storONE™
THINK RESULTS™

Intel Optane Flash Array Powered by S1: Lab Report

The Optane Flash Array combines the power and performance of Intel Optane Drives with Intel QLC Drives StorONE's S1:Optane Flash Array (OFA) was recently tested by both StorONE and Intel. The purpose of this test was to validate StorONE's claims of delivering optimal per drive performance while still providing enterprise-class storage services. As a result of the testing exercise, StorONE believes there is a significant opportunity to broaden the appeal of Optane and make it more viable for mainstream enterprise use.

WHO IS STORONE?

StorONE is the developer of S1, the first Enterprise Storage Platform. An enterprise storage platform is a hardware-independent software solution that consolidates all storage use cases (file, block, object, cloud) into a single operating environment. Unlike legacy storage software written over a decade ago, S1 delivers the maximum rated performance of the hardware it runs on. The Platform also provides the complete set of storage services that enterprise data centers demand, like protection from media failure, accidental or malicious data deletion, and data center-wide disaster.

THE PURPOSE OF OFA

Data Centers of all sizes face constant pressure to deliver an architecture that can meet the ever-increasing expectation of high-performance. Intel's Optane Storage Class Memory is an ideal answer to these demands. Public Cloud Providers, Hyperscalers, and Organizations with considerable artificial intelligence and deep learning workloads are already integrating Optane into their environments because the performance gains lead to better decisions that directly impact the organization's bottom line.

Mainstream workloads can also benefit from Optane. These organizations, however, need to balance the gains in performance with the overall cost. To justify the investment in Optane, these organizations need to leverage Optane across as many workloads as possible and reduce the investment in other parts of the infrastructure. QLC flash is an ideal technology to partner with Optane because of its low cost and high density. The challenge is QLC's low write durability threshold. Optane can act as a shock absorber so that QLC's low endurance does not impact the data center's use of it. At the same time though, IT can't afford, yet another process to manage. As a result, the movement of data between these two tiers must be seamless. Once implemented, IT Architects can realize higher performance and even higher cost savings by using a large tier of QLC and a small tier of Optane versus the cost of a TLC-only all-flash array.

UNDERSTANDING THE STORAGE SYSTEM ECO-SYSTEM

A Storage System is a combination of several parts, storage media, storage network, CPUs, and storage software. Extracting maximum performance out of the storage eco-system means removing as many bottlenecks as possible. Before enterprise solid-state drives (SSD) based on flash, storage media was always the most significant source of IO bottlenecks. Flash and now especially, Intel Optane moves that focus elsewhere. The internal network and the protocol the storage system uses to communicate

with the media is also no longer a source of the bottleneck thanks to NVMe which offers PCIe access as well as higher command count and higher queue depth. Even the storage network is no longer the primary bottleneck, thanks to the latest high-bandwidth advances in Ethernet and Fibre-Channel, as well as the potential of NVMe over Fabrics (NVMe-oF).

There has always been enough CPU power to drive the storage software. In fact, until the recent advancements in media and networking, there was almost too much CPU power. Now, however, thanks to the reduction in latency by these advances, the CPU power can be leveraged, but the storage software has to take advantage of modern CPU design. The problem is few vendors have refactored their code to take advantage of modern multi-threaded CPUs or have redeveloped decade old algorithms to take advantage of memory-based storage. Without the software to properly drive the eco-system, vendors and customers are forced to over-provision resources to drive the storage media which makes taping into the potential of these technologies an even more expensive investment.

USE CASE: CREATING A SIMPLE SINGLE SYSTEM TO MEET ALL STORAGE NEEDS

Storage infrastructures within data centers are becoming increasingly complex as more and more storage silos appear for each IT Stack (VMware, Hyper-V, KVM, Kubernetes). Vendors have tried to solve the storage sprawl problem by delivering scale-out architectures. These scale-out designs, however, don't efficiently use storage resources making them more expensive over time. Scale-out architectures also add a layer of networking and management complexity as they scale. The reality is that, if a scale-up architecture can scale to meet the data center's capacity requirements, then those architectures already meet most organizations' current and future performance demands. Lastly, scale-up architectures are easier for most organizations to manage since there is just one physical system instead of multiple storage nodes. Scaling beyond one system, if needed, can eventually be managed by leveraging tiering or automated volume movement between systems.

