YCSB workload approximates a session store recording recent actions, or a metadata server under a heavy load operation.

Testing for Performance and Average Latency

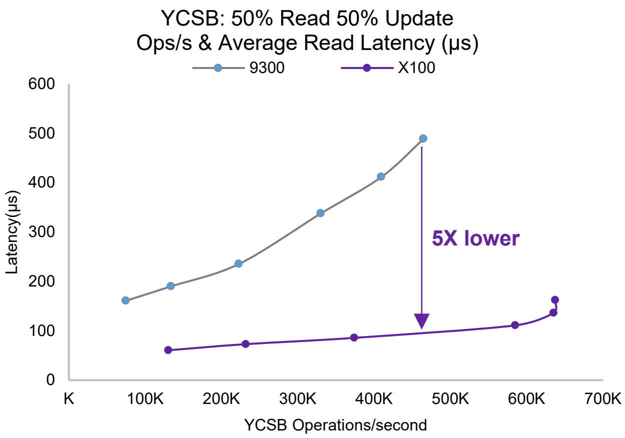

Figure 1: Micron X100 and 9300 – 50% Read, 50% Update

Figures 1 through 3 show average read latency on the y-axis (microseconds) and database operations per second on the x-axis. In all figures, lower is better.

Each point shows the database performance at a specific YCSB thread count, scaling across 8, 16, 32, 64, 96 and 128. This shows the benefit of the X100 in 50% Read, 50% Update, both in increased database operations per second and reduced average latency across a range of results.

Note that at 460K ops/s, the Micron 9300 latency is 500 µs while the Micron X100 is 100 µs, a fivefold improvement.

The Micron X100 reaches maximum performance at 128 threads and shows 162 µs latency. This is equal to the lowest latency value measured on the Micron 9300. Comparing similar thread counts, the X100 achieves up to 1.8 times the operations per second (at thread count 64) at one-quarter the read latency across the range of thread counts shown.

Testing for 99% Read Tail Latencies

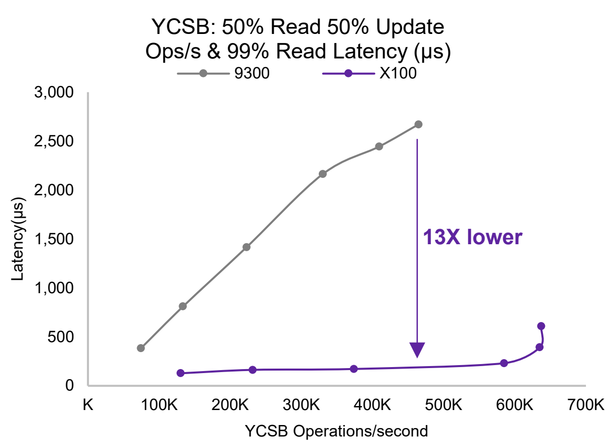

Figure 2: Micron X100 and 9300 – 99% Read Tail Latency – 50% Read, 50% Update

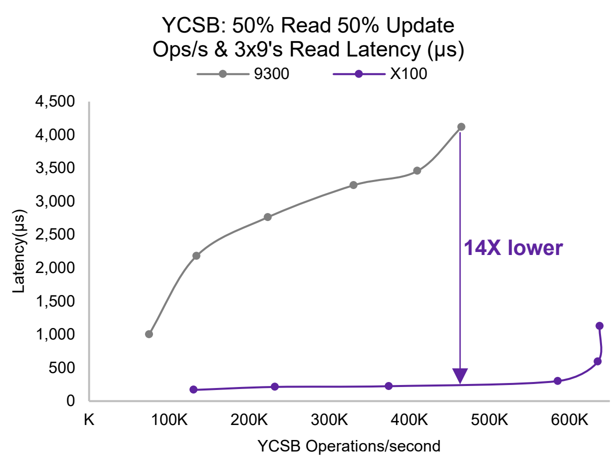

Figure 3: Micron X100 and 9300 – 99.9% Read Tail Latency – 50% Read, 50% Update

Tail latency values (for all YCSB workloads) should be compared to see the application's QoS as the database becomes more heavily loaded. For example, at 460K ops/s, the maximum performance for the NVMe SSD, we see 4 milliseconds 99.9% read latency. On the X100 at the same 460K ops/s, we have 290 µs latency. The NVMe SSD read latency is 14 times higher at that load. We also see the X100 maintains low 99.9% read latency as load increases, while the NVMe SSD's 99.9% read latency increases somewhat linearly.

The performance of the X100 translates to a more stable application QoS as load increases on the system.

Workload: 95% Read, 5% Write

This YCSB workload approximates photo tagging. Adding a tag is considered an update, but most operations are to read tags.

Testing for Performance and Average Latency

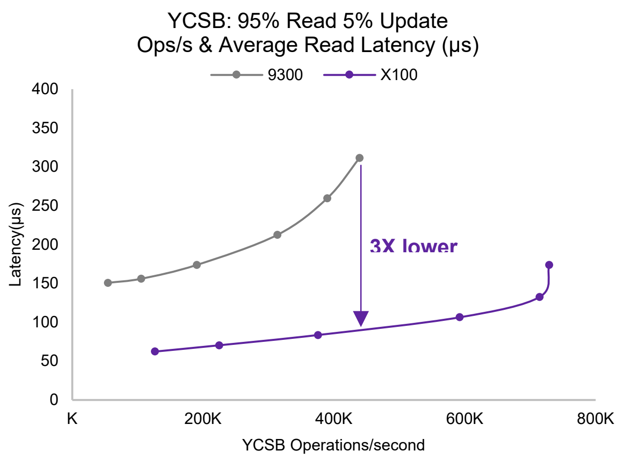

Figure 4: Micron X100 and 9300 – 95% Read, 5% Write

Figures 4 through 6 show read latency in microseconds on the y-axis and database operations per second on the x-axis for the 95% Read, 5% Write workload. In all figures, lower is better.

