Solution Brief

Understanding the Software Stack of the VAST AI Operating System

A Newcomer's Guide

Welcome to VAST! If you're looking to understand our core product, the VAST AI Operating System, you're in the right place. This document will break down what it is, its key components, and how they work together to power the next generation of artificial intelligence applications.

The Evolution: From Revolutionary Architecture to an AI Operating System



VAST began its journey by revolutionizing data storage for large scale high performance workloads. As shown in the first image, our foundation was laid with DASE (Disaggregated and Shared Everything architecture) in 2016. This groundbreaking approach allowed us to build Universal Storage by 2019 – a single tier of storage designed to handle all data types with extreme performance and scalability.

Recognizing the growing needs of AI, we evolved our offering into an AI Data Platform by 2023, providing structured and unstructured data services specifically for AI workloads. Today, this has culminated in the VAST AI Operating System, launched in 2025. It's a comprehensive AI platform designed for distributed agentic computing, aiming to provide a unified environment for data, an engine for insights, and a framework for intelligent automation.

What is the VAST AI Operating System?

At its core, the VAST AI Operating System is a software platform that simplifies and accelerates the entire AI pipeline in constant realtime flow, from raw data ingest, through context analysis and ultimately, intelligent action. It's designed to be the foundation upon which AI-driven applications and autonomous agents are built, deployed, and managed efficiently.

Think of it like the operating system on your computer, but built for the massive distributed scale and unique demands of AI. It manages the underlying resources across hundreds to thousands of systems, provides essential services, and offers a consistent environment for AI applications to run.

The VAST AI Operating System offers total flexibility with no vendor lock-in. As a software-defined platform, it runs on your preferred hardware or within the cloud of your choice—whether on-premises, hybrid, or public. It remains entirely model-agnostic, letting you deploy any AI model you need. VAST acts as the ultimate orchestrator, bridging your infrastructure and intelligence with a unified data layer and management framework necessary to ensure seamless performance at scale.

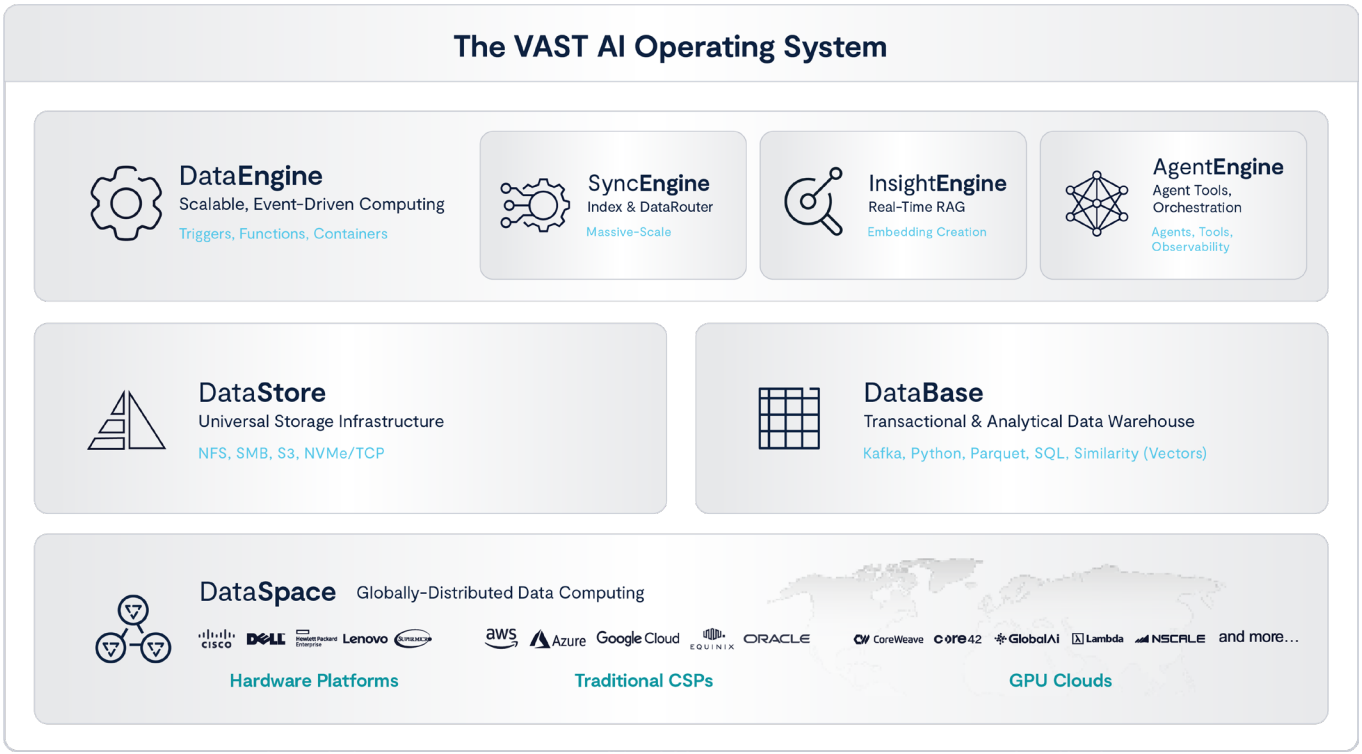
Deconstructing the VAST AI Operating System: Key Components and How They Interact