The Intel Optane / StorONE design intends to cross the bridge between scale-up and scale-out architectures by enabling data centers to select one physical system that meets all of their capacity and performance needs across a wide variety of IT stacks. The solution combines Intel Optane Memory Class Storage and QLC-based flash to properly balance the customer need for performance and capacity, as well as maintain very appealing economics.

UNDERSTANDING THE INTEL OPTANE - STORONE ARCHITECTURE

The Intel Optane - StorONE architecture use Intel Optane as a large shock absorber to the less durable QLC-Flash tier. The solution does not use Optane as a cache in this configuration, but as a safe, non-volatile tier of storage where data persists for some time, potentially days, if not weeks. The overwhelming majority of modifications to data is only for a short time after creation and can occur mostly on the Optane tier. After the initial period of activity, the data reaches a reference state and often never changes again. At that point, the majority of IO is reads of that data.

The use of Optane as a tier instead of cache not only enables very high-performance read/write operations during this active time, but it also allows data to go through this initial lifecycle to its reference state before the solution moves the data to QLC. QLC can provide more than acceptable read performance. The StorONE software provides the customer with automated movement of data between the Optane and QLC tiers while at the same time delivering the enterprise features, they expect, like snapshots and built-in data protection. It also extracts maximum life out of the QLC tier by protecting it from unnecessary write IO and optimizing the IO operations to be as sequential as possible to increase the life of the QLC drives.

TESTING SUMMARY

In StorONE's testing, detailed below, we were able to prove the platform's ability extract maximum per drive performance using modest server hardware and server RAM. Using only three Intel Optane drives, the S1 Enterprise Storage Platform generated over one million read IOPS and over 400K Write IOPS. During these tests, our erasure coding-based data protection was active, and the software was taking snapshots.

The testing leads StorONE to the creation of S1:Optane, the first Optane Flash Array. The first has the greatest potential for volume sales of Intel Optane, the Optane Flash Array (OFA). This configuration combines three Optane drives with 5 QLC Flash Drives to deliver 40TBs of capacity. Depending on customer requirements both the Optane tier and the QLC tier can be expanded. It is a solution designed for the mainstream data center, providing them with the first viable upgrade to seven-year-old all-flash arrays. For Intel, it means selling Optane in volume. S1, because of its maximum per-drive-performance, and intelligent, QLC aware, auto-tiering makes the solution possible.

The over 2TBs of Optane capacity ensures that the in most cases, all of an organization's write operations will go to the high-performance tier. The "direct write" design of the StorONE S1 software means there is no requirement for a RAM-based cache layer. Requiring a write-cache overlooks the critical advantage of the write-optimized Intel Optane technology and increases design complexity because of RAM's volatility. Requiring a write cache also increases the cost of the overall solution.

Since S1 does not treat Optane as a cache, read operations are in-place. Data does not need to move from QLC to Optane. Because of the number of drives in the QLC tier, the tier generates excellent read performance and does not always require data promotion. S1 will only promote data from the QLC tier to the Optane when it identifies a read performance advantage. This QLC intelligence prolongs the tiers flash endurance and accelerates user response times.

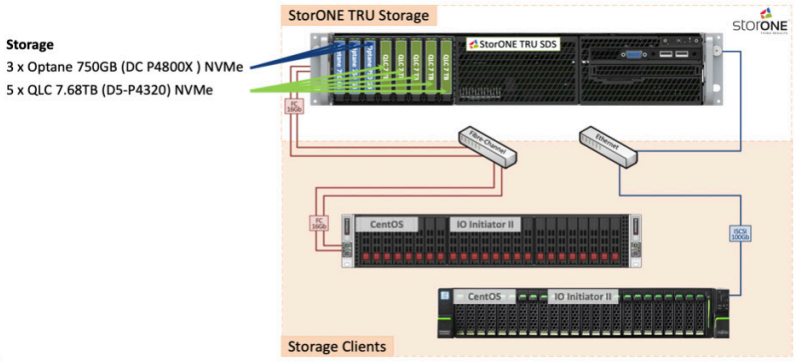
The second solution is the All-Optane Array, which is an Optane only configuration ideal for extreme performance workloads. Although we require further testing, we believe the S1 Software will continue to scale as additional Optane drives added to the configuration. The eventual bottleneck will be the PCIe bus.