Each point shows the database performance at a specific YCSB thread count, scaling across 8, 16, 32, 64, 96 and 128.

Figure 4 shows the benefit of the Micron X100 with 95% Read, 5% Write both in double the average increased database operations per second and reduced average latency.

The average read latency of the X100 is one third of the 9300 at its highest performance point at 430K ops/s.

Testing for Read Tail Latencies

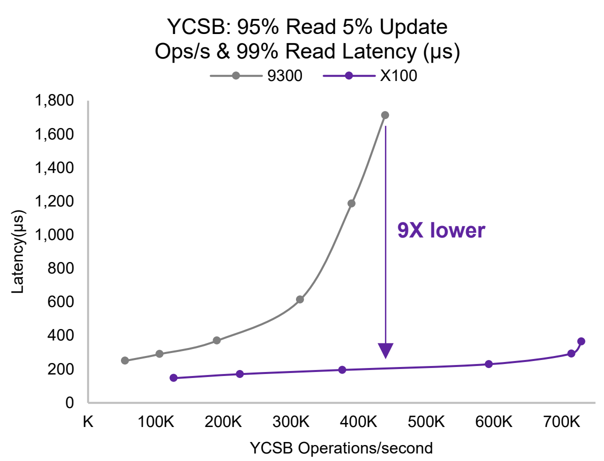

Figure 5: Micron X100 and 9300 – 99% Read Tail Latency – 95% Read, 5% Write

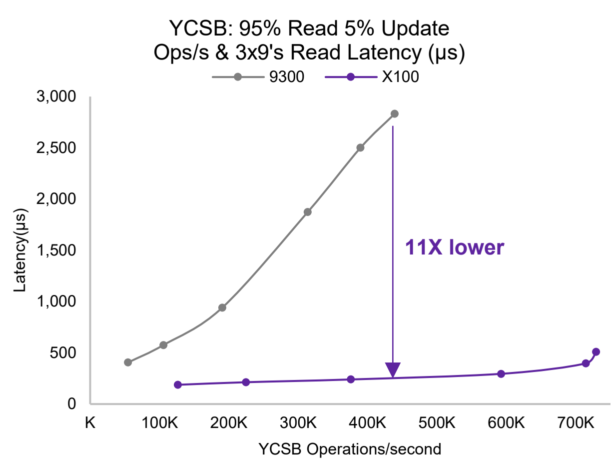

Figure 6: Micron X100 and 9300 – 99% Read Tail Latency – 95% Read, 5% Write

Figures 5 and 6 show that the Micron X100 enables significantly higher operations per second while maintaining much lower 99% and 99.9% tail latency than the Micron 9300.

Workload: 100% Read, 0% Write

This YCSB workload approximates reading user profiles from cache (when the profiles are stored elsewhere).

Testing for Performance and Average Latency

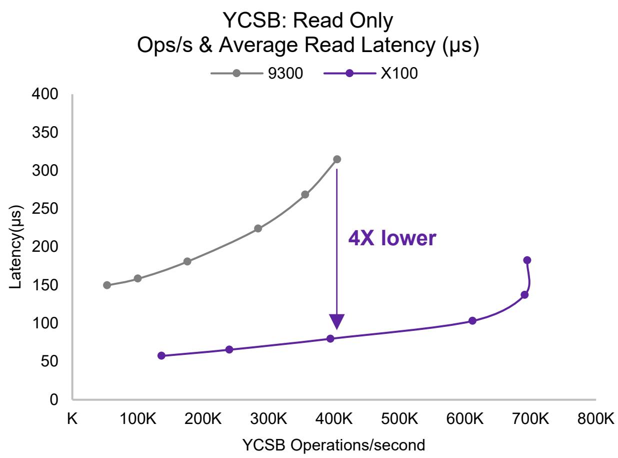

Figure 7: Micron X100 and 9300 – 100% Read

Figures 7 through 9 show read latency in microseconds on the y-axis and database operations per second on the x-axis. In all figures, lower is better.

Each point shows the database performance at a specific YCSB thread count, scaling across 8, 16, 32, 64, 96 and 128.

With 100% Read, 0% Write, the Micron X100 showed 1.7 to 2.6 times the performance of the Micron 9300. The Micron X100 also showed an average of 50% lower read latencies than the Micron 9300.

The X100's average read latency is one quarter of the 9300's latency at its maximum performance level, 128 threads, 400K ops/s.

