Enterprise AI Infrastructure

Bring data lake, training, inference and retention together - with the performance, scalability and control enterprise AI demands.

Let AI ambition grow without creating new data silos.

Every successful AI initiative depends on data that is ready, accessible and protected. When datasets, models and outputs remain tied to individual GPU servers, teams lose time moving data, infrastructure becomes harder to scale and valuable compute resources are pulled into storage operations.

What holds AI back

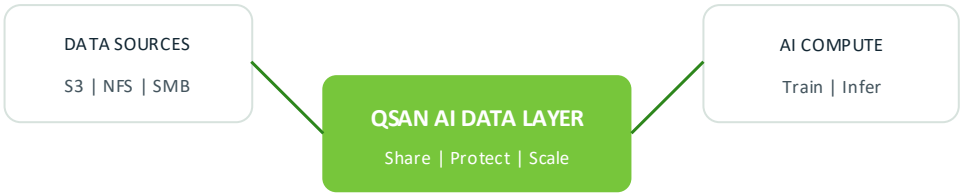
Data is tied to compute

Storage growth can force changes to GPU nodes, while every local copy adds cost, risk and management effort.

Workflows remain disconnected.

Data lake, training, inference, analytics and backup depend on separate paths and repeated transfers.

The QSAN advantage: One Data Foundation



QSAN brings datasets, model files, vector data, checkpoints and generated outputs into one enterprise AI Data Layer. GPU nodes receive a high-speed data path, teams gain native access to shared data, and the business retains control of its most valuable AI assets.

Business advantages built into the solution

01 Keep AI engines moving

QSAN combines GPUDirect Storage-ready data paths with RDMA connectivity to move data efficiently and reduce CPU involvement. Shared file and high-throughput block access give each AI workload the right path to the data it needs.

02 Create one connected workflow

Native S3, NFS and SMB services connect ingestion, analytics and backup with the same data foundation used by training and inference. QSAN helps teams collaborate without building another storage island for every new AI project.

03 Grow at the pace of opportunity

QSAN separates data growth from GPU acquisition. Expand performance or capacity as datasets, models and users increase, then connect additional AI nodes to the shared foundation without redesigning the environment around local storage.

04 Make every storage tier count

Place active datasets, current models and checkpoints on high-performance resources. Retain historical models, logs and generated content on capacity-efficient tiers, so performance stays available where the business needs it most.

QSAN

More value from every stage of enterprise AI

QSAN turns storage into a strategic AI foundation - accelerating access, simplifying collaboration and protecting data from the first dataset through production and long-term retention.

01	02	03	04	05	06
Ingest	Prepare	Train	Checkpoint	Infer	Retain
S3 / file	Shared data	GPU feed	Save / restore	Model load	Protect / Archive

More value from every GPU

QSAN keeps data moving so accelerators can focus on AI. Validated GPUDirect Storage performance over NFS-RDMA creates a direct storage-to-GPU path and reduces CPU involvement. Shared NFS access supports datasets and model repositories, while NVMe-oF/RDMA serves high-throughput checkpoint and block workloads. Teams gain the flexibility to match each workload to the right data path.

Infrastructure that grows with AI

AI data and GPU demand grow at different rates. QSAN lets organizations expand storage independently and connect new compute nodes to existing datasets. Unified data services, disaggregated NVMe performance and flexible hybrid capacity expansion support a larger shared AI environment. A hybrid architecture aligns high-performance resources with active work and capacity-efficient storage with long-term retention.

Shared data. Faster collaboration.

QSAN connects data engineering, training, inference, analytics and backup through a shared data foundation. Native S3, NFS and SMB services let teams use the access method their workflow requires, with S3 and NFS seamless operation on concurrently consolidated platform. Centralized models, vector data and checkpoints help teams reuse AI assets and connect new GPU nodes without another full data copy.

Protection across the AI lifecycle

AI value extends from source datasets to tuned models, checkpoints and generated outputs. QSAN combines not only essential RAID, high available hardware, but mature software centralized protection strategy. Keep active work on high-performance resources and retain older content on capacity-efficient tiers in multiple failover mechanism. This lifecycle approach balances performance and cost while keeping sensitive AI assets under organizational control on premises.

99.83% AU
Training utilization

Up to 3x
Checkpoint restore

65.1% faster
32B model loading

11.7 GiB/s
Peak GDS read

Read the full lab report

Explore QSAN AI Solution

RECOMMENDED PRODUCTS

XN SERIES	Unified storage	Shared AI data services and native multiprotocol access
XE SERIES	NVMe-oF JBOF	Disaggregated NVMe performance and independent expansion
XD SERIES	DAS / JBOD	Flexible capacity expansion for growing AI data