In both cases, the S1 Enterprise Storage Platform can open up new windows of opportunity for Intel's Optane technology without forcing complicated and expensive workarounds to hide software inefficiency. Both configurations can start small and scale large, making them appealing to data centers of all sizes. It is Intel's opportunity to sell Intel Optane in volume.

TESTING ENVIRONMENT

The StorONE test environment includes our S1 Enterprise Storage Platform software running on one storage server. The server has two Intel Xeon Platinum 8260L processors, 64GB Memory when testing the All-Optane Array, and 128GB when testing the Optane-Flash Array. The network for the testing uses a 16Gbps Fibre Channel with a QLogic QLA2672 adapter and 100Gbps iSCSI using a Mellanox ConnectX5 adaptor. The media inside the storage server is three NVMe-based Intel Optane 750GB (DC P4800X) and Five NVMe-based 7.68TB (D5-P4320) QLC Drives.

There are two clients involved in the testing. The first is a Supermicro (SYS-2028U-TN24R4T+) Server configured with an Intel Xeon E5-2650 v4, 128GB Memory, and 1 Fibre-Channel Dual-port 16Gb Qlogic QLA2672. The second is a Fujitsu (RX2540) Intel Xeon E5-2680 v3, 128GB Memory, 1 Ethernet Dual-port 100Gb Mellanox ConnectX5.



All-Optane Array Testing

The first step in StorONE's testing was to validate our claim of maximum performance per drive. To confirm this claim, StorONE ran a series of tests on the Optane only configuration, including sequential reads, random reads, random mix. In all tests, data protection remains active. The results reported are client-side numbers. The S1 Server on write tests was writing data and metadata twice for redundancy.

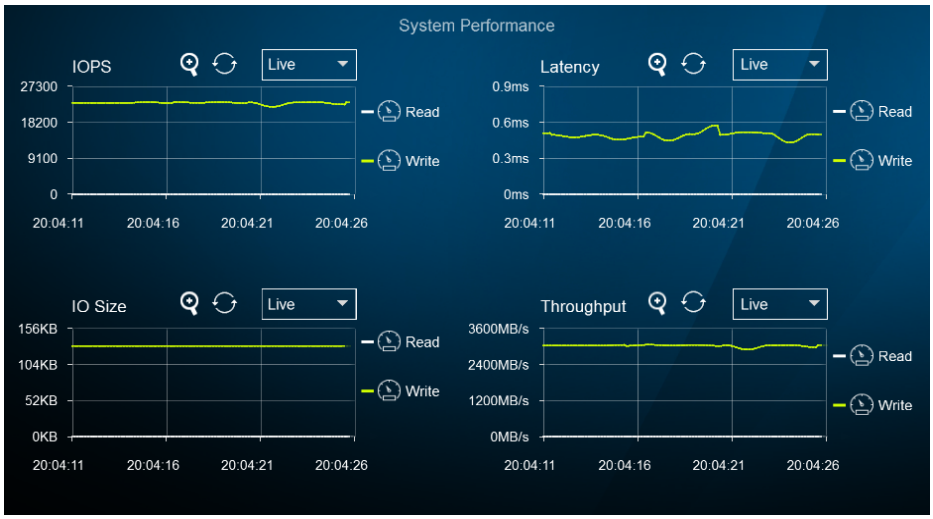
In all cases, the StorONE S1 Enterprise Storage Platform delivered between 85% to 100% of the raw performance of the physical drives.

