Solution Brief

VAST DataEnclave

Bring State-of-the-Art AI to the World's Most Sensitive Data



Enterprise AI faces a fundamental trust problem. Regulated organizations cannot move regulated data outside their security boundaries, while AI model providers cannot expose proprietary model weights inside infrastructure they do not control. Roughly half of the world's data lives on premises or in colocation facilities, beyond the reach of today's leading AI models, but the data cannot go to the model, and the model cannot safely come to the data.

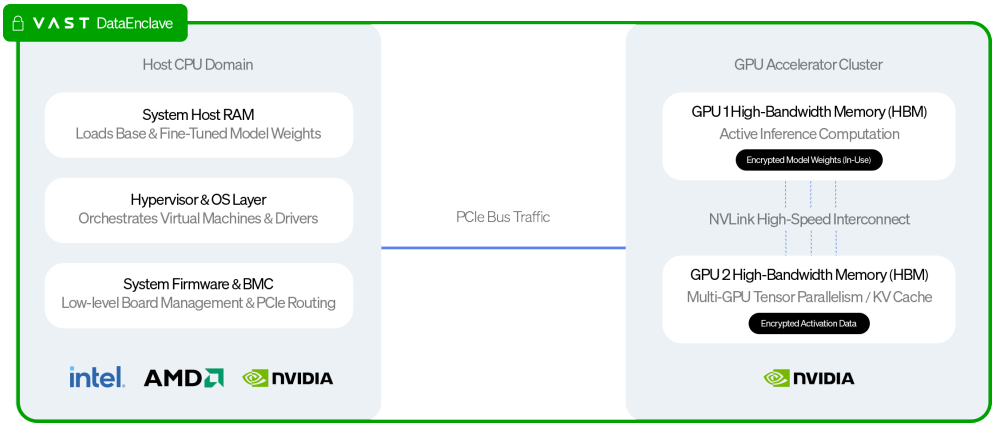
VAST DataEnclave, the confidential AI capability of the VAST DataEngine, combines a secure container runtime in VAST DataEngine, cryptographic attestation, and customer-controlled key management to create verified, hardware-isolated environments for sensitive AI workloads. Through a verify-before-decrypt architecture, data, models, keys, and secrets remain protected from infrastructure operators, administrators, and other third parties, even while AI workloads are actively processing them.

How It Works

Hardware-Isolated Execution for AI Workloads

VAST DataEnclave extends VAST's protection of data at rest and in transit to data and models while they are actively being processed. AI workloads run inside hardware-isolated confidential environments that protect sensitive data and model assets from the underlying infrastructure, including host administrators, cloud operators, and other privileged users.

The DataEnclave secure container runtime is compatible with the CNCF Confidential Containers (CoCo) ecosystem, including runtimes such as Kata Containers. This combines familiar container-based deployment with hardware-backed isolation spanning CPU memory and GPU execution, enabling organizations to run AI workloads in protected execution environments without changing how applications are packaged and deployed.



Verify Before Decrypt with Cryptographic Attestation

Hardware isolation alone does not establish that an environment should be trusted with sensitive data or proprietary models. Before access to protected assets is granted, DataEnclave cryptographically verifies that the environment conforms to an approved security policy.

The DataEnclave attestation service validates evidence covering the underlying CPU and GPU, firmware, software environment, and workload configuration against approved reference values. This ensures that protected workloads execute only within environments that meet the security requirements established by the data and model owners.

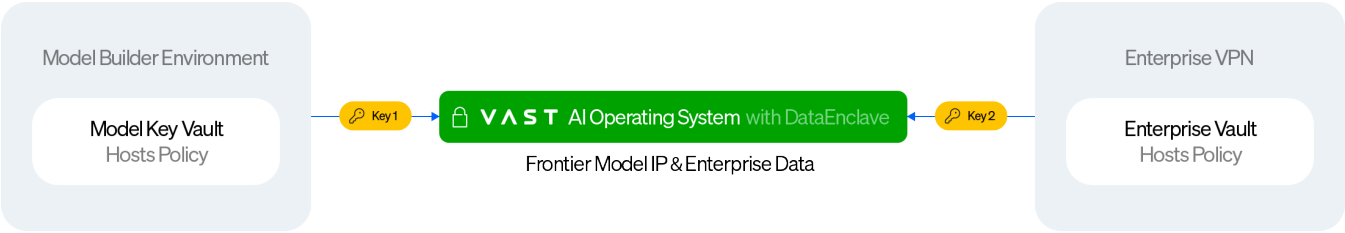
Only after this evidence passes verification are the cryptographic materials required by the workload released. If the hardware, software, or workload configuration has been modified or does not conform to policy, attestation fails and the secrets remain unavailable. The principle is simple: no attestation, no access. Sensitive assets are accessible only inside an environment that has cryptographically proven both its identity and its integrity.

Independent Control of Enterprise Data and Model IP

DataEnclave creates a neutral execution environment in which enterprises and AI model providers can collaborate without requiring either party to surrender control of its most valuable assets.

Enterprise data remains encrypted under customer-controlled keys, while proprietary model weights can remain encrypted under keys controlled by the model provider. Neither party needs to provide its keys to the infrastructure operator. Following successful attestation, the appropriate key material can be securely delivered to the verified confidential environment.

This separation of trust enables proprietary models to operate against regulated enterprise datasets without exposing the underlying data to the model provider or the model weights to the enterprise or infrastructure operator. Rather than relying on organizational trust between parties, DataEnclave establishes trust in the execution environment through cryptographic proof.



