Solution Brief

# VAST AI Operating System on Amazon Web Services

VAST

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-prem capacity.

The challenge is that moving or replicating data to the cloud introduces friction. Transfers 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, now available on AWS.

## A Unified Platform for AI Data and Compute

VAST AI OS enables customers to extend their existing VAST environments into Amazon Web Services. The software deploys directly on AWS i3en.24xlarge and or i7ie.48xlarge instances and connects back to on-prem or cloud VAST clusters through the Global Namespace. The result is a single, consistent data view across all environments that VAST end users compute in.

VAST 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-prem or in AWS.

This configuration forms a single federated cluster that spans where the data lives and where the compute runs, even if the source data is distributed across multiple geographic locations. Several on prem, colo, and cloud locations can be brought together in one Namespace, creating a unified operational model for all environments.

### Benefits for Enterprise AI Deployments

**Performance parity**

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

**Global availability**

Access and interact with their data anywhere without copies or complexity

**Security**

Unified global permissions model ensures no drift between environments

**Compatibility**

Works natively with AzureML, PyTorch, Tenserflow, JAX, Docker, Hugging Face

**Scalability**

Extends to additional clusters for federated and multi-region AI

# The VAST Data Platform

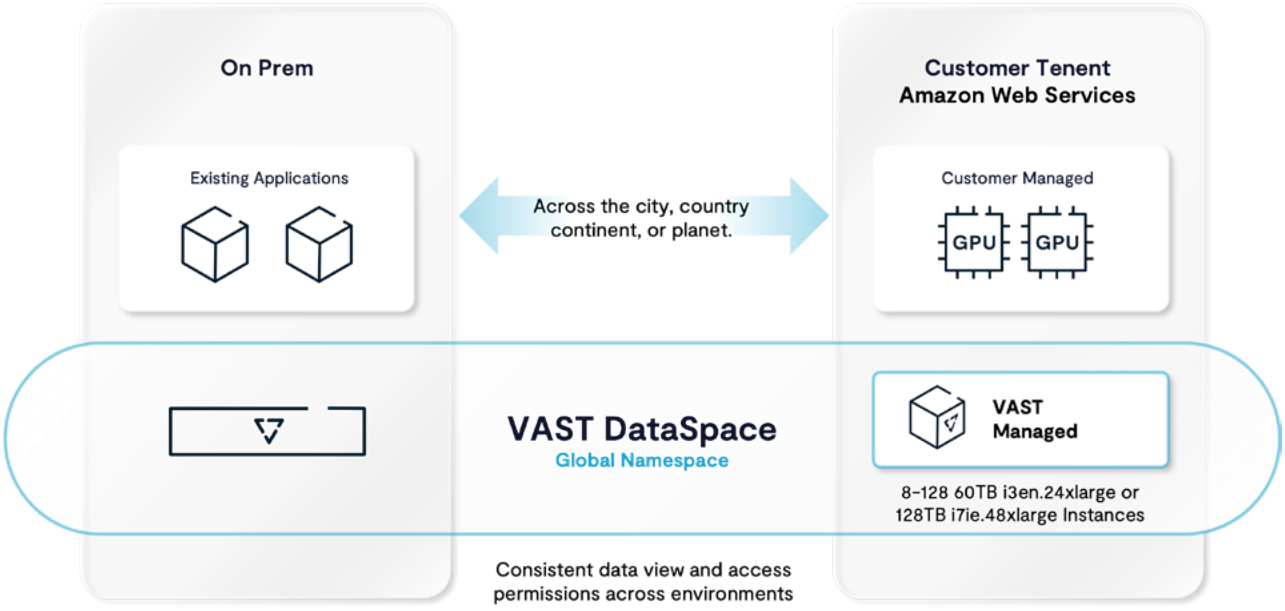



**Figure 1. VAST's Global Namespace creates a single, consistent data environment that spans on-prem, or cloud locations, even across vast distances. Existing applications can keep working locally while cloud-based AI workloads run against the same dataset without data movement, drift, or reconfiguration.**

The VAST DataSpace layer 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 environments.

## Getting Started

VAST AI OS is available now in AWS. Customers can deploy a managed VAST cluster on i3en.24xlarge and or i7ie.48xlarge instances, connect it to their on-prem 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 enforces a single trust zone across environments. Data permissions remain consistent everywhere.

**Explore the VAST Possibilities – VASTdata.com**

Scalable to Exabytes · Performance for AI Workloads · Federated AI Architecturesecurity policies

## About VAST Data

VAST Data is the leading data platform software company bringing businesses into the AI era. The VAST Data Platform delivers scalable performance, simplifies data management, and enhances productivity by accelerating time-to-insight for workloadintensive applications. Launched in 2019, VAST has become history's fastest-growing data infrastructure startup. Looking to simplify your data center and discover insights in all of your data? **Contact us at hello@vastdata.com**



©2025 VAST Data, Inc. All rights reserved. All trademarks belong to their respective owners.
01/2026