ALL-OPTANE ARRAY TEST SUMMARY

IO SIZE	IO PATTERN	THROUGHPUT	IOPS	LATENCY MS	QD
128KB	Sequential Write	3,100	MBPs	0.5	4
128KB	Sequential Read	6,800	MBPs	0.3	32
4KB	Sequential Write	540,000	IOPs	0.6	32
4KB	Sequential RE-write	450,000	IOPs	0.5	32
4KB	Random Write	480,000	IOPs	0.55	32
4KB	Random RE-write	420,000	IOPs	0.6	24
4KB	Random Read	1,150,000	IOPs	0.035	64
4KB	Mixed 80/20 Random read/Re-write	850,000 (680,000 / 170,000)	IOPs	0.08 / 0.58	32
128KB	Mixed 80/20 Sequential Read/Re-Write	4,100 / 1,100	IOPs	0.07 / 0.7	32

ALL-OPTANE ARRAY TEST DETAILS

TEST 1: 3 Optane drives: sequential write 128k



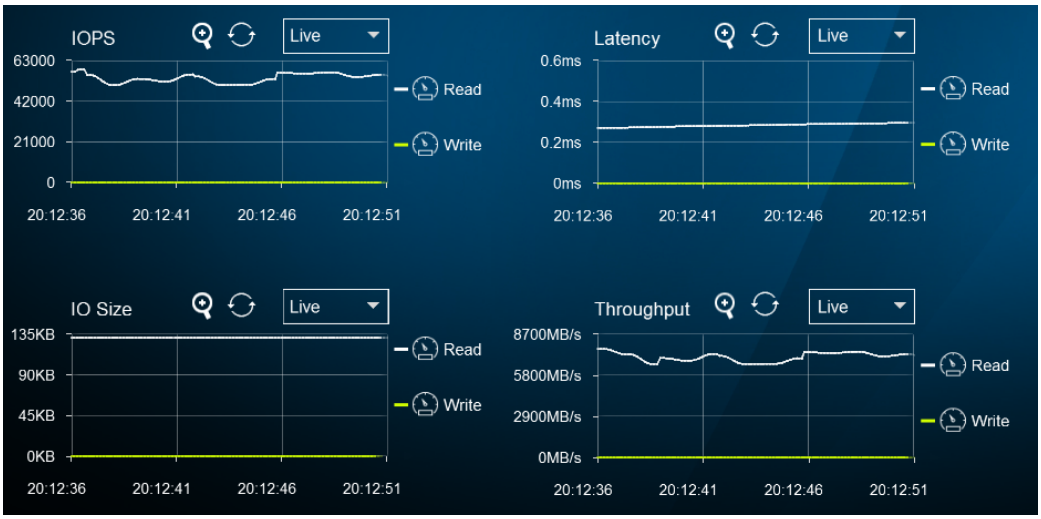
CUSTOMER

- 10 volumes
- 1 drive redundancy
- Throughput: 3.1GB/s
- Latency 0.5ms

StorONE (Background)

- Data is written twice (for redundancy)
- Metadata written twice (for redundancy)
- Drives real throughput 6.5GB/s