Testing for Read Tail Latencies

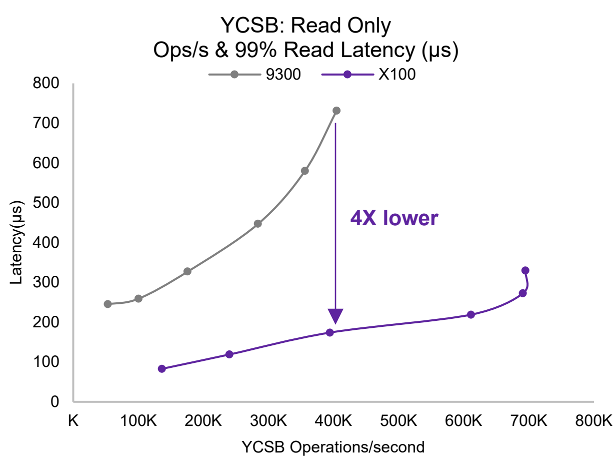

Figure 8: Micron X100 and 9300 – 99% Read Tail Latency – 100% Read

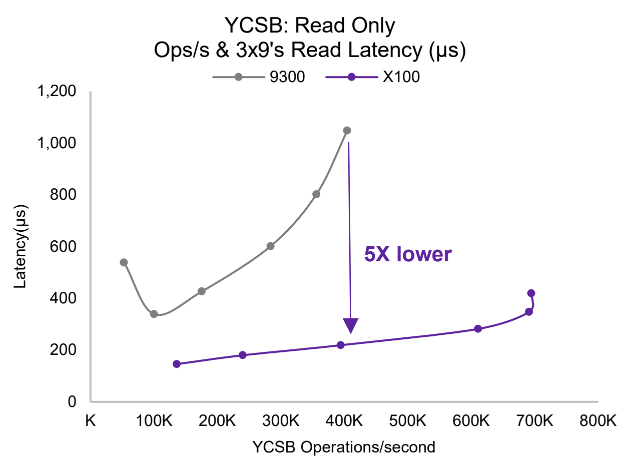

Figure 9: Micron X100 and 9300 – 99.9% Read Tail Latency – 100% Read

Workload: Read Latest

In this YCSB workload, new records are inserted, and then the most recently inserted records are the most popular. This workload models user status updates in which people want to read the latest update.

Testing for Performance and Average Latency

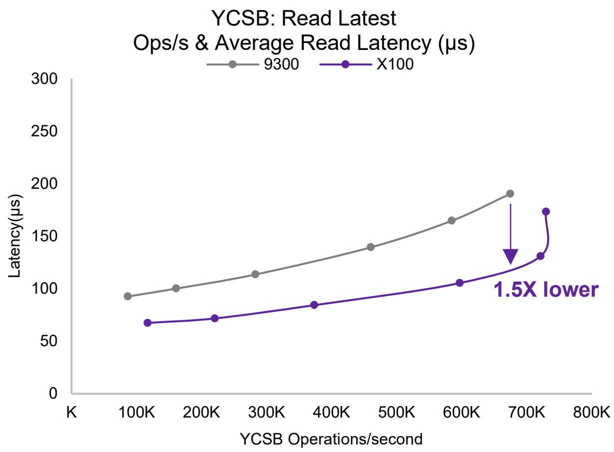

Figure 10: Micron X100 and 9300 – Read Latest

This workload has a Zipfian distribution, meaning the active dataset for the Read Latest workload is much smaller than the dataset used to test other workloads, fitting into main system memory. In all figures, lower is better.

Because of this, the differences in database operations per second for the Read Latest workload database for the Micron 9300 and the Micron X100 are relatively small (the Micron X100 shows about 1.3 times the Micron 9300 database operations per second). However, the 99.9% latency results for the X100 are about 10 times better than those for the Micron 9300.

Figures 10 through 12 show read latency in microseconds on the y-axis and database operations per second on the x-axis.

Each point shows the database performance at a specific YCSB thread count, scaling across 8, 16, 32, 64, 96 and 128. Figure 10 shows that the operations per second differences between the Micron X100 and Micron 9300 are somewhat smaller when compared to the other workloads.

Testing for Read Tail Latencies

In general, the Micron X100 shows about 1.3 times the Micron 9300 database operations per second. Still, the Micron X100 shows a strong benefit in reduced latency when compared to the Micron 9300. For example, the 99.9% latency results for the Micron X100 are about 10 times better than those for the Micron 9300 (Figure 12).

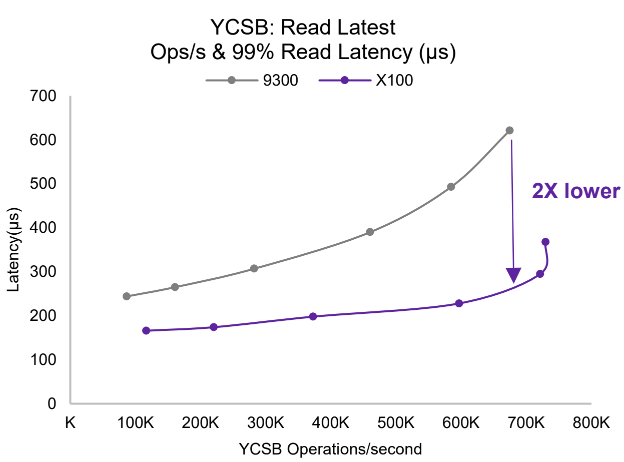

Figure 11: Micron X100 and 9300 -- 99% Read – Read Latest

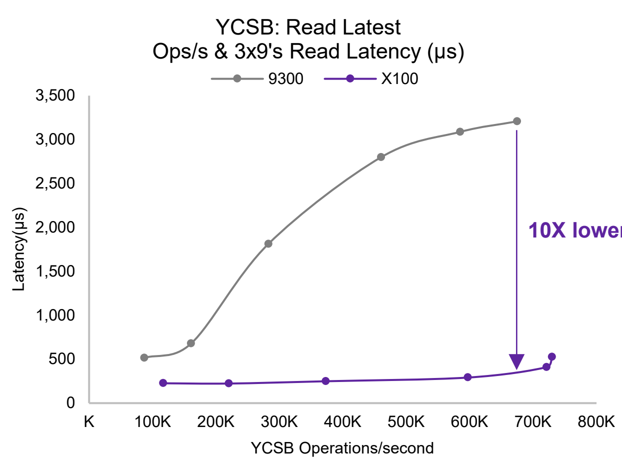

Figure 12: Micron X100 and 9300 – 99.9% Read – Read Latest

Workload: 50% Read, 50% Read-Modify-Write

In this YCSB workload, the client reads a record, modifies it, and writes back the changes. This models a user database application, where user records are read and modified by the user, or user activity is recorded.

