White paper

# VAST DataBase: Unifying Transactions, Analytics, and AI at Scale

09/25 | Version 1.0

# Table of Contents

# Executive Summary

Modern enterprises are struggling to operationalize AI and analytics at scale due to fragmented and complex data architectures. The VAST DataBase solves this problem by providing a single, hyper-scale platform for structured data. By breaking the traditional tradeoffs between transactional and analytical workloads, VAST DataBase simplifies the data landscape and delivers unprecedented performance for the AI era.

This document provides a technical overview of the VAST DataBase architecture and its core capabilities, demonstrating how it enables real-time insights, simplifies operations, and delivers a streamlined path to enterprise AI.

# The AI Data Challenge

Traditional data systems—including relational databases, data warehouses, and data lakes—were not designed for the real-time, high-concurrency demands of modern AI and analytics. Enterprises face three primary obstacles:

**Fragmented architectures:** Data is siloed across separate systems, requiring complex ETL pipelines that introduce latency and overhead.

**Performance bottlenecks:** Legacy “shared-nothing” designs move massive amounts of data across networks, creating latency that stalls real-time AI.

**Operational complexity:** Data teams spend time on manual tasks such as table maintenance, partitioning, and compaction, distracting them from high-value projects.

Enterprises need a new class of database that collapses silos, eliminates ETL, and delivers unified performance across both transactional and analytical workloads. This is the problem VAST DataBase was built to solve.

## The Limits of Sharded Database Architectures

Most modern databases attempt to scale through sharding, but this shared-nothing model introduces inefficiencies. Queries must fan out across shards, generating heavy east-west traffic, while developers and DBAs constantly rebalance partitions and manage hotspots. As data grows, coordination overhead mounts and performance flatlines.

VAST’s Disaggregated Shared-Everything (DASE) architecture removes these constraints entirely.

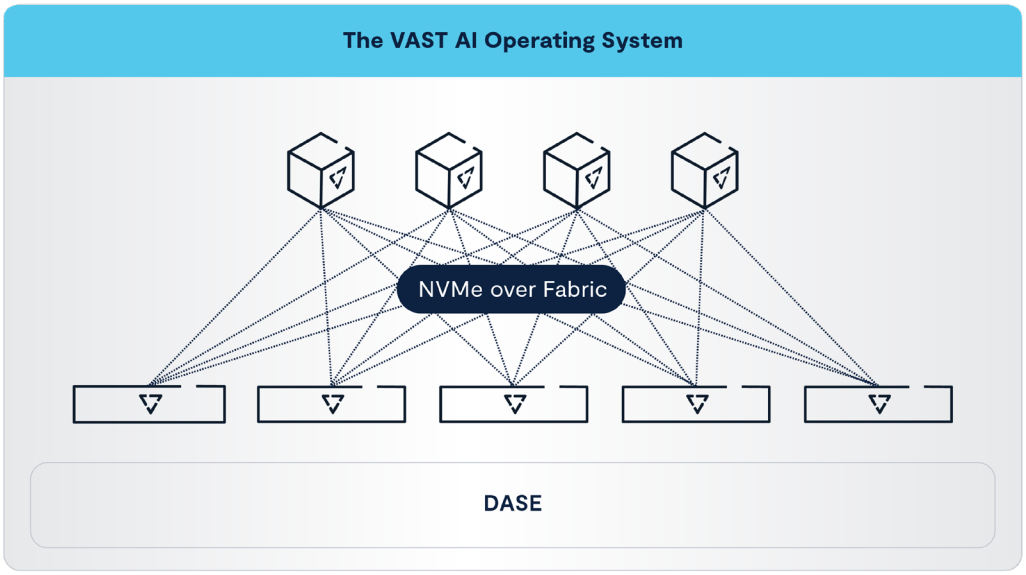
# The VAST DASE Architecture

VAST DataBase is built on VAST's DASE architecture, which fundamentally redefines data architecture. DASE decouples the compute tier from the storage tier, eliminating bottlenecks and allowing both to scale independently. Every compute node can access the full global dataset with single-millisecond latency—even at exabyte scale.