TEST 2: 3 Optane drives: sequential read 128k



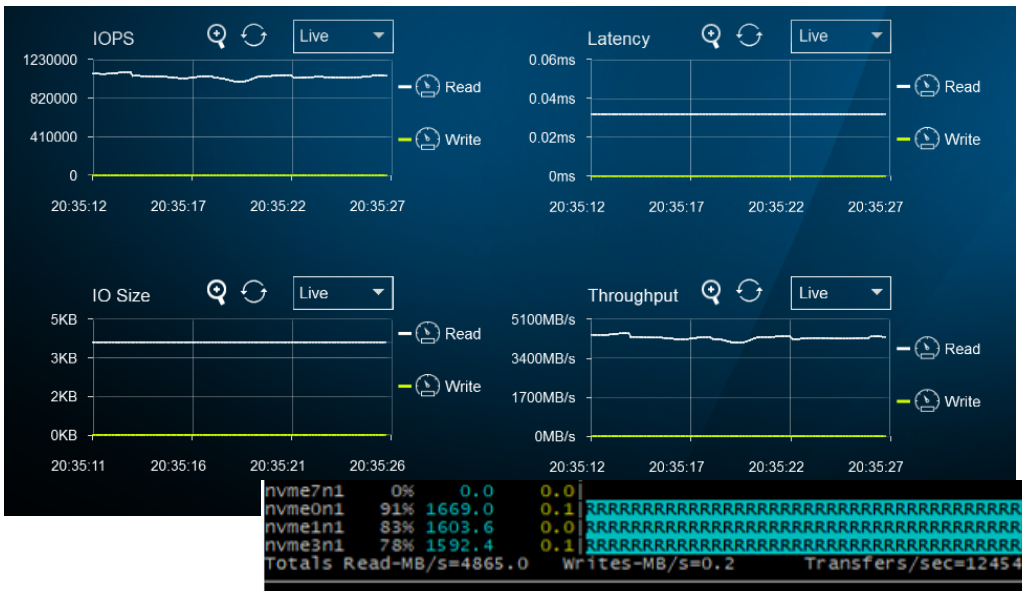
CUSTOMER

- 1 drive redundancy
- Throughput: 7.1GB/s
- Latency 0.3ms

StorONE (Background)

- Data is read from all drives
- Metadata is read from all drives
- Drives real throughput 7.2GB/s