Testing for Performance and Average Latency

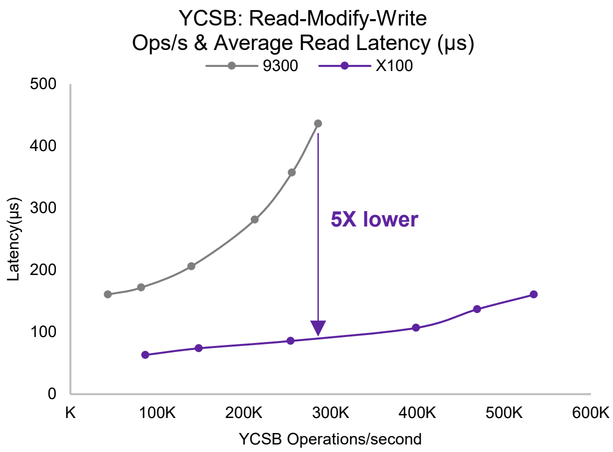

Figure 13: Micron X100 and 9300 – 50% Read, 50% Read-Modify-Write

Figures 13 through 15 show read latency in microseconds on the y-axis and database operations per second on the x-axis. In all figures, lower is better.

Each point shows the database performance at a specific YCSB thread count, scaling across 8, 16, 32, 64, 96 and 128.

The Micron X100 50% Read, 50% Read-Modify-Write workload, performance improvement is very consistent, ranging from a minimum of 1.8 times (16, 32 and 96 threads) to a maximum of 2 times (8 threads).

At 290K ops/s, the max for the 9300, the average read latency on the Micron 9300 is five time higher than the X100.

Testing for Read Tail Latencies

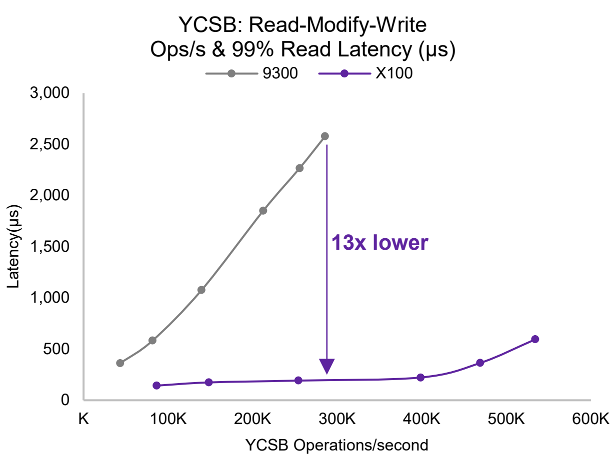

Figure 14: Micron X100 and 9300 – 99% Read – 50% Read, 50% Read-Modify-Write

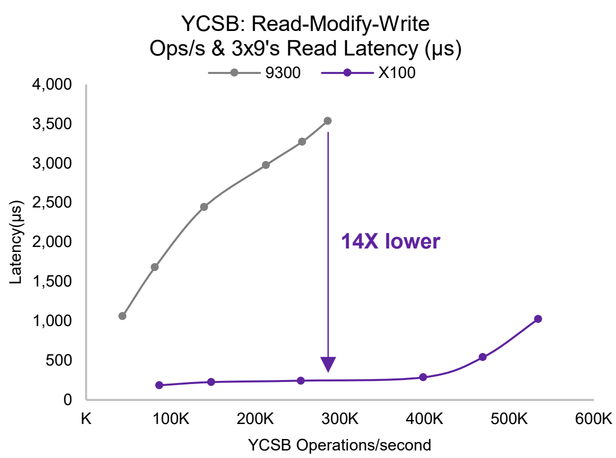

Figure 15: Micron X100 and 9300 – 99.9% Read – 50% Read, 50% Read-Modify-Write

Power Efficiency

The equation of database operations per watt (system power) helps evaluate power efficiency, the amount of useful work completed (database operations) divided by power consumed to accomplish that work (watts). This metric is well-suited when comparing NoSQL deployments where power may be limited or expensive. It also helps analyze additional factors like total cost of ownership (TCO).

When we measured database operations per second for each of the YCSB workloads, we also measured instantaneous power consumption during each test. These two values are used to calculate power efficiency as shown in the following equation:

$$\text{Power Efficiency} = \frac{\text{Database Operations}}{\text{Power Consumed}}$$

Figures 16 through 20 show threads along the x-axis and calculated database operations per watt (ops/watt) along the y-axis. In these figures, higher is better.

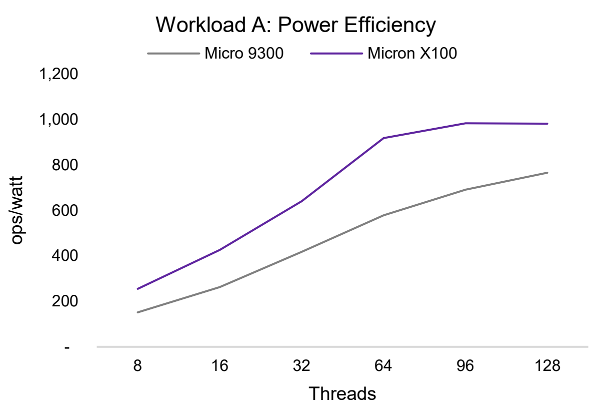

Figure 16: Micron X100 – 1.3x to 1.7x Greater Energy Efficiency – 50% Read, 50% Update

