

NVMe-over-Fabrics: Accelerating Data Center Innovation in the AI Era

NVMe-over-Fabrics (NVMe-oF™) is a networked storage protocol that enables storage to be disaggregated from compute, making it widely available to multiple applications and servers. By enabling shared access to a common storage pool, data can be easily utilized across applications and capacity can be allocated regardless of location.

Leveraging NVMe™ device-level performance, NVMe-oF delivers the lowest end-to-end latency from application to shared storage. It enables a composable storage infrastructure that can deliver the low latency and performance benefits of direct attached storage, while providing the agility and flexibility of sharing storage and compute. From transaction processing to real-time analytics to machine learning and beyond, NVMe-oF can elevate your data strategy to the next level.

Ethernet as the Fabric of Choice

Western Digital's OpenFlex® Data24 and RapidFlex™ product lines use Ethernet as the fabric for NVMe-oF, delivering high bandwidth, low latency, and future scalability. As Ethernet replaces SAS in AI data centers, it enables unified block, file, and object storage across disk and flash, simplifying infrastructure with a single, widely adopted interface. Ethernet's broad ecosystem and compatibility across hardware platforms support flexible, cost-effective storage solutions and pave the way for dynamic, software-defined architectures tailored to evolving workload demands.

Accelerating AI and HPC Workloads

AI and HPC pipelines are increasingly data-intensive, requiring rapid access to vast datasets for training, inference, and simulation. Traditional storage protocols like SATA and SAS struggle to keep up with the throughput and latency demands of modern compute environments. NVMe-oF extends these benefits across networked environments, enabling scalable, high-performance storage access.

Supporting both RDMA and TCP transport protocols, NVMe-oF allows organizations to tailor deployments based on latency, throughput, and infrastructure requirements.

A Scalable Foundation For the Future

As AI and HPC workloads continue to evolve, NVMe-oF will remain an integral part of AI-native infrastructure. Its ability to deliver high-speed, low-latency access to data makes it ideal for disaggregated environments where storage and compute need to scale independently. With technologies like GPUDirect Storage enabling direct data paths between NVMe storage and GPU memory, NVMe-oF will play a critical role in eliminating bottlenecks and accelerating performance. NVMe-oF will integrate more deeply with orchestration platforms, AI frameworks, and composable infrastructure strategies - empowering agile, scalable, and intelligent data ecosystems