TEST 3: 3 Optane drives: random read 4K



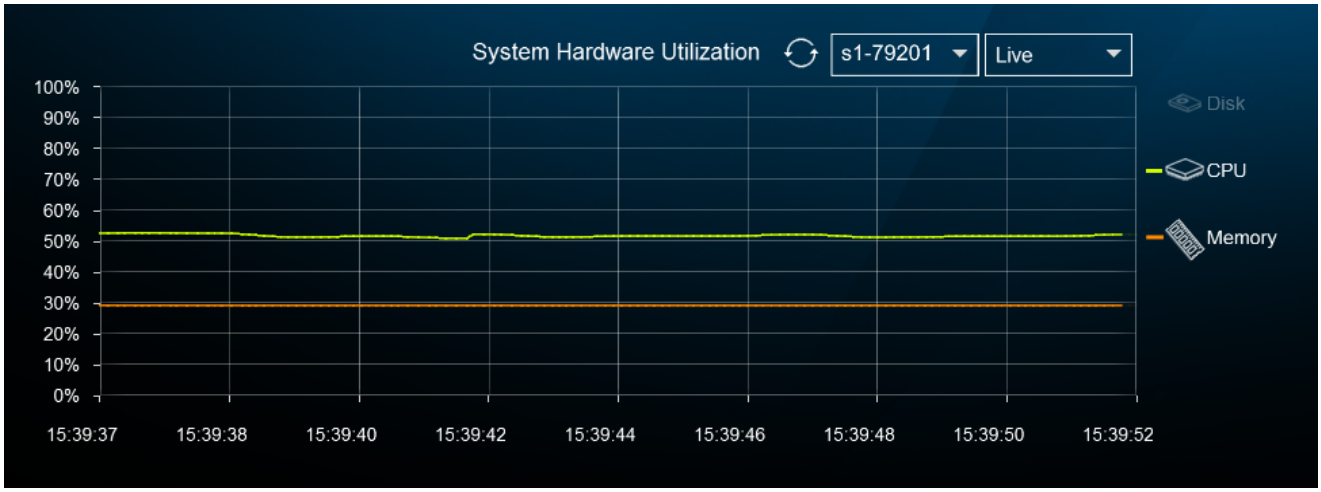
CUSTOMER

- 1 drive redundancy
- IOPS 1,150,000
- Throughput: 4.3GB/s
- Latency 0.3ms

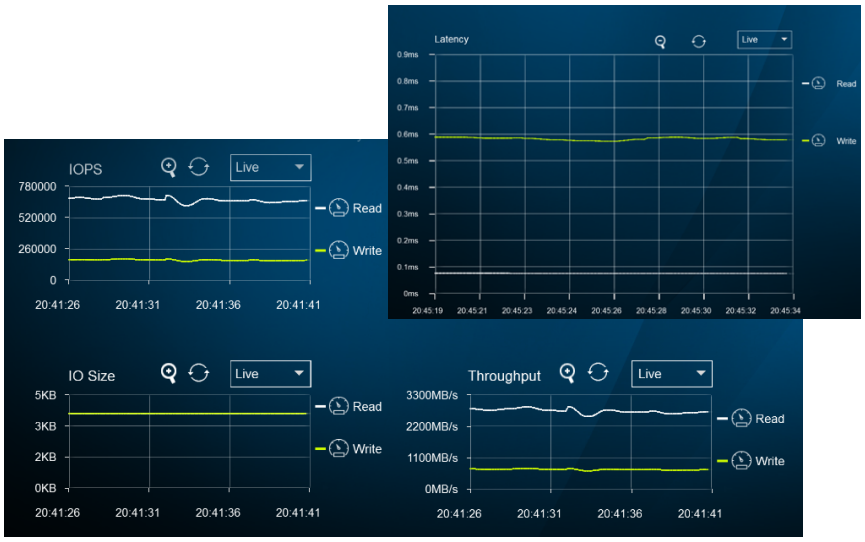
StorONE (Background)

- Data is read from all drives
- Metadata is read from all drives
- Drives real throughput 4.865GB/s

TEST 4: 3 Optane drives: Random Read 4K – Low Hardware Utilization



TEST 5: 3 Optane drives: 4K Random 80% Read 20% RE-Write



CUSTOMER

- 1 drive redundancy
- Read IOPS: 680K @ 0.08ms
- Write IOPS: 170K @ 0.58ms

StorONE (Background)

- Data is read from all drives
- Metadata is read from all drives
- Data is written twice (for redundancy)
- Drives real IOPS 1,280,000

TEST 6: 3 Optane drives: 128K Sequential 80% Read 20% RE-Write

CUSTOMER

- 1 drive redundancy
- Read TP: 4.1GB/s @ 0.4ms
- Write TP: 1.1GB/s @ 0.9ms

StorONE (Background)

READ

- Data is read from all drives
- Metadata is read from all drives

WRITE

- Data is written twice (for redundancy)
- Metadata is written twice (for redundancy)
- Drives real TP 6.5GB/s

OPTANE FLASH ARRAY TESTING

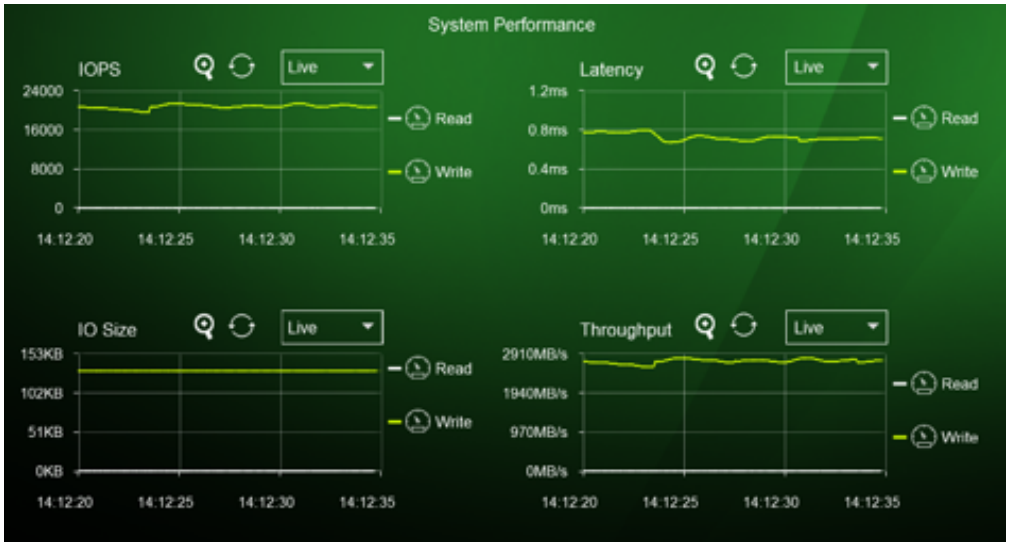
The Optane Flash Array testing uses the same physical configuration for testing. The auto-tiering tests use ten volumes. Each volume consists of a 25GB upper tier (Optane) and a 100GB lower tier (QLC).