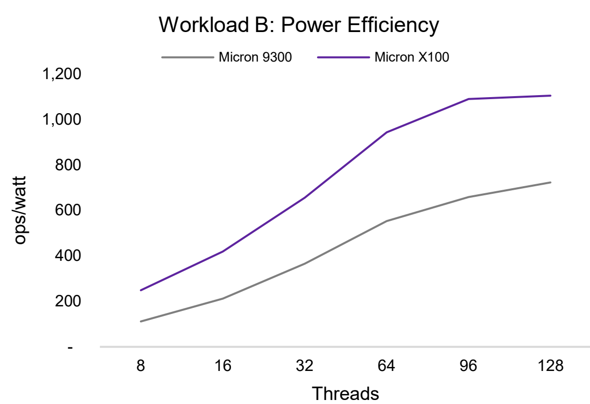

Figure 17: Micron X100 – 1.5xX to 2.2x Greater Energy Efficiency – 95% Read, 5% Write

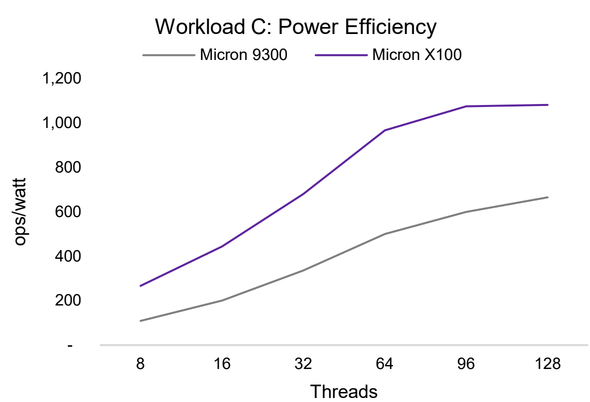

Figure 18: Micron X100 – 1.6x to 2.4x Greater Energy Efficiency – 100% Read, 0% Write

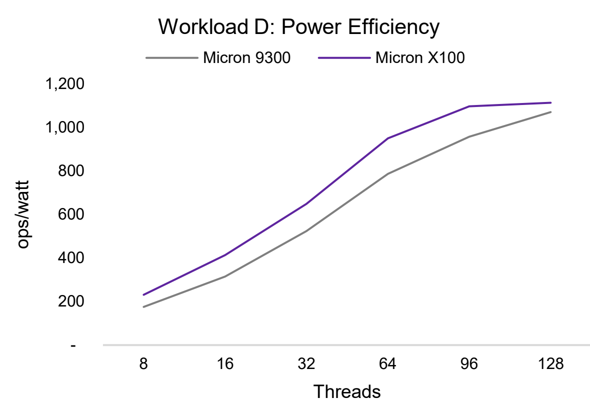

Figure 19: Micron X100 – 1.0x to 1.3x Greater Energy Efficiency – Read Latest Workload

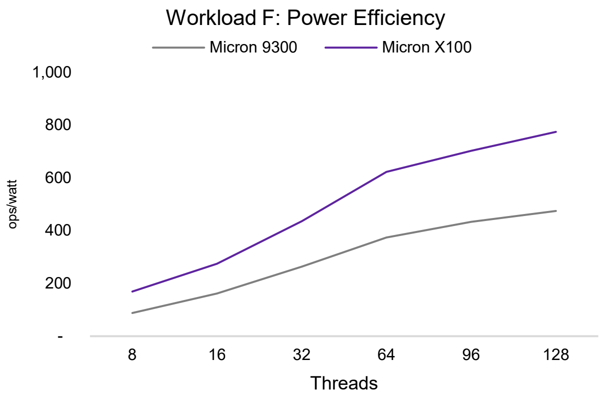

Figure 20: Micron X100 – 1.6x to 1.9x Greater Energy Efficiency – 50% Read, 50% Read, Modify, Write

While the maximum power draw of the Micron X100 is higher than the Micron 9300, the X100 approximately doubles the performance advantage in database operations per watt for most workloads in YCSB. This performance increase results in the observed higher power efficiency for the X100.

While the performance difference in the read latest workload is smaller than the other workloads, the Micron X100 performance increase yields slightly better power efficiency than the Micron 9300 SSD there as well.

CPU Utilization

Figures 21 through 25 show the dramatic difference in the percentage of CPU utilization by workload for the Micron X100 and Micron 9300. The IOWAIT% (the percentage of time the CPU is waiting) consumes up to 50% of the available CPU resources with the Micron 9300, while the IOWAIT% is only a maximum 5% with the Micron X100. This translates to a difference in total CPU utilization: The Micron 9300 uses up to 67% of the CPU resources while the X100 uses only up to 40%. Note that the reduction in X100 CPU utilization also includes almost double the higher database operations per second across most tests.

