Solution Brief

VAST AI OS on AWS

One Data Estate, AWS GPUs



Enterprises everywhere are accelerating their AI programs. They have years of valuable data stored on-premises or across distributed environments, but access to large-scale compute is limited. As demand for AI training and inference grows, many organizations want to use cloud AI accelerators to supplement on-premises capacity.

The challenge is that moving or replicating data to the cloud introduces friction.

Full copies are slow, expensive, and create new compliance risks. Each additional copy of data brings the potential for drift, not only in content, but in permissions and security posture.

VAST AI Operating System changes that with a unified data view, intelligent caching, and consistent security across every environment.

A Unified Platform for AI Data and Compute

VAST on Cloud (VoC) for AWS provides companies with a consistent data foundation across AWS and on-premises VAST deployments. The VAST AI Operating System (VAST AI OS) is a software-defined, high-performance data platform specifically designed for enterprise AI, machine learning, analytics, and agentic AI. It delivers secure, governed access to enterprise knowledge.

Running on Amazon EC2 I8ge storage-optimized instances equipped with locally attached NVMe (Non-Volatile Memory Express), VAST on Cloud delivers the high-throughput performance that AI workloads demand.

The VAST DataStore unifies structured and unstructured data using the Element Store, which manages data in a protocol-independent format. Likewise, VAST on Cloud stores file and directory permissions internally in a protocol-independent format and translates them according to the security model configured for each view, helping to maintain consistent authorization across NFS, SMB, and S3.

VAST AI OS integrates directly into container environments, allowing GPUs to access data without staging or manual replication. There are no sync jobs or special workflows to maintain. Once configured, AI workloads read and write data through the same VAST mount, regardless of whether the storage is on-premises or in AWS.

VAST AI Data Platform

Performance Parity

Maximize GPU utilization without having to pre-stage a copy of all your data.

Global Availability

Access and interact with your data anywhere without copies or complexity.

Security

A unified global permissions model ensures no drift between environments.

Compatibility

Works natively with Amazon SageMaker, PyTorch, TensorFlow, JAX, Docker, and Hugging Face.

VAST DataSpace in AWS deployments connects all sites into one logical namespace within your own tenant, maintaining unified access controls, encryption, and compliance policies everywhere. You experience a single, seamless data fabric across all your VAST AI OS environments.

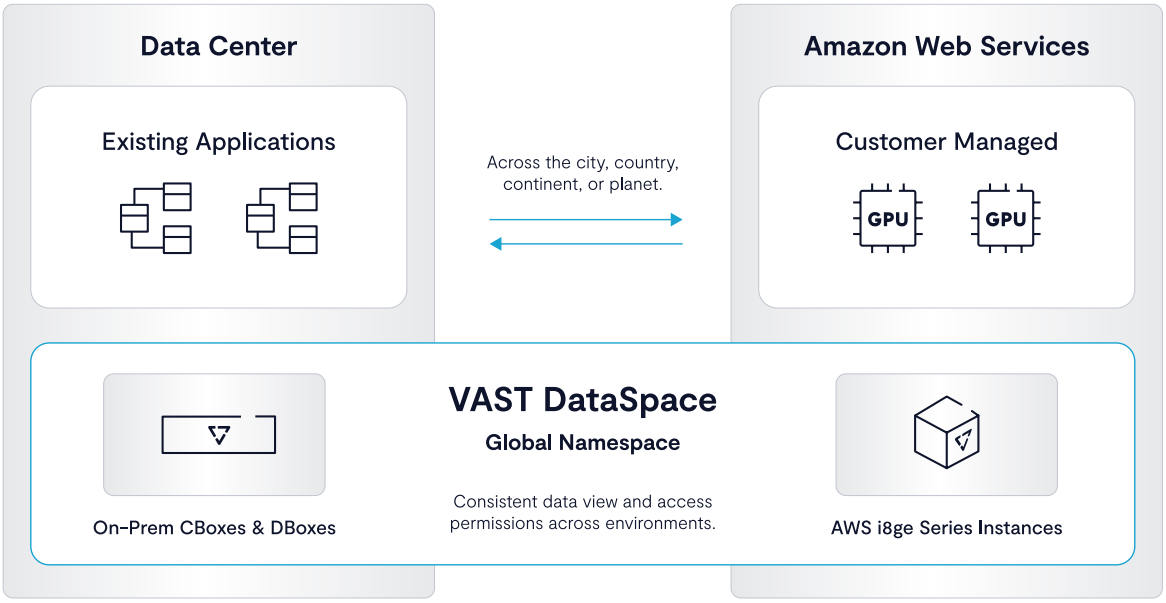


Figure 1. VAST's Global Namespace creates a single, consistent data environment that spans on-prem, or cloud operations.

Tying this together, VAST Polaris gives you a single control plane over the hybrid estate. Rather than managing each cluster, region, and location as its own silo, your team sees and governs every site through one interface, turning what would otherwise be a sprawl of independent deployments into a single, coherently operated system.

Getting Started

VAST on Cloud is available on AWS. Customers can deploy a managed VAST cluster using AWS i8ge series instances, connect it to their on-premises VAST environment, and begin running GPU workloads on existing datasets immediately. Learn more by visiting the Marketplace listing.

One Data Fabric

AI workloads can run wherever compute capacity is available while maintaining control of data and security policies.

Maximum Utilization

GPUs can access and load models from VAST storage with performance consistent with local NVMe devices.

Consistent Policies

VAST DataSpace and Polaris enforce a single trust zone across environments. Data governance remains consistent everywhere.

About VAST Data

VAST Data is the AI Operating System company – powering the next generation of intelligent systems with a unified software infrastructure stack that was purpose-built to unlock the full potential of AI. The VAST AI OS consolidates foundational data and compute services and agentic execution into one scalable platform, enabling organizations to deploy and facilitate communication between AI agents, reason over real-time data, and automate complex workflows at global scale. Built on VAST's breakthrough DASE architecture – the world's first true parallel distributed system architecture that eliminates tradeoffs between performance, scale, simplicity, and resilience – VAST has transformed its modern infrastructure into a global fabric for reasoning AI. Learn more at vastdata.com and follow VAST Data on LinkedIn, YouTube and X.

vastdata.com/demo hello@vastdata.com

©2026 VAST Data, Inc. All rights reserved. All trademarks belong to their respective owners.

11/26 I 16813