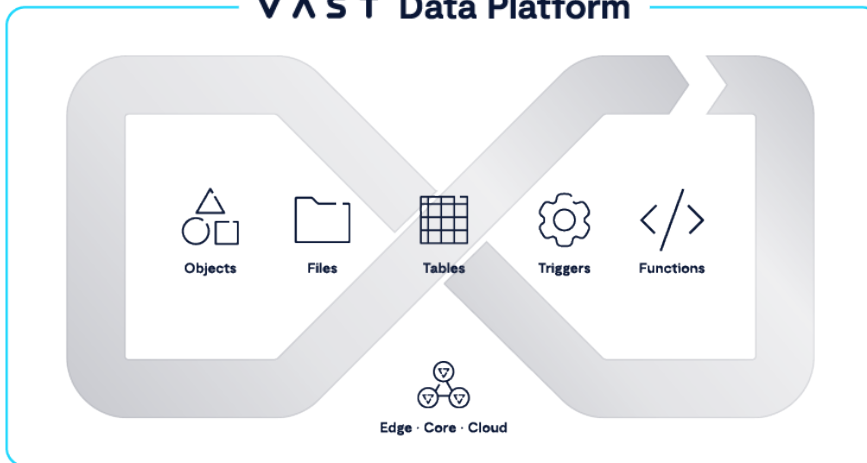


Solution Brief

High-Performance Storage Solution Brief for NVIDIA Cloud Partners



VAST Data Platform



The VAST Data Platform For NVIDIA Cloud Partners

Certified as a high-performance storage solution for the NVIDIA Partner Network, the VAST Data Platform empowers cloud service providers to deploy AI models at an unprecedented scale with confidence and security. With VAST Data, Cloud Solution Providers can help their client organizations seamlessly scale their AI operations across thousands of GPUs, ensuring industry-leading uptime and reliability.

Building on extensive experience with large-scale NVIDIA HGX-based server deployments, VAST Data supports the NCP RA with a validated, fully integrated architecture backed by our global support services team. The VAST Data Platform combines the high performance and scalability required for AI workloads with the simplicity and resiliency needed by CSPs. This platform is ideal for service operators, offering industry-leading uptime and cost efficiency.

Benefits for NVIDIA Cloud Partners:



Service provider-grade reliability

Observed production systems realizing 99.9999% availability



Secure multi-tenancy

Per tenant data encryption, flexible network segmentation, robust audit capabilities



Fine-grained workload isolation

Granular QoS policies prevent multi-tenant IO contention ("noisy neighbor")



Infrastructure cost savings

Consolidate storage silos into a single, affordable all-flash solution and permanently eliminate the overhead of data copies and sprawl



Improved operational efficiency

Online operations (such as upgrades) never require downtime, leading to better customer experience with fewer admins required

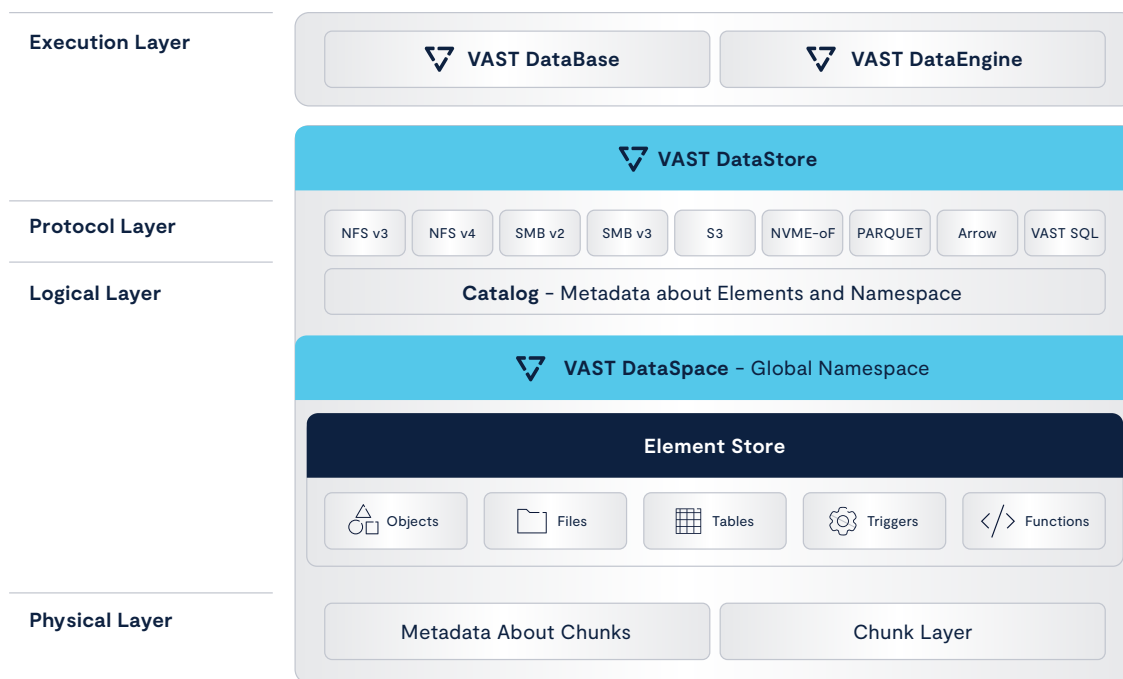


Figure 2. VAST Data Platform is a single namespace solution that provides multi-protocol file and object storage (DataStore), high-performance data warehousing (DataBase), and multi-site global data access (DataSpace)

The VAST Data Platform

The VAST Data Platform pioneered an architecture that delivers not only performance and scale but also operational efficiency and ease of use. With flexible configurations to support over 16K GPUs in a single cluster, NVIDIA Cloud Partners (NCPs) can confidently pair VAST Data solutions with any scale of large AI infrastructure deployments. Based on the NCP Reference Architecture (RA), featuring NVIDIA HGX H100 platforms, this jointly validated VAST Data reference architecture meets and exceeds the intensive demands of AI, high-performance computing, and other data-intensive applications.

The VAST Data Platform provides additional data services—such as multi-protocol data ingest, accelerated ELT for feature engineering—useful in helping CSPs accelerate data pipelines for their customers, all underpinned by a multi-tenant Zero Trust security model. This holistic approach not only streamlines AI pipelines by consolidating storage across various data stages but also significantly minimizes complexity and costs, thereby boosting overall system performance and security for CSPs.

Discover how VAST Data Platform helps Cloud Service Providers

Optimize AI Workflows Achieve seamless scalability and uptime, supporting the entire AI pipeline from data ingest to model serving.	Enhance Reliability Deploy AI models at scale across thousands of GPUs with secure, efficient operations.	Integrate with NVIDIA Utilize unified storage and data services deeply integrated with NVIDIA technologies for next-gen AI factories.
Infrastructure Scalable to Exabytes · Performance for AI workloads · Archive Economics · Full ACID transactions		

Explore the VAST Possibilities: Download the full Reference Architecture [here](#)