The Architecture: Disaggregated and Shared Everything (DASE)

At the heart of VAST's technological innovation is its Disaggregated and Shared Everything (DASE) architecture, first introduced in 2016. This architecture was a fundamental rethinking of how distributed systems should be built to overcome the limitations of traditional designs, particularly as the scale of data and computing continued to explode.

DASE is built on the principle of disaggregating (separating) compute logic from the physical storage media, while allowing all compute nodes to access all storage devices in parallel. Using high-performance NVMe-over-Fabrics, CPUs can access remote SSDs as if they were local. This "shared everything" approach, combined with novel transactional data structures, eliminates the need for data partitioning and complex inter-server coordination typical in "shared-nothing" architectures. The result is a highly scalable and resilient system that breaks traditional tradeoffs between performance and capacity. DASE enables terabyte-to-exabyte linear scalability, simplifies data management, and allows for unprecedented efficiency in utilizing all-flash storage. As a software-defined system, able to run on your choice of vendor server or cloud platform, this architectural foundation was crucial for VAST's evolution from Universal Storage to a comprehensive AI Data Platform and ultimately, the VAST AI Operating System.

Let's look at the key layers and components of the VAST AI Operating System built upon this DASE foundation, as illustrated in the provided architecture diagrams:



The Foundation Data Platform Services: DataSpace, DataStore, and DataBase



VAST DataSpace: This is the globally-distributed data computing layer. It allows organizations to manage and access data that might be spread across different locations, whether on-premises, in the cloud, or at the edge. It provides a unified view and access layer for all your data.



VAST DataStore (Universal Storage Infrastructure): Building on our Universal Storage roots, the DataStore is where your data lives. It's a high-performance, scalable, and resilient file, object and block storage system. It's designed to ingest, store, and provide rapid access to the massive datasets required for AI, including files, objects, and volumes.



VAST DataBase (Universal Database Infrastructure): This isn't just a traditional database. It's a database built for AI, capable of handling diverse data types including structured data (like tables in an Enterprise Data Warehouse - EDW), contextual meta data about unstructured data (like text and images), vectors (critical for many AI algorithms like similarity search), streams (for real-time data processing), catalogs, and logs.

How they work together:

The DataSpace provides the overarching fabric. Within it, the DataStore efficiently holds all types of raw and processed data, while the DataBase organizes and makes this data queryable and usable for AI model training and inference.

The Engines: DataEngine, SyncEngine, InsightEngine, and AgentEngine

Building on the foundational data platform services, the VAST AI Operating System provides specialized engines to transform raw data into intelligent, autonomous action. The following sections outline the core execution environments and how they unify the AI lifecycle.



