WHITE PAPER

Xsight Labs E1: A Storage Acceleration Performer for the Cloud and Data Center

Growing data demands and the increase in data volumes and throughput, drive storage providers to embrace scalable, flexible, secure and future-proof infrastructure for cloud and enterprise deployments.

JULY 2025 | BY XSIGHT LABS, LTD.

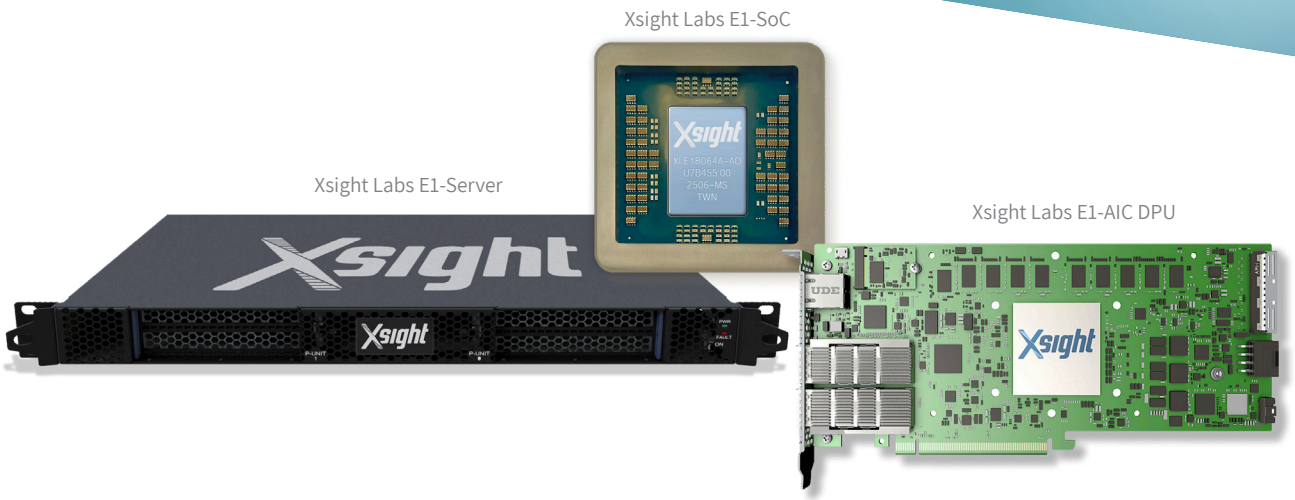


Figure 1 - Xsight Labs E1-SoC Infrastructure Processor Powering DPUs and Servers

In today's landscape, especially in enterprise, hyperscale, cloud environments, or at the edge, storage is no longer a backend concern.

Rather, it is a performance bottleneck, a security risk, and a scalability limiter. Data volumes surge in size and workloads become more distributed.

Traditional architectures can't keep up, and CPU overhead from storage protocols (e.g., TCP/IP, NVMe-oF) drains compute resources.

Across the industry, from cloud providers and hyper-converged infrastructure vendors to high-performance storage players, the consensus grows. Storage processing should run decoupled from main processing to enable better control, better security, improved workload density, and more efficient resource utilization.

Data processing unit (DPU) offloading has been utilized to provide performance benefits, and to ensure isolation. More often than not, however it came at the cost of flexibility and upgradability. Future ability to support emerging standards often meant forklift upgrade.

Xsight Labs is now breaking this equation.

This paper explores in detail how the recently announced Xsight Labs E1 infrastructure processor – developed and designed from the bottom up – faces these challenges, addressing them in a scalable, future-proof and power efficient manner.

The Xsight Labs E1 infrastructure processor arrives as a System-on-a-Chip (E1-SoC), and spawns two lines of products. One, is the Xsight Labs E1 Add-In Card (E1-AIC) DPU, bringing new and unmatched levels of elasticity and programmability to infrastructure design. It ensures performance at scale, future-proof support for evolving standards, isolation and security.

The second is the Xsight Labs E1-Server, leveraging the Xsight Labs E1-SoC as the main server host CPU and supporting storage target and edge deployments.