Confidential GPU-Accelerated AI

DataEnclave extends confidential execution to GPU-accelerated AI workloads, protecting sensitive data and models throughout accelerated processing. GPUs operating in Confidential Compute mode are incorporated into the attested environment, with hardware-backed protections for GPU execution and attestation verifying that CPU and GPU resources meet approved security requirements before keys and secrets are released.

This enables organizations to run high-performance AI inference while keeping model weights, enterprise data, and processing isolated from the underlying infrastructure and its administrators. Applications continue to use familiar AI frameworks and development environments, combining accelerated AI performance with the confidentiality and integrity required by both enterprise data owners and model providers.

Flexible Key Management and Trust

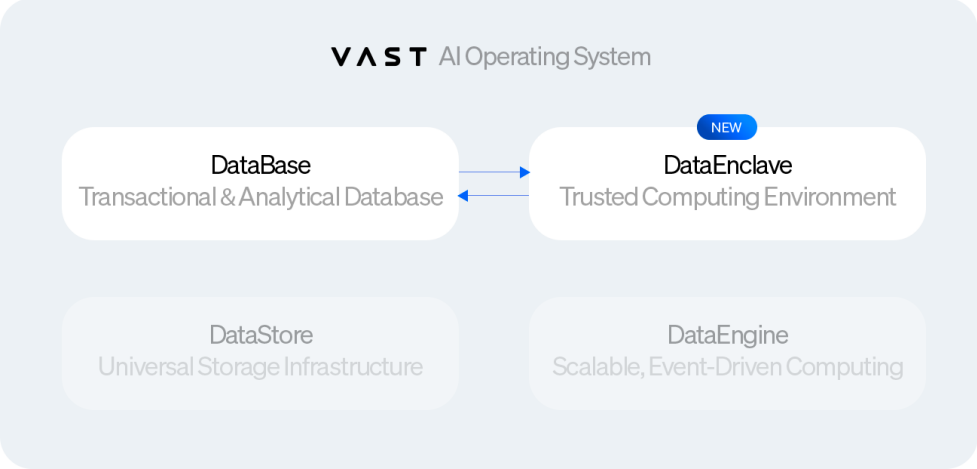
DataEnclave is designed to preserve control over cryptographic keys rather than centralizing that control within the infrastructure platform. Rather than introducing a proprietary VAST key store, DataEnclave integrates with customer-controlled and third-party key management infrastructure.

This separation between execution, attestation, and key ownership enables organizations to align confidential AI deployments with existing security boundaries while ensuring that control of the underlying infrastructure alone never grants access to protected data or models.

Governed and Auditable by Design

DataEnclave maintains security and compliance visibility without compromising the confidentiality of protected workloads. Every attestation event, key release, policy decision, and workload lifecycle action is captured in a tamper-proof audit trail in VAST DataBase, providing a queryable record of what ran, where, and under what verified policy.

Security, compliance, and operations teams can query this history for regulatory audits, incident response, and forensic investigation without exposing the sensitive data or proprietary models protected inside the confidential environment. This provides accountability across the confidential AI lifecycle while preserving the isolation that DataEnclave is designed to enforce.



Confidential AI in Connected and Air-Gapped Environments

Some of the organizations with the strongest need for advanced AI also operate under the strictest infrastructure constraints. Government, defense, financial services, healthcare, and other regulated environments may require workloads to operate behind enterprise firewalls or inside completely disconnected infrastructure. DataEnclave is designed to support data sovereignty and regulatory requirements in these environments while keeping data and models under their owners' control.

For connected deployments, the DataEnclave attestation service is compatible with the open CNCF Confidential Containers Trustee stack and integrates with standard cloud and on-premises key management systems. For fully air-gapped environments, DataEnclave supports on-premises attestation and key brokering without requiring an external network connection.

This enables DataEnclave to extend the same verify-before-decrypt trust model from connected environments to fully air-gapped infrastructure, bringing advanced AI to sensitive data wherever it must remain.

Build a Confidential AI Factory

VAST DataEnclave enables enterprises, model builders, AI cloud providers, and infrastructure partners to build confidential AI factories where state-of-the-art AI can operate on the world's most sensitive data without requiring either the data owner or model owner to relinquish control.

- **Enterprises** can run proprietary foundation models and fine-tuned AI against regulated data inside their own security domain, while securely deploying their own models wherever they need them.
- **Model builders** can bring closed-weight models to sensitive enterprise workloads that previously required local or open-weight alternatives while protecting model IP.
- **AI cloud and sovereign cloud providers** can host proprietary models and regulated enterprise data without gaining access to either, opening workloads and markets where operator access would otherwise violate security or sovereignty requirements.
- **OEM partners** can deliver turnkey infrastructure with confidential compute, accelerated AI, and VAST data services pre-integrated to meet model providers' deployment requirements.

By bringing together secure execution, cryptographic attestation, customer-controlled key management, and auditable governance, VAST DataEnclave extends VAST's data protection model across the complete data lifecycle, protecting data at rest, in flight, and now in action so organizations can securely operationalize AI on data that could never leave their control.



Why VAST

Protect Data in Action

Extend VAST data protection from data at rest and in flight to sensitive data and models while they are actively being processed.

Verify Before Decrypt

Cryptographic attestation ensures the environment meets approved security policies before keys or other secrets are released.

Protect Both Sides

Enterprise data and proprietary model IP remain under their respective owners' control.

Confidential Accelerated AI

Run GPU-accelerated AI inside attested environments within NVIDIA confidential compute mode.

Deploy Anywhere

Support connected, private, sovereign, edge, and fully air-gapped confidential AI environments.

Auditable by Design

Maintain a tamper-proof audit trail of attestation, key releases, policy decisions, and workload activity without exposing protected assets.

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**.

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

09/26 | 14139