KIOXIA

Top 5 Reasons

Optimize Data Center Server Architectures with EDSFF E1.S SSDs

The main building block of cloud computing is the hyperscale data center - an infrastructure built on a scale many orders of magnitude greater than a standard data center. They enable cloud providers to deliver extremely efficient, scalable and cost-effective platforms that meet today's big data needs, but differ from traditional data centers in several ways. One key differential is how they carefully manage storage drives to meet user requirements, reduce customer storage acquisition costs and enable customer scalability on an as needed basis. With their vast resources, hyperscalers develop custom server designs enabling them to quickly transition and take advantage of the latest storage technologies. The benefits afforded by hyperscale data centers are very compelling to customers.

While the hyperscale data center value proposition is quite compelling, many of its benefits are limited to only a small number of cloud providers able to operate at this hyperscale level. For the larger demographic of standard data centers who would like to emulate the hyperscale value proposition, the Open Compute Project (OCP) is enabling similar benefits and economies of scale as the largest hyperscalers. OCP takes an open source, industry collaborative approach to the data center landscape, addressing such storage challenges as drive performance and latency, power loss protection[1], drive serviceability, signal integrity, power envelop delivery, thermal management, security, and Quality of Service (QoS), to name a few.

One of the primary areas of OCP focus is on storage device form factors that, in many cases, could address the storage challenges mentioned above within a data center landscape. After their evaluation, OCP standardized on the Enterprise and Datacenter Standard Form Factor (EDSFF) E1.S and E3 form factors developed by the Small Form Factor Technical Affiliate (SFF-TA) working group as part of the Storage Networking Industry Association (SNIA).

EDSFF form factors enable strong throughput, input/output operations per second (IOPS) and latency performance with the PCIe® interface and the NVMe® specification. From an interface perspective, PCIe 4.0 provides a much larger pipe for data to move through versus the SATA interface. While many infrastructures still utilize SATA-based storage, hyperscalers have either deployed or plan to deploy NVMe/PCIe 4.0 storage.

KIOXIA Corporation provides next-generation PCIe 4.0 data center NVMe SSDs with its XD6 Series, which are designed to the OCP NVMe Cloud SSD specification and available in E1.S form factors. There are a number of key reasons why these new EDSFF E1.S SSDs can enable a data center server architecture to adapt to changes, but the top five supported by KIOXIA include:

1. **Higher Performance / Higher Power Budget vs M.2 Devices**
   More than doubles the power budget versus M.2 devices, enabling E1.S SSDs to saturate PCIe Gen4 performance
2. **Standardized Thermal Solutions:**
   Improves interoperability across vendors and platforms while providing the flexibility to select the right balance of cooling and storage density through different E1.S heatsink options
3. **Improved Physical Serviceability**
   Vastly improves serviceability with hot plug support that no longer requires an entire server to be taken down in order to replace a single SSD
4. **Designed to Better Accommodate NAND Flash Memory Packages**
   Wider PCB design enables optimized orientation of the NAND flash memory packages, providing headroom for higher capacity drives
5. **Supported by Leading Hyperscalers**
   Facebook™ and Microsoft®, leading authors of the OCP NVMe Cloud SSD specification, are using E1.S designs on new and upcoming platforms that has resulted in industry-wide support and adoption

1. Higher Performance / Higher Power Budget vs M.2 Devices

KIOXIA XD6 Series PCIe 4.0 data center NVMe SSDs are representative of a new category of OCP cloud-optimized SSD designs and are based on the EDSFF E1.S form factor. The XD6 Series product specifications[2] include:

XD6 Series NVMe/PCIe 4.0 Data Center SSD Specifications:

Form Factor	E1.S (9.5/15/25mm)	
Flash Memory Type	BiCS FLASH™ 3D flash memory	
Interface Specification	PCIe 4.0 x 4L, NVMe 1.4	
User Capacities *(in gigabytes[3])*	1,920 GB	3,840 GB
Performance		
Sequential Read: 128 kibibytes[4] (KiB), Queue Depth (QD) =32	6,500 MB/s*	6,500 MB/s
Sequential Write: 128 KiB, QD =32	1,200 MB/s	2,350 MB/s
Random Read: 4 KiB, QD =256	660,000 IOPS	880,000 IOPS
Random Write: 4 KiB, QD =128	50,000 IOPS	90, 000 IOPS
Power		
Supply Voltage	12V	
Active Power Consumption	<14.0 watts	
Endurance *(per 5 years)*	1 Drive Write Per Day[5] (DWPD)	
MTTF Reliability	2.0 Million Power-On Hours (MPOH)	
Operating Temperature	0 to 70° C	
Security Option(s)	Self-Encrypting Drive: TCG-Opal 2.0	

**MB/s = megabytes per second;*

Utilizing a purpose-built controller with the latest PCIe 4.0 interface and 3,840 GB capacity, the KIOXIA XD6 Series offers 4x greater improvement in random write performance, 3.5x greater improvement in random read performance, 2.5x greater improvement in sequential write performance and 2x greater improvement in sequential read performance over the previous generation (XD5 Series M.2 data center SSDs). The sequential read bandwidth nearly saturates the PCIe 4.0 bus, transferring data at a speed up to 6,500 MB/s (Figure 1).

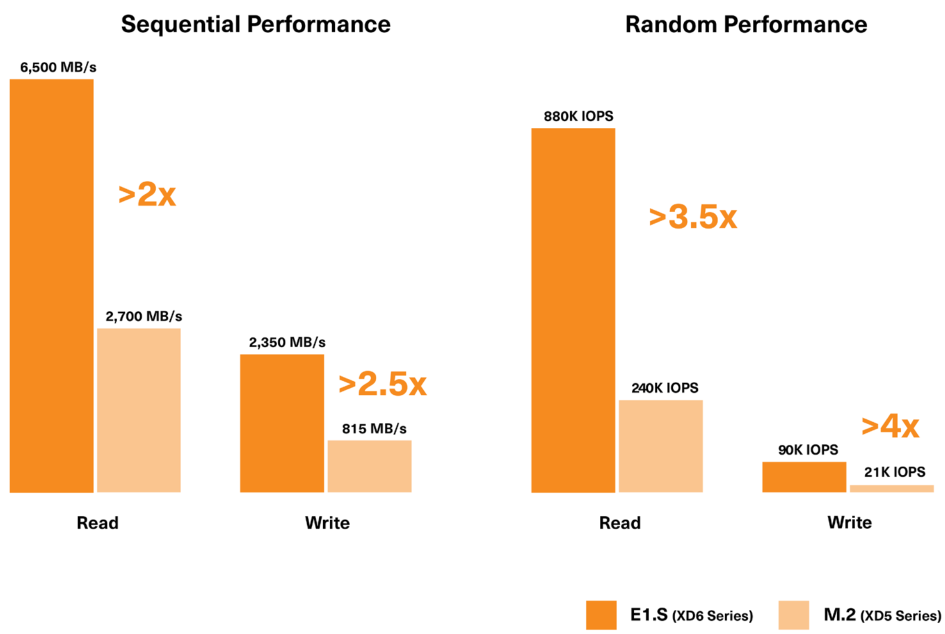

Figure 1: E1.S SSD (XD6 Series) vs M.2 SSD (XD5 Series) performance comparisons

In addition to raw performance, power efficiency is a key metric for data center administrators who care about total cost of ownership (TCO). The XD6 Series offers PCIe Gen4x4 bandwidth saturation performance while maintaining a low 12-watt active power draw during these sequential read operations. Additional OCP power states are supported enabling a host server to set the right balance of power to performance for each application.