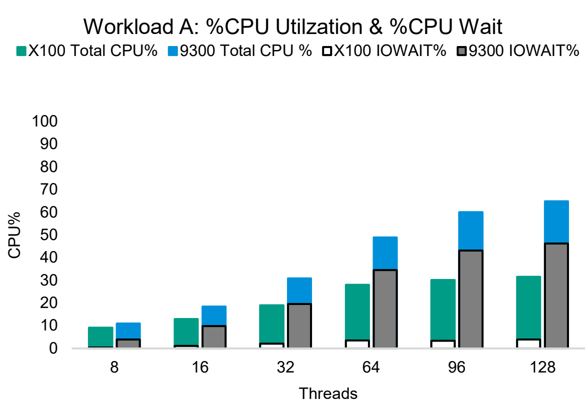

Figure 21: Total %CPU vs. %IOWAIT – 50% Read, 50% Update

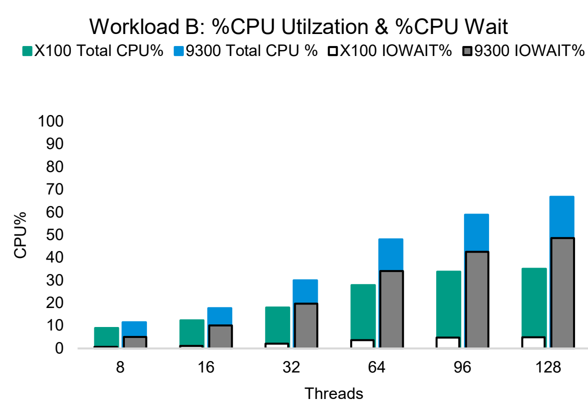

Figure 22: Total %CPU vs. %IOWAIT – 95% Read, 5% Write

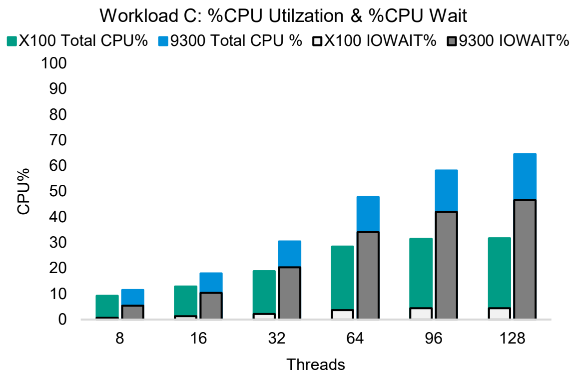

Figure 23: Total %CPU vs. %IOWAIT – 100% Read, 0% Write

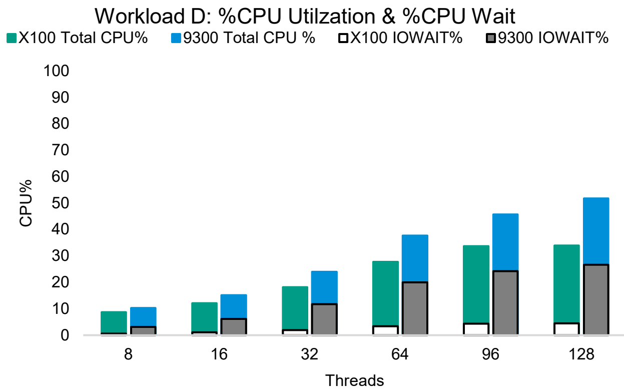

Figure 24: Total %CPU vs. %IOWAIT – Read Latest Workload

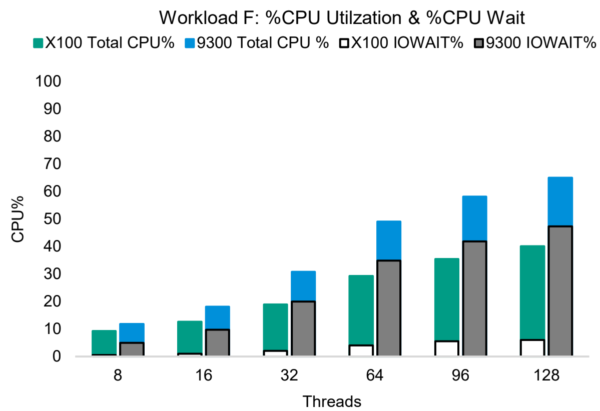

Figure 25: Total %CPU vs. %IOWAIT – 50% Read, 50% Read, Modify, Write

Conclusion

The Micron X100 NVMe SSD with 3D XPoint and an x16 Gen3 PCIe interface demonstrated superior performance across a variety of YCSB workloads compared to the Micron 9300, despite the Micron 9300 being among the fastest NAND-based SSDs certified for use with Aerospike.

Using the Aerospike database, a flash-optimized NoSQL database, the Micron X100 achieved 8 GB/s of storage throughput at 19 µs of latency and far outperformed the Micron 9300 in benchmark operations per second at dramatically lower latency with up to 10 times the improvement in 99.9% latency along with more than two times higher database operations per second.

In addition, the Micron X100 shows lower tail 99% and 99.9% latency across nearly all the tested workloads and uses half the CPU/power resources of the Micron 9300 with far greater power efficiency.

The Micron X100 is a remarkable storage-class memory device that excels when used with applications that are tuned for flash like the Aerospike database.

Learn More

Micron X100 SSD with NVMe: www.micron.com/x100
Micron 9300 SSD with NVMe: www.micron.com/9300
Aerospike database: www.aerospike.com
Yahoo! Cloud Serving Benchmark (YCSB): github.com/brianfrankcooper/YCSB

Appendix A: Storage Device Performance

Aerospike database incorporates a defragmentation operation which triggers when the storage device is getting close to full. With the Micron X100 SSD and a 400GB dataset, this defragmentation operation triggers; it does not trigger with a 400GB dataset on the 3.2TB Micron 9300 SSD (due to lower performance and additional capacity).

Table 1 shows the effects of the Aerospike database defragmentation operation. When executing the same workload, the X100 average input/output (I/O) request is nearly double that of the Micron 9300 average I/O request size This I/O request size doubling translates into these test workloads reflecting a best-case storage I/O for the Micron 9300 and a worst-case storage I/O for the Micron X100.

Figures 26 through 31 show storage device performance while the Aerospike defragmentation process is running.