The paper continues to discuss a range of storage use cases, including storage initiators and storage targets in a storage area network, and concludes with E1 performance benchmarking data.

The Storage Challenge - Perform, Be Scalable, Yet Remain Agile

As data volumes continue to grow, storage solutions providers face increasing pressure to deliver systems that are scalable, performant, and secure. Data Processing Units (DPUs) are critical enablers to meet these demands, in the context of storage initiators in particular.

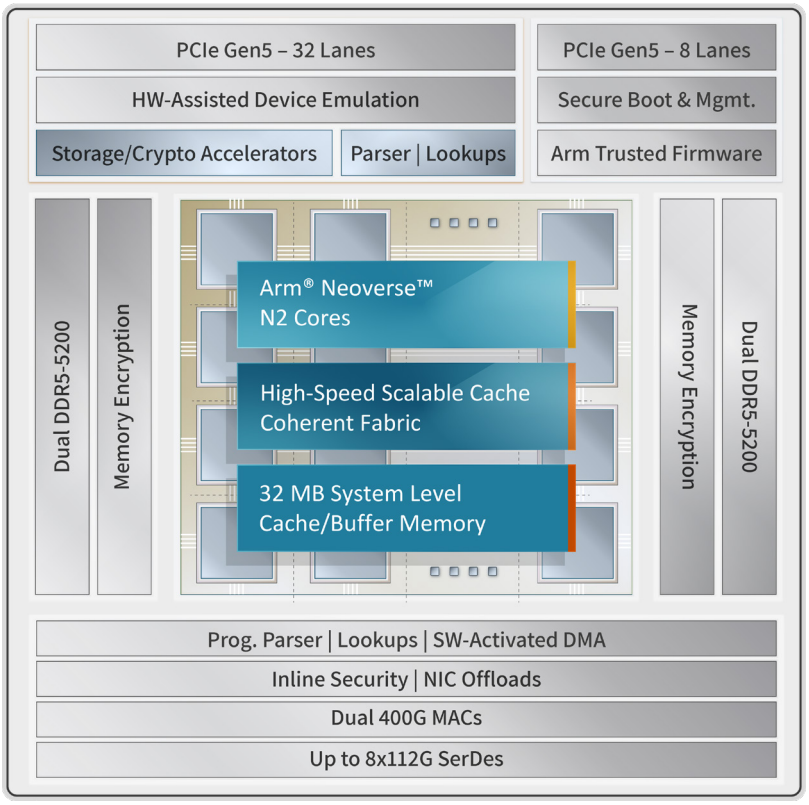


Figure 2 - Xsight Labs E1-SoC Architecture

The following challenges depict the reason why offloading infrastructure tasks from the host CPU to a DPU is no longer optional, but rather a strategic necessity.

- **Resource utilization –** Storage I/O operations on the initiator side, especially with NVMe-oF, TCP, or iSCSI, intensively consume a significant portion of CPU cycles. In high-throughput environments, such as HPC or data lakes, and large-scale analytics platforms, dedicated CPU cores are often required to manage networking stacks (e.g., RDMA, TCP/IP), encryption, compression or storage protocols. Not only does it limit scalability, it also reduces efficiency of compute resources intended for application processing.

- **Performance –** In hyperconverged and cloud-native architectures, mixing compute and infrastructure tasks can result in unpredictable performance. Unless isolated, storage traffic competes with application workloads for CPU time and memory bandwidth, increasing latency and degrading overall system responsiveness.

- **Scalability –** Tier 1 and tier 2 cloud providers must scale storage to support thousands of tenants while ensuring performance and isolation. Unless workloads are offloaded, CPU bottlenecks lead to noisy-neighbor problems, potentially violating SLA. Tenant-level isolation, and secure access require even more CPU cycles.

- **Security –** Modern storage systems must enforce strict access control, encryption, and data integrity on data at rest and in transit. When implemented on host CPU, attack surfaces widen and increase the risk of lateral movement in compromised environments. Offloading these tasks to DPUs enhances security by isolating control and data planes. System's vulnerability footprint is reduced.