Performance on a basic M.2 (22x110) steady state SSD is typically limited at 8.25W, especially with higher capacities such as 4 terabytes[3] (TB) and beyond. This form factor cannot achieve full Gen4x4 bandwidth with an 8.25W power envelope. On the other hand, depending on the heat sink, capacity and performance requirements, an E1.S SSD can achieve full Gen4x4 bandwidth and even Gen5x4 bandwidth with support for up to a 25W power envelope (depending on drive capacity).

2. Standardized Thermal Solutions

OCP provides the ecosystem with a standardized, open source specification for data center hardware, enabling similar thermal benefits and economies of scale used by the world's largest hyperscalers.

E1.S SSD Type	Width	Length	Thickness
E1.S (9.5mm)	33.75mm	118.75mm	9.5mm
E1.S (15mm)	33.75mm	118.75mm	15mm
E1.S (25mm)	33.75mm	118.75mm	25mm

The KIOXIA E1.S SSDs[6] include:

XD6 Series – 9.5mm[7]

XD6 Series – 15mm

XD6 Series – 25mm

For OCP cloud-specified SSD designs based on the E1.S form factor, the XD6 Series supports a number of thermal management features that monitor drive health and data integrity. It is available in 9.5mm, 15mm and 25mm heatsink sizes as defined in the EDSFF E1.S standard, while its performance profile and power envelope are identical across all XD6 Series variants. The heatsink options provide data center architects with the ability to select the thermal profile that is best suited for cooling their systems.

Additionally, a thermal throttling mechanism has been implemented to protect XD6 Series drives if they detect operating temperatures beyond the specified operating range (0 to 70°C). The thermal throttling is engaged and a '**WCTEMP**' Asynchronous Event Notification (AEN) is issued to the host when the operating range is reached. If the temperature continues to rise, as detected by thermal sensors and reported via the Self-Monitoring, Analysis and Reporting Technology (SMART) health monitoring system, the XD6 Series SSD will throttle its performance, reducing heat that was generated by NAND flash memory die activation, with the goal of reducing the temperature. If the temperature continues to rise in the event of a system fan failure, accidental fire or catastrophic hot conditions, a CCTEMP AEN is issued to the host and the XD6 Series SSD will perform a shutdown to preserve the data that was written to flash memory. After the thermal shutdown has been issued, the SSD will remain in power off mode until power is turned back on and the system cycles.

3. Improved Physical Serviceability

The EDSFF E1.S specification allows for hot swap (also known as hot plug support) which is featured within XD6 Series SSDs. E1.S hot plug support directly resolves physical serviceability concerns associated with previous M.2 products. To replace an M.2 22110 SSD, normally an entire server would need to be powered down in order to replace a single drive - a process that is time-consuming and potentially challenging to work around without impacting serviceability. Once powered down, the process requires removing the system from the rack, pulling out the AVA-4 M.2 carrier card, removing the heatsink and M.2 SSD, putting in a new M.2 SSD and thermal interface material and redoing the entire assembly to get the server back in operation.

With E1.S SSDs, the drive swap is simple and can occur on the fly. As data center administrators manage an overabundance of systems, the cliché that ‘time is money’ holds true. The ability to improve physical serviceability is critical for saving time and reducing total cost of ownership (TCO).

4. Designed to Better Accommodate NAND Flash Memory Packages

M.2 SSD packaging includes a 22mm wide printed circuit board (PCB) that constrains NAND flash memory placements and limits higher drive capacities. As defined by the EDSFF E1.S specification, the PCB width dimensions have increased to 33.75mm. This wider width provides more space on the PCB for additional flash memory chips, which in turn enables higher capacity SSDs and provides more capacity per allowable space. When compared to one of the highest capacity M.2 SSDs available[8] at 3.84 TB, E1.S SSDs can support up to 15.36 TB, or 4x that of M.2 SSD capacity.