	Average Request Size	
YCSB Workload	**X100**	**9300**
A – 50% Read, 50% Update	20KB+	10KB
B – 95% Read, 5% Write	8KB	4KB
C – 10% Read, 0% Write	8KB	4.6KB
D – Read Latest	8KB	5KB
F – 50% Read, 50% Read-Modify-Write	15KB	8KB

Table 1: Average I/O Request Size

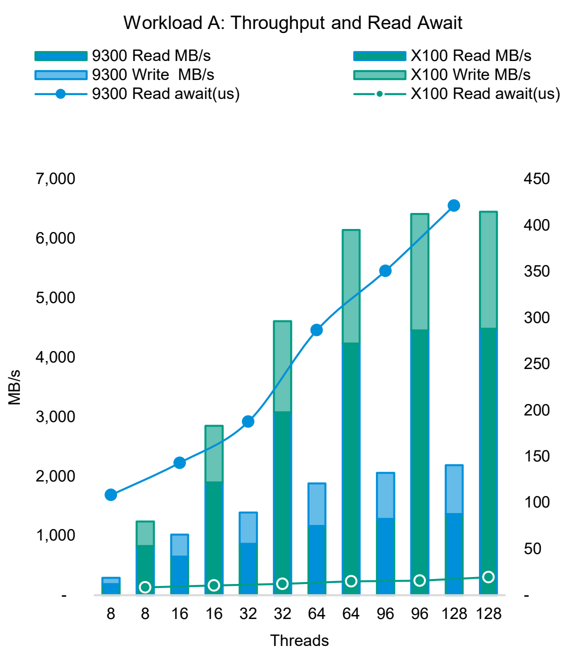

Figure 26: Throughput and Read Await, Workload A

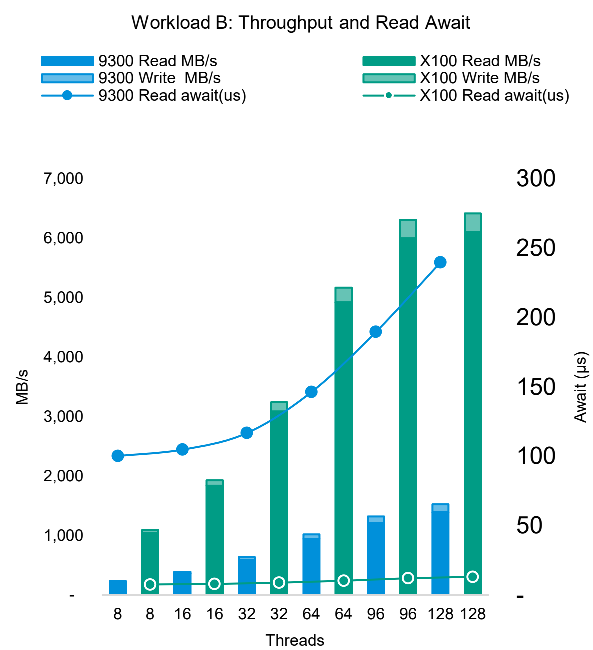

Figure 27: Throughput and Read Await, Workload B

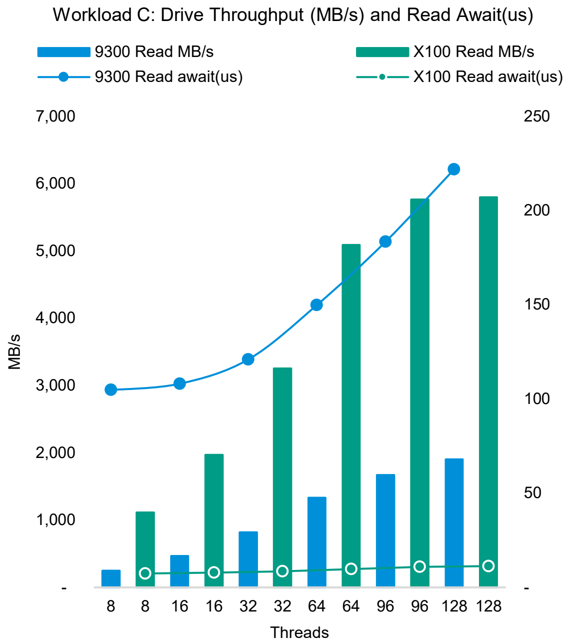

Figure 28: Throughput and Read Await, Workload C

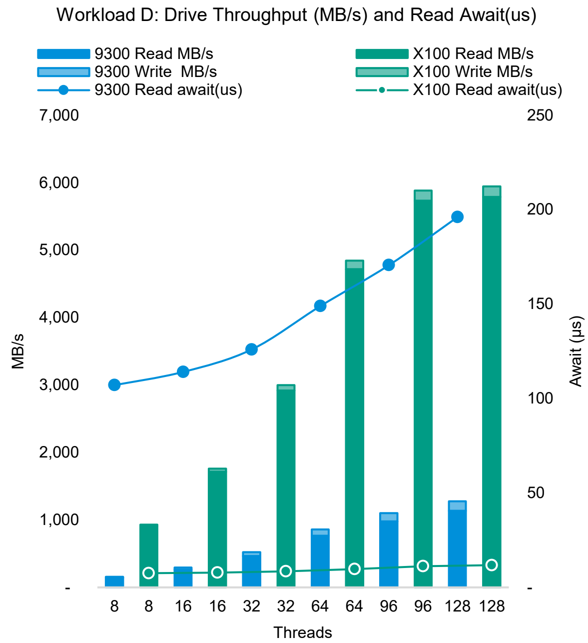

Figure 29: Throughput and Read Await, Workload D

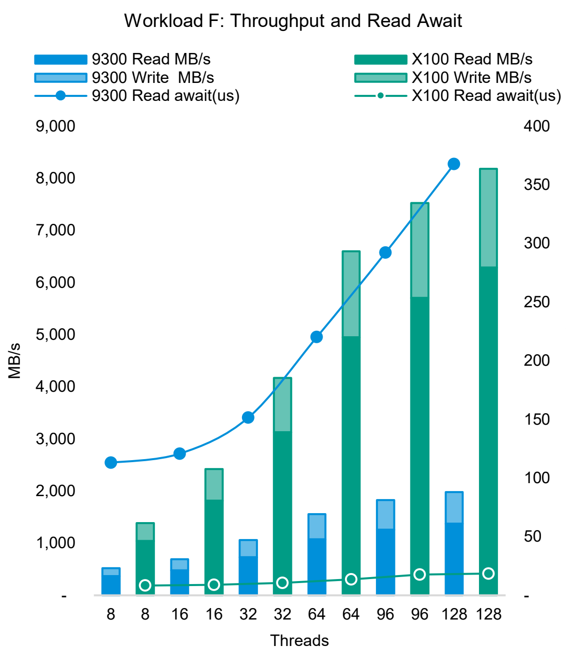

Figure 30: Throughput and Read Await, Workload F

Despite the difference in I/O request size between the Micron X100 and Micron 9300 noted earlier, results show the Micron X100 maintaining an extremely high read and write throughput during testing, roughly double the YCSB performance, while also maintaining extremely low storage device latency (measured by await).

Appendix B: Aerospike Database Configuration Files

B.1 Micron X100 Configuration

```
service {
	paxos-single-replica-limit 1 # Number of nodes where the replica count is automatically reduced to 1.
	proto-fd-max 15000
	feature-key-file /etc/aerospike/features.conf
}

logging {
	file /var/log/aerospike/aerospike.log {
		context any info
	}
}

network {
	service {
		address any
		port 3000
	}

	heartbeat {
		mode multicast
		multicast-group 239.1.99.222
		port 9918

		interval 150
		timeout 10
	}

	fabric {
		port 3001
	}

	info {
		port 3003
	}
}

namespace ycsb {
	replication-factor 1
	memory-size 128G

	storage-engine device {
		device /dev/nvme1n2p1
		device /dev/nvme1n2p2
		device /dev/nvme1n2p3
		device /dev/nvme1n2p4
		device /dev/nvme1n2p5
		device /dev/nvme1n2p6
		device /dev/nvme1n2p7
		device /dev/nvme1n2p8
		device /dev/nvme1n2p9
		device /dev/nvme1n2p10
		device /dev/nvme1n2p11
		device /dev/nvme1n2p12
		device /dev/nvme1n2p13
		device /dev/nvme1n2p14
		device /dev/nvme1n2p15
		device /dev/nvme1n2p16
		write-block-size 128K
	}

	transaction-pending-limit 100
}
```

B.2 Micron 9300 Configuration

```
service {
        paxos-single-replica-limit 1 # Number of nodes where the replica count is automatically reduced to 1.
        proto-fd-max 15000
        feature-key-file /etc/aerospike/features.conf
}

logging {
        file /var/log/aerospike/aerospike.log {
                context any info
        }
}

network {