The first step in the Optane Flash Array testing is to fill the Optane tier with data, forcing the S1 software to move older data to the QLC tier. It is challenging to simulate real-world data aging in a lab environment, so the results below, while impressive, represent a worst-case scenario for the configuration. The tests show a continuous write pattern that results in constant tiering activity. In real-world use, tiering occurs during off-hours when the configuration is not busy. It is reasonable to assume that during the workday, the customer will experience full Optane performance for all read and write activity, with no tiering overhead.

OFA TEST SUMMARY

10 VOLUMES 1+1	AUTO-TIERING
Random Write 4k	158K IOPs @ 1.1ms *
Random Read 4k	1,200K IOPS @ 0.4ms
Sequential Write 128K	2,800 MB/s
Sequential Read 128K	10,000 MB/s

OFA TEST PHASE 1: WRITE TO UPPER TIER



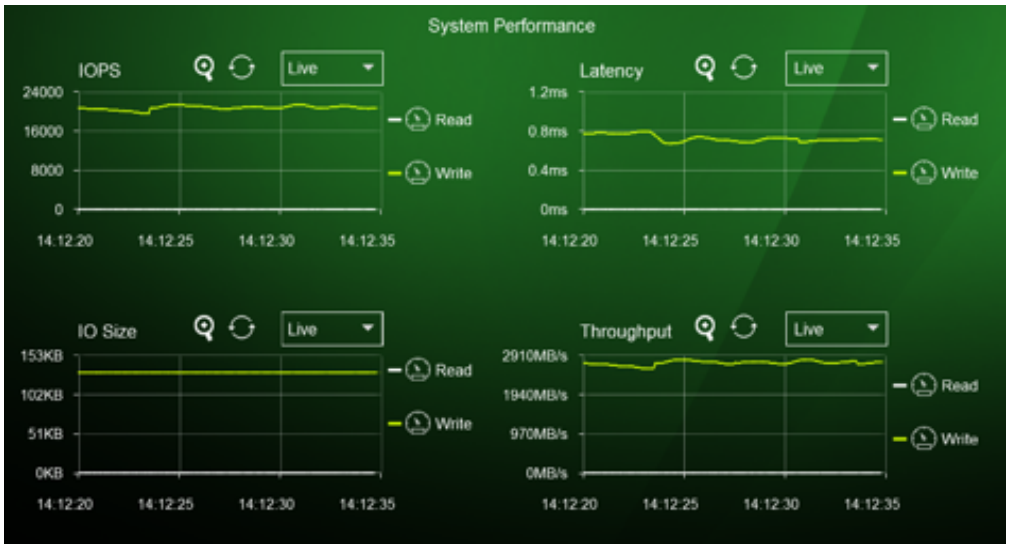
CUSTOMER

- 1 drive redundancy
- Read TP: 2.8GB/s
- Latency .07ms

StorONE (Background)

- Data is written twice (for redundancy)
- Metadata is written twice (for redundancy)
- Drives real throughput 6.5GB/s

OFA TEST PHASE 2: PERFORMANCE WHILE EVACUATING DATA TO LOWER TIER



CUSTOMER

- 1 drive redundancy
- Read TP: 2.2GB/s
- Latency 4.8ms

StorONE (Background)

- Data is written twice (for redundancy)
- Metadata is written twice (for redundancy)
- Data is being evacuated from upper to lower tier
- Drives real throughput 8.5GB/s

CONCLUSION

In real-world use, the S1 powered OFA enables the customer to experience all the benefits of Optane at a price lower similar to an All-Flash Array. It also provides optimizations for and protection of the less durable QLC technology. The time of the All-Flash Array has come to an end. The future is the Optane-Flash Array.


storONE™
THINK RESULTS™