**Stateless Compute Nodes (CNodes):** Containers that run the VAST DataEngine, compute, and query execution but do not own specific data slices. Each CNode connects to all Data Nodes (DNodes) over NVMe-oF, eliminating the east-west traffic that drives latency in shared-nothing architectures. CNodes eliminate the need for local caches or shard coordination, enabling true linear scalability without complex rebalancing.

**Highly Available Data Nodes (DNodes):** All-flash data enclosures that combine Storage Class Memory (SCM) for ultra-fast writes and metadata with high-capacity flash for data. SCM absorbs writes globally at low latency, eliminating hotspots before data is durably striped across SSDs.

**Similarity-Based Data Reduction:** Inline compression, deduplication, and similarity reduction cut physical storage by 3:1 or more—even for embeddings or pre-compressed files. Every write is examined in memory to identify common patterns across the dataset.




This design delivers three key benefits:

1. **Linear scalability:** Add CNodes to increase throughput or query capacity instantly, with no repartitioning. All nodes share the load for all data.
2. **Simplified management:** No shard ownership, resharding, or distributed transaction coordination. Writes complete without CPU-to-CPU coordination, ensuring resilience even at extreme scale.
3. **Global efficiency:** Similarity-based reduction lowers storage costs, while SCM ensures sustained write performance across diverse workloads.

For engineers, this means no rebalancing headaches. For enterprises, it means predictable performance and effortless scale to exabytes.

## Why Cloud Architectures Fall Short

Most cloud data lakes and warehouses are built on blob storage, designed for durability and capacity—not low-latency access. This makes them inherently inefficient for real-time analytics and AI, where every query incurs high-latency storage calls.

VAST takes a different approach. With a flash-first design and the DASE architecture, every compute node accesses all data directly over NVMe with single-millisecond latency—delivering performance cloud-blob-based systems cannot match.

## VAST DataBase: Breaking Tradeoffs

For decades, enterprises have faced a false choice: systems that excel at transactions but struggle with analytics, or systems that deliver analytics at scale but cannot handle transactional workloads. Bridging these worlds required duplicate infrastructure, constant data movement, and complex ETL pipelines that slowed insights and inflated costs.

The VAST DataBase ends this compromise. It is the first system to bring transactional consistency and analytical scale together on a single foundation.

Incoming transactional data is written in a row-based format directly to the high-speed SCM buffer, ensuring near-instant ingestion with full ACID compliance. In the background, CNodes asynchronously reorganize this data into a columnar format stored on lower-cost flash, where it is optimized for analytical queries. This seamless conversion process eliminates the need for separate systems or ETL pipelines, while guaranteeing integrity and reliability across both workload types.

Instead of running separate databases, warehouses, and pipelines, organizations can consolidate onto one system where transactions and analytics coexist without compromise. The result is **real-time AI and analytics on live data, simplified operations, and faster time to insight.**

In practice, VAST DataBase delivers the speed of a transactional database, the scale of a data warehouse, and a trusted, future-ready foundation for AI workloads—all without ETL-driven delays or infrastructure sprawl.

## Integrated AI-Native Capabilities

Unlike bolt-on systems, VAST DataBase is built for AI from the ground up:

**Integrated vector store:**
Stores embeddings natively alongside structured and unstructured data, scaling to trillions of vectors without a separate vector DB.

**Real-time event analytics with VAST Event Broker:**
Converts streams from the VAST Event Broker into queryable tables, enabling analytics on both live and historical data—without ETL.

**Unified governance:**
Provides role-based, fine-grained access (FGAC) controls applied consistently across raw data, structured tables, and vectors. This ensures secure and compliant analytics and Retrieval-Augmented Generation (RAG) workflows.