        service {
                address any
                port 3000
        }

        heartbeat {
                mode multicast
                multicast-group 239.1.99.222
                port 9918

                interval 150
                timeout 10
        }

        fabric {
                port 3001
        }

        info {
                port 3003
        }
}

namespace ycsb {
        replication-factor 1
        memory-size 128G

        storage-engine device {
                device /dev/nvme0n1
                device /dev/nvme0n2
                device /dev/nvme0n3
                device /dev/nvme0n4
                device /dev/nvme0n5
                device /dev/nvme0n6
                device /dev/nvme0n7
                device /dev/nvme0n8
                device /dev/nvme0n9
                device /dev/nvme0n10
                device /dev/nvme0n11
                device /dev/nvme0n12
                device /dev/nvme0n13
                device /dev/nvme0n14
                device /dev/nvme0n15
                device /dev/nvme0n16
                write-block-size 128K
        }

        transaction-pending-limit 100
}
```

Appendix C: Hardware Configuration

Micron's testing lab used the following hardware specifications:

- Database server
  - Dual socket R740XD
    - 2x Intel Xeon Platinum 8168 CPU @ 2.7GHz
    - 24 cores per socket
  - 384GB memory
  - Boot disk: Micron 5100 240GB
  - Database storage:
    - Micron 9300 MAX 3.2TB
    - Micron X100 800GB
  - Network: Mellanox ConnectX-4 100Gbps
- Loadgen
  - Dual socket R740XD
    - 2x Intel Xeon Gold 6142 CPU @ 2.6GHz
    - 12 cores per socket
  - 384GB memory
  - Boot disk: Micron 5300 480GB
  - Network: Mellanox ConnectX-4 100Gbps

Entity	Configuration Description
System software versions	Dell 740XD BIOS version: 2.8.2 Backplane FW: 4.35
BIOS configuration	High-performance mode enabled Intel VT disabled Processor hyper-threading enabled
Dataset storage	9300 configuration: • 3.2TB drive with 16 namespaces (200GB each) X100 configuration: • 800GB drive with 16 partitions (50GB each)
OS settings	sysctl.conf: • vm.min_free_kbytes = 1153434 • vm.swapiness = 0
Aerospike configuration	Records: 97 million Record size: 4096 bytes Storage-engine: Device (data on device, Index in memory) Replication factor: 1 Memory size: 128G Write block size: 128K Transaction pending limit: 100

Table 2: Hardware and Software Configurations for Testing X100 and 9300 SSDs on Aerospike