- **Edge and micro data centers –** Form factor and physical space are limited in edge data center installations. Each additional server, installed to accommodate infrastructure tasks, puts pressures on cost and complexity. DPUs enable higher functionality within compact footprints, allowing more capability per rack unit.

- **Power efficiency –** Power consumption and thermal management remain critical concerns for large-scale data centers. Metrics such as power-per-IOPS or power-per-Gbps efficiency are key benchmarks for modern storage systems.

- **AI factory workloads –** AI introduces new challenges for storage. Training and inference workflows require ultra-low latency and extremely high IOPS, particularly in networks leveraging NVMe and RoCEv2. AI storage pipelines are multi-tiered, spanning across a chain of different technologies (hot SSD, cold HDD, archive). These workflows demand compliance with auditing and data anonymization requirements. Managing such complexity within host CPUs is inefficient and unsustainable.

Meeting the demands of modern storage infrastructure, across performance, scalability, security, and power efficiency, requires a strategic architectural shift. DPUs offer a compelling solution by offloading critical tasks from the host CPU, enabling infrastructure that is both agile and sustainable.

We shall expore how Xsight Labs E1 architecture addresses these challenges and redefines what's possible in storage system design.

E1 Architecture – Software Defined Storage Solution, in Hardware

Unlike competing DPUs, the Xsight Labs E1-AIC DPU is a fully programmable, high-performance 2 x 400GbE DPU solution, sustaining bidirectional 400GbE line-rate throughput. Delivered as an add-in card, it integrates seamlessly using PCIe Gen5 x16 host interface to modern server nodes, delivering advanced storage initiator functionality.

Supporting a choice of industry standard host interfaces and networking stacks, including NVMe and virtio-blk, the Xsight Labs E1-AIC DPU adapts to varying deployment models, from locally attached storage, to remote network or shared storge; with a standard storage stack running on host. Legacy iSCSI is also fully operable over the Xsight Labs E1-AIC DPU, utilizing its network host interface (virtio-net or XNA).

Alternatively, the Xsight Labs E1-Server working as a storage target, brings additional value with its programmability, lookaside offload (e.g., encryption) sub-engines, and flexible architecture.

Straight Forward and Standard

At its core, the Xsight Labs E1-SoC was designed with open standards in mind, as evidenced by its standard programming model, its host interfaces (relevant for E1-AIC DPU), and system integration methods. This ensures compatibility, ease of integration, and long-term sustainability in modern data centers.

- **Programming Model –** The Xsight Labs E1 architecture supports standards-based software defined development model, enabling developers to use familiar tools and frameworks without proprietary lock-in.
- **Host interface –** Over PCIe host interface, the Xsight Labs E1-AIC DPU delivers line-rate NVMe or virtio-blk performance. It can support locally attached storage (NVMe SSD optionally available on E1-AIC DPU) as well as remote storage access over the network (RoCEv2, or TCP) or both simultaneously, if desired.
- **Storage device interface –** The Xsight Labs E1-AIC DPU features industry-standard storage interfaces, including M.2 / U.2 for local NVMe SSDs, and 100GbE, 200GbE and 400GbE for high-speed network interfaces for storage over the network.

Performance without compromise

Featuring clear-cut, no frills internal data path with an array of high-performance processing cores (for example, no exception jumping to secondary data processing plane, the way other DPUs operate), the Xsight Labs E1 architecture provides the benefit of measurable, scalable, and most importantly, predictable storage performance.

- **Line rate –** Storage IOPS, throughput, and latency, from the perspective of the compute or hypervisor node, are consistent and maintained at high-performance levels, thanks to the Xsight Labs E1-AIC DPU's full line-rate support. Internally, the Xsight Labs E1 processing subsystem reserves sufficient compute capacity to reliably handle 2 × 400GbE traffic, sustaining performance even under intensive data and storage workloads.

- **Accelerators sub-system –** To ensure sustained line rate without compromising on storage feature set, hardware sub-engines, including AES-XTS and CRC64 accelerators are available internally on the Xsight Labs E1-SoC to be seamlessly engaged, to guarantee privacy, integrity and SLA.

- **Scaling factors –** The Xsight Labs E1 infrastructure processor architecture is fundamentally designed for scale, leveraging high core density to handle complex and parallelized workloads efficiently, enabling per-tenant workload isolation and horizontal expansion across data center topologies.