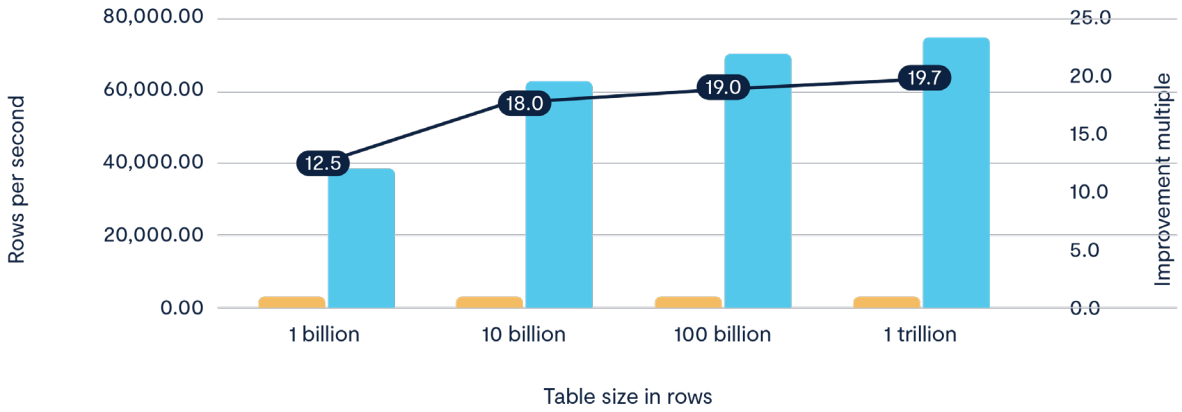
By delivering these capabilities natively, VAST DataBase eliminates the need for fragmented pipelines, reducing both latency and operational complexity.

# Unprecedented Performance

The VAST DataBase is designed to deliver the world's fastest performance beyond what traditional architectures can provide. By eliminating bottlenecks and enabling in-place execution, it accelerates both transactional and analytical workloads at exabyte scale.

Performance by Design: VAST's DASE architecture and 32KB element store minimize read amplification, reduce memory pressure on compute, and cut bandwidth between storage and compute. These efficiencies reduce CPU time and deliver consistent single-millisecond latency with up to 20x faster queries and 40x faster transactions than legacy systems.

### Rows/s on 0.001% filter, unsorted data, AWS/Iceberg, VDB




### Data rate on 0.001% filter, unsorted data, AWS/Iceberg, VDB

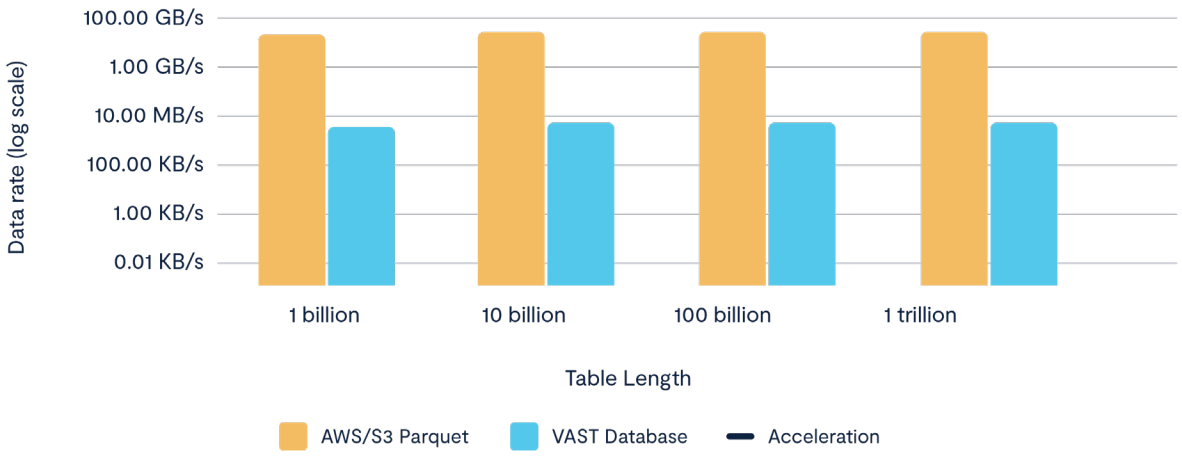



**Table: VAST DataBase Performance vs. Iceberg with .001% filtering is 50x faster**

#### Optimized Query Engine

The native VAST Query Engine runs directly on CNodes, supporting both SQL queries and vector search with in-place execution—processing data where it lives for maximum efficiency. Arrow Database Connectivity (ADBC) API compliance ensures seamless integration with the Apache Arrow ecosystem, enabling zero-copy data transfers and a standard, high-performance API. This architecture delivers up to 20x faster queries and 40x faster transactions than legacy systems, with consistent single-millisecond latency for real-time performance at scale

### Sorted Tables for Logarithmic-