VAST DataEngine (Universal Computing Environment for Agentic Systems): The DataEngine is the execution and orchestration supervisor for the AI OS. It transforms distributed computing resources into an "integrated thinking machine" by managing serverless function pipelines. It uses Event Triggers (CRUD events, table updates, or analytics counters) and Functions (Python-based microservices) to process data. By decoupling applications from "data gravity," it intelligently moves compute to where the data resides, optimizing for cost, execution time, and hardware requirements like GPUs. It includes a native, Kafka-compatible Event Broker that stores message topics as VAST Tables, allowing for global message ordering and real-time SQL querying of live streams.



VAST SyncEngine: The SyncEngine is a highly scalable data discovery and migration service that deploys within the DataEngine. It acts as a universal "bootstrapper," indexing and synchronizing data from external sources—including S3 buckets, file systems, and SaaS platforms like Confluence or Google Drive—into the VAST DataSpace. It maintains a durable, searchable index in the VAST DataBase and triggers enrichment pipelines automatically as new data is onboarded, ensuring datasets stay current without manual scripts.



VAST InsightEngine: Focused on extracting intelligence, this engine provides real-time embedding microservices that transform data into AI-readable formats. It handles agent and model orchestration, tool services, and pipeline observability, effectively turning raw data into actionable insights for RAG and agentic workflows.



VAST AgentEngine (Available 2026): Arriving in 2026, the AgentEngine is the production-grade deployment and orchestration layer for AI agents. It bridges the gap between notebook prototyping and enterprise scaling by providing a robust runtime for long-running, containerized "services." It manages the agent lifecycle, persistent state, and secure service discovery. Supporting any Python-based framework (like LangGraph or Agno), it utilizes the Model Context Protocol (MCP) to allow agents to interact with VAST native services and custom tools. Every action within the AgentEngine is auditable and traceable, ensuring compliance and observability across complex, multi-agent "stacks."

How they work together:

The SyncEngine discovers and onboards data, which the DataEngine then processes via event-driven functions. The InsightEngine contextualizes this data, providing the embeddings agents need to reason. Finally, the AgentEngine provides the stable, secure environment where these agents execute complex tasks by leveraging tools and data across the entire VAST ecosystem.

Putting It All Together: A Unified Platform for AI

The VAST AI Operating System replaces fragmented toolchains with a cohesive environment built on the Disaggregated and Shared Everything (DASE) architecture. By consolidating infrastructure that traditionally required separate silos—object stores, compute clusters, Kubernetes pods, and message brokers—into a single pool of resources, VAST eliminates performance bottlenecks and improves utilization across the datacenter.

The journey from data to action is seamless:

01

Ingest & Discover
The SyncEngine pulls data from across the enterprise into the DataStore, preserving identity and continuity.

02

Process & Stream
The DataEngine orchestrates serverless functions and manages real-time event streams via the Event Broker, preparing data for AI without the constraints of "data gravity."

03

Contextualize
The InsightEngine extracts deep patterns and generates vector embeddings within the DataBase, making the information immediately queryable and "readable" for LLMs.

04

Execute/Action
The AgentEngine (launching 2026) sits at the top of the stack, empowering organizations to build and manage AI Agents that perform complex, multi-step tasks with full observability and governance.

By providing a comprehensive and unified platform, the VAST AI Operating System empowers organizations to unlock the full potential of their data and accelerate their AI initiatives, moving from complex, fragmented toolchains to a cohesive environment for building and running intelligent systems.

About VAST Data

VAST Data is the leading data platform software company bringing businesses into the AI era. By accelerating time-to-insight for workload-intensive applications, the VAST Data Platform delivers scalable performance, simplifies data management, and enhances productivity. Launched in 2019, VAST has become the fastest-growing data infrastructure startup in history.

Looking to simplify your data center and unlock insights from all of your data? Contact us at hello@vastdata.com

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

02/2026