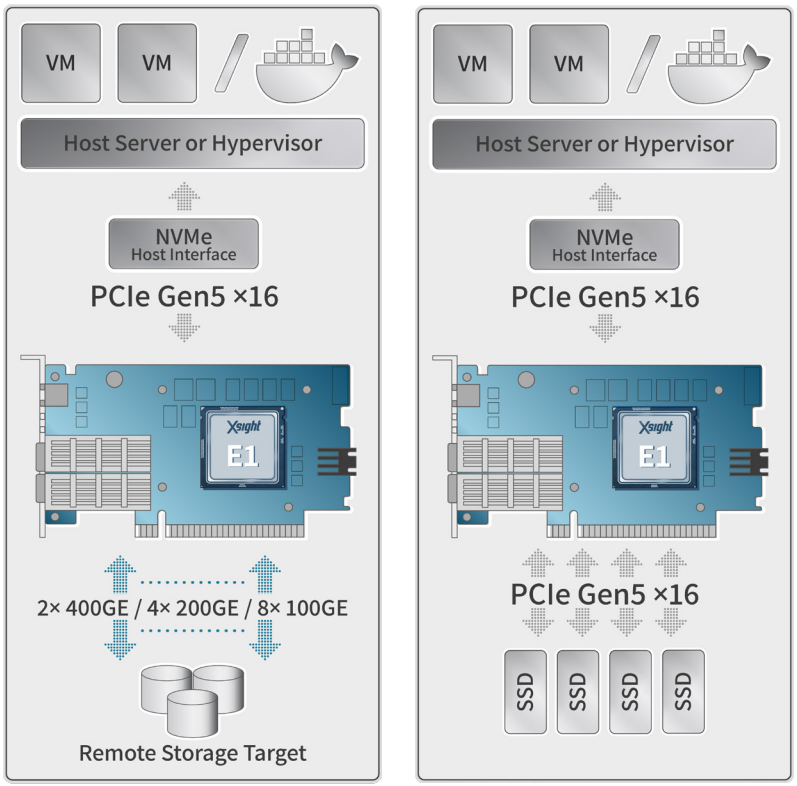


Figure 3 (left) - The Xsight Labs E1-AIC DPU Server Front-End Networking Layout
Figure 4 (right) - Locally Attached Storage Layout using M.2/U.2 NVMe SSD

Long-term scalability is warranted by elastic software programmability based on open standards, enabling organizations to future-proof their infrastructure investments (scaling for AI workloads, cloud-native services, or storage-intensive applications without architectural overhauls).

Isolated and Secure

The isolation of infrastructure processing workloads offered by the Xsight Labs E1-AIC DPU, warrant predictability and control.

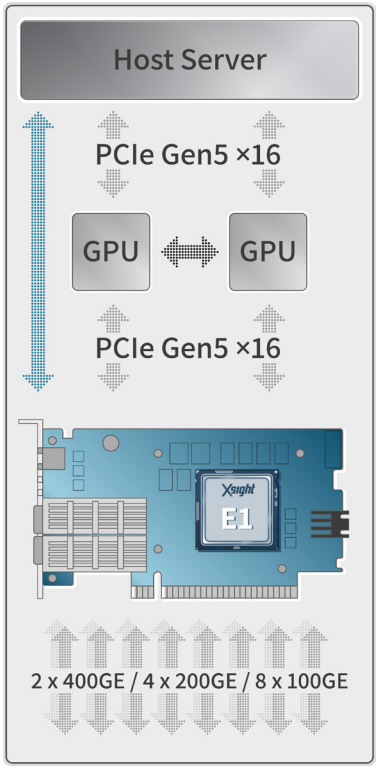
When storage and networking infrastructure tasks are performed separate from main CPU, with resources separately allocated, high IOPS rate comes not at the expense of core processing. This guarantees predictable performance and predictable and stable latency.

Siloed infrastructure processing, as done within the Xsight Lab E1-AIC DPU provides stronger security posture and mitigates the risk of lateral movement in case of a breach in a multi-tenant environment. Attack surfaces are minimized to failures on the storage or networking stack on the DPU, away from business workloads on the main host CPU, improving overall system resilience.

Powerful, Yet Power Saver

The Xsight Labs E1-AIC DPU delivers significantly higher power efficiency, measured in IOPS per watt, compared to general purpose x86 CPU handling similar workloads, especially on high NVMe-oF throughput scenarios.

By delegating infrastructure workload to Xsight Labs E1-AIC DPU, main host CPU remains focused on revenue generating user logic workload, improving overall power efficiency and thermal load management, across the data center.



Future Proof

The flexibility and elasticity of Xsight Labs E1 infrastructure processor standard-based architecture positions it as one of the most future ready DPUs available and ensures it remains relevant in a fast-evolving infrastructure landscape.

AI workloads intensify and bring unique demands on networking and storage, to which the Xsight Labs E1-AIC DPU is built to adapt. AI-generated data spans multiple lifecycle stages, from ingestion and training artifacts to RAG vectors and audit logs, requiring seamless tiered storage management. With its programmable pipeline, the Xsight Labs E1-AIC as a drop-in DPU can intelligently manage automated tiering across SSD, HDD, and object archive storage, without interrupting active workloads.

The Xsight Labs E1-AIC DPU delivers lasting value, ready for what comes next.

Figure 5 - Xsight Labs E1-AIC DPU in an AI Cluster

Architectural Flexibility, From Inline Accelerator to Storage Brain

As storage architectures evolve to support increasingly diverse workloads, the Xsight Labs E1-SoC supporting Xsight Labs E1-AIC DPU or Xsight Labs E1-Server stands out as a highly adaptable building block. Its key strength lies in its flexible deployment options. It integrates in multiple configurations, paired with compute server, embedded in a storage node, or deployed as a standalone appliance in the data path. Each setup offers unique advantages in performance, isolation, offload efficiency, and overall system scalability, allowing architects to tailor infrastructure design without compromise.

Server Add-On, DPU Configuration

In a NIC-based configuration, the Xsight Labs E1-AIC DPU is deployed as an intelligent add-on card in the server. It exposes standard network and storage interfaces (e.g., NVMe, virtio-blk) to the host, enabling applications to interact with the DPU as if it were a local storage. The DPU handles read/write requests by accessing remote storage targets over RDMA/RoCEv2 or TCP, or by interfacing with locally attached SSDs on the DPU itself.

This architecture is host-transparent. It means that no changes are needed to host applications or the OS. Standard Linux storage stacks and drivers are fully supported, and orchestration frameworks (e.g., Kubernetes, OpenStack) can manage the storage interfaces natively. Pulling the infrastructure processing away from the host enhances infrastructure efficiency, lowers CPU contention, and simplifies multi-tenant performance management.

On the DPU side, the Xsight Labs E1-SOC runs standard target-side software such as SPDK.

In DPU configurations, Xsight Labs E1 infrastructure processor supports:

- **Headless storage** (e.g., JBOF/JBOD), where the DPU handles volume management and enforces access control
- **Managed storage heads,** where storage logic resides remotely
- **Local storage,** accessed directly via U.2/M.2

In addition to line-rate data processing services, that include encryption and error correction sub-engines to reduce host CPU overhead, the Xsight Labs E1-AIC DPU platform is capable of co-hosting additional service-layer workloads, including key/value stores or metadata engines, offloading them from the main server CPU for better efficiency and performance. Here to, many standard open-source solutions, such as MongoDB and others, are operable on the DPU itself.

Leveraging the in-SoC SRAM-based system-level-cache (SLC), and onboard DRAM as a cache layer, the Xsight Labs E1-AIC DPU boosts data locality and accelerates data-heavy applications without modifying the main host workload.

Locally Attached Storage

The Xsight Labs E1-AIC DPU supports optional M.2 or U.2 connectors, enabling direct attachment of high-speed SSDs. These devices are locally managed by the DPU, and can also be exposed to the host system as standard NVMe or virtio-blk block devices, effectively allowing the Xsight Labs E1-AIC DPU to act as a standalone storage target within the server.

This configuration minimizes storage access latency by eliminating network hops, allowing in-line data services, such as encryption, compression, or deduplication, to be directly performed on the DPU. Additionally, it creates opportunities for advanced tiered storage models. Combining its local storage and its RDMA NIC capabilities, the Xsight Labs E1-AIC DPU can coordinate between local SSDs (hot tier), remote HDD arrays (cold tier), and object-based archives, optimizing storage pipelines for AI/ML workloads such as training datasets, inference caching, or RAG memory layers.

This architecture offers a compelling path for performance-sensitive applications and power or space-constrained environments like edge data center.

Potentially, it can also boost ephemeral storage service performance, where temporary storage is tied to the lifecycle of a virtual machine (VM) instance in cloud environments.

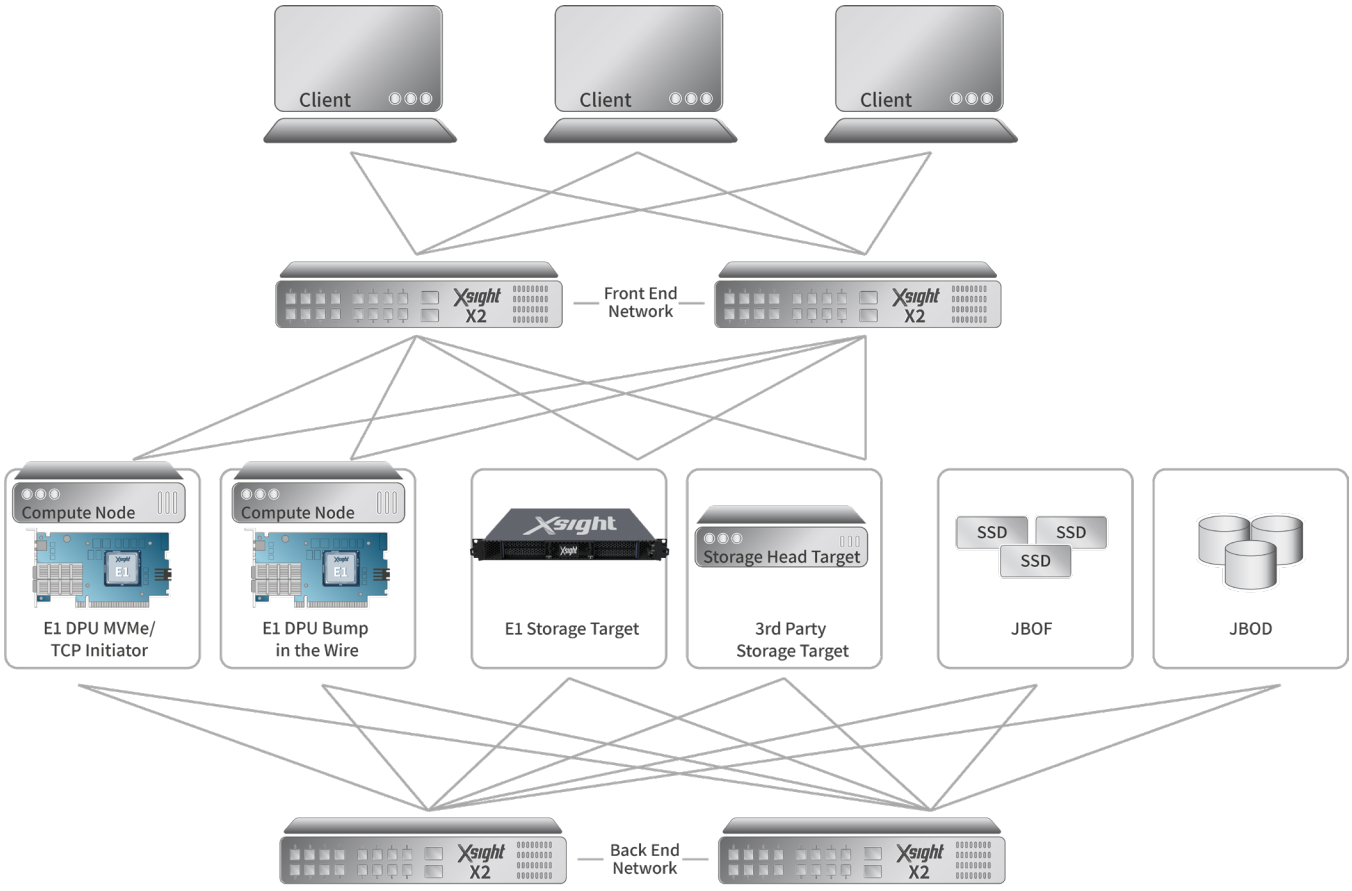


Figure 6 - The Xsight Labs E1 is a valuable element anywhere in the storage area network. As an initiator add-in DPU (left), the Xsight Labs E1-AIC supports all the networking storage initiator standards, including NVMeoF (RDMA), NVMeoTCP, and iSCSI. The Xsight Labs E1 DPU either accesses storage heads (middle) or even directly manages JBOD/JBOF headless storage over the network (right). As a bump in the wire DPU add-in card, (2nd from left) the Xsight Labs E1 injects new functionality to storage traffic, such as compression, encryption of de-duplication, without changing the topology. Finally, the Xsight Labs E1-SOC, as a main CPU, can independently manage a storage target head (3rd from left), providing unprecedented storage feature support.

Transparent In-Line Accelerator

The Xsight Labs E1-AIC DPU can seamlessly be integrated into third-party storage servers as a transparent in-line accelerator, without any change to the existing network topology or host software. This enables organizations to modernize and scale their existing infrastructure incrementally, without the need for disruptive forklift upgrades; while preserving their current investments. This is a non-evasive approach, that adds no friction.

In this role, the Xsight Labs E1-AIC DPU functions as a complementary accelerator for existing storage appliances. It enhances any storage stack by enabling advanced data services such as encryption, compression, deduplication, and in-line RDMA processing. These capabilities are delivered without impacting upstream systems, ensuring compatibility and smooth integration across diverse storage environments.

RDMA-Enabled DPU at the Center of AI Cluster Fabrics

Within AI clusters, the Xsight Labs E1-AIC DPU is a strategic interconnect, bridging three critical fabrics:

- **Host Interface Fabric –** A PCIe connection between the DPU and the host CPU, where the Xsight Labs E1-AIC DPU exposes a standard storage or network interface (e.g., virtio-blk, NVMe).
- **GPU Interface Fabric –** RMDA direct over PCIe links from the Xsight Labs E1-AIC DPU to one or more GPUs, enabling high-bandwidth data flow from host to accelerator without routing via main CPU.
- **Network Fabric –** 100/200/400GbE RDMA capable interfaces that connect the DPU to the cluster's external fabric for distributed training and inference.

A fourth fabric, GPU-to-GPU interconnects, such as NVLink or UALink, is managed outside of the DPU domain but complements the Xsight Labs E1-AIC DPU's role in the data pipeline.

An array of 16 configurable PCIe Gen5 lanes (in addition to the 16-lane host interface), is central to this architecture. It allows fine-grained allocation of PCIe resources, to meet topology requirements. This flexibility is essential for AI workloads, where maximizing bandwidth and minimizing latency between compute, storage, and networking are key requirements.

Storage Head Server

The Xsight Labs E1-Server reference platform brings a compelling blend of power efficiency, deterministic performance, and built-in data services, as a storage head. Its high-density, low-power core architecture is ideal for running in-line operations such as encryption, compression, and deduplication, while concurrently serving as a high-performance RDMA endpoint.

Unlike general-purpose x86 platforms that in many cases overprovision for baseline storage workloads, the Xsight Labs E1-Server delivers optimized compute-per-watt, seamlessly integrating with standard storage stacks (e.g., SPDK or NVMeoF). This is particularly attractive for use in disaggregated storage architectures, AI pipelines, and micro data center deployments, where footprint, power, and scalability matter most.

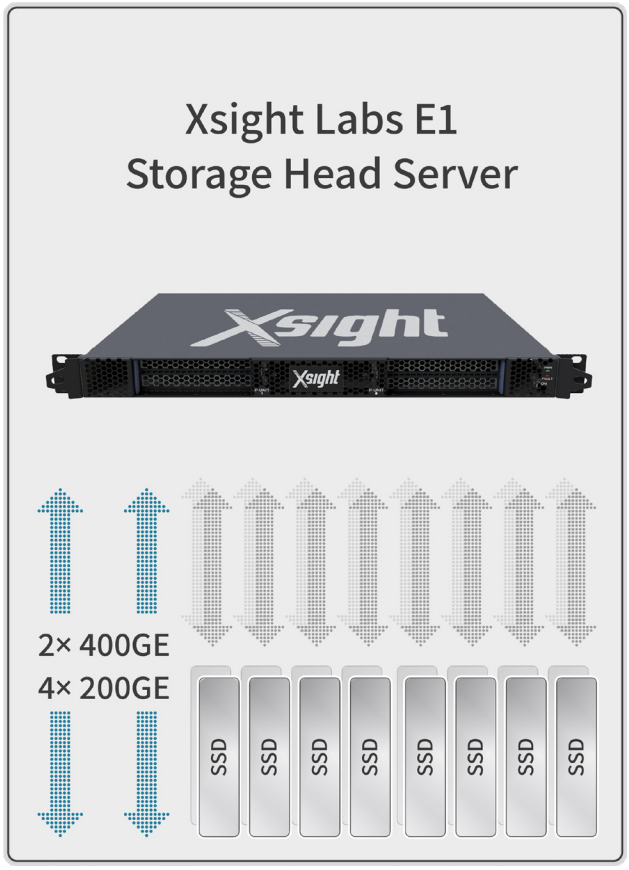


Figure 7 - Xsight Labs E1-Server as a Storage Head Server

Measurements and Core Capabilities

To evaluate the Xsight Labs E1's real-world performance as a DPU or a storage target processor, this section presents key metrics, reflecting its impact across public cloud, enterprise environments, and AI workloads. These measurements highlight how the Xsight Labs E1 accelerates storage, networking, and security tasks, enabling scalable, high-performance infrastructure with efficient resource utilization. Table 1 summarizes performance across various deployment models, demonstrating how the Xsight Labs E1 infrastructure processor delivers practical value from edge to hyperscale.

Function	Performance Target	Target Workload
DPDK TestPMD	43 Mpps per core	Bump in the wire
E-unit Packet Rate	200Mpps	Storage initiator or target
E-unit throughput	400Gbps at 460B packets	Storage initiator or target
VirtIO/XNA	15 Mpps per core	Storage initiator or target
RDMA	5 Mpps per core	Storage initiator or target
NVMe	3 MIOPS per core	Storage initiator
NVMe/TCP	2.8 MIOPS with 10 cores	Storage initiator or target
Bump-In-Wire Latency vSwitch + Network side Security	< 3usec latency	Bump in the wire
XNA PCIe Attached Host vSwitch + Network side Security	< 7usec latency	Locally attached storage/initiator
NGINX Reverse Proxy TPS (no authentication)	17500 per core	Bump in the wire

Table 1 – The Xsight Labs E1 infrastructure processor foundational capabilities, mapped to storage use cases

Summary

Storage architectures are under growing pressure from AI-driven data pipelines, cloud-scale demands, and power efficiency goals. In this evolving landscape, DPUs are emerging as foundational components, and the Xsight Labs E1-AIC exemplifies this shift.

By offloading intensive storage and networking tasks from the host CPU, the Xsight Labs E1-AIC DPU delivers measurable improvements in performance, security, and resource utilization, while significantly reducing power consumption. In addition, the Xsight Labs E1-Server provides a reference of the E1-SoC as a feature rich, dense and future proof main processor.

With support for a broad range of deployment models, including hyperconverged infrastructure, AI clusters, and edge environments, the Xsight Labs E1 infrastructure processor architecture offers flexibility without sacrificing standards compliance. Its programmable architecture and native support for open storage stacks make it a future-proof investment.

Now is the time to rethink how your storage architecture scales. Discover how the Xsight Labs E1-AIC and E1-Server unlock new levels of infrastructure efficiency and resiliency.

Contact us

Contact us at **sales@xsightlabs.com** to learn more or request a demo.
www.xsightlabs.com