Supported Capacities: M.2 SSDs vs E1.S SSDs

M.2: 960 GB, 1.92 TB, 3.84 TB

E1.S: 960 GB, 1.92 TB, 3.84 TB, 7.68 TB, 15.36 TB

5. Supported by Leading Hyperscalers

Leading authors of the OCP NVMe Cloud SSD specification, Facebook and Microsoft are using E1.S designs on new and upcoming platforms with scalability and versatility relating to capacity, thermal considerations, performance, hot-plug capabilities, etc. These large volume deployments from these key hyperscalers are receiving industry-wide adoption that positions E1.S as a preferred form factor for future storage projects, particularly 1U deployments.

Facebook’s first experimental OCP-designed data center in Prineville, Oregon reportedly resulted in ‘a building that is 38% more efficient and 24% less expensive to build and run[9] than other state of the art data centers.’ Other data center companies took note, and today OCP has over 195 members, including the world’s largest hyperscalers.

Jason Adrian, Senior Director of Azure® Platform Architecture, Microsoft and OCP Storage Chair

“Microsoft and the OCP storage workgroup demonstrated how an open collaboration across the industry could align hyperscalers, system designers and SSD vendors around next-generation storage form factors. The EDSFF E1.S form factor is the future of flash storage in hyperscale data centers, including Azure platforms. SSDs designed to the OCP NVMe Cloud SSD specification, such as the KIOXIA XD6 Series data center SSDs, will power the next-generation of EDSFF E1.S based servers.”

Additional XD6 Series SSD information is available here.

Notes:

[1] Power Loss Protection (PLP) support helps to record data that resides in buffer memory to NAND flash memory by utilizing the backup power of the solid capacitor in case of a sudden power supply shut down.

[2] Product testing was conducted in a lab environment by KIOXIA Corporation. Tested content are believed to be current and accurate as of the date that the document was published, but is subject to change without prior notice. Read/write sequential and random performance results may vary depending on the host device, read and write conditions, and file size.

[3] Definition of capacity - KIOXIA Corporation defines a megabyte (MB) as 1,000,000 bytes, a gigabyte (GB) as 1,000,000,000 bytes and a terabyte (TB) as 1,000,000,000,000 bytes. A computer operating system, however, reports storage capacity using powers of 2 for the definition of 1Gbit = 2^{30} bits = 1,073,741,824 bits, 1GB = 2^{30} bytes = 1,073,741,824 bytes and 1TB = 2^{40} bytes = 1,099,511,627,776 bytes and therefore shows less storage capacity. Available storage capacity (including examples of various media files) will vary based on file size, formatting, settings, software and operating system, and/or pre-installed software applications, or media content. Actual formatted capacity may vary.

[4] KiB: A kibibyte (KiB) means 2^{10}, or 1,024 bytes, a mebibyte (MiB) means 2^{20}, or 1,048,576 bytes, and a gibibyte (GiB) means 2^{30}, or 1,073,741,824 bytes.

[5] Drive Write(s) per Day: One full drive write per day means the drive can be written and re-written to full capacity once a day, every day, under the specified workload for the specified lifetime. Actual results may vary due to system configuration, usage, and other factors.

[6] Each XD6 Series E1.S SSD product image may represent a design model.

[7] The ‘millimeter (mm)’ size for each XD6 Series E1.S SSD indicates the form factor of the SSD and not its physical size.

[8] Based on the capacity of a leading M.2 SSD with information publicly available at the time of this publication.

[9] Source: Architect, Facebook Prineville Data Center, December 20, 2012, https://www.architectmagazine.com/project-gallery/facebook-prineville-data-center.

Trademarks:

Facebook is a trademark of Facebook Inc. Microsoft and Azure are registered trademarks of Microsoft Corporation in the United States and/or other countries. NVMe is a registered trademarks of NVM Express, Inc. PCIe is a registered trademark of PCI-SIG. All other trademarks or registered trademarks are the property of their respective owners. All other company names, product names and service names may be trademarks or registered trademarks of their respective companies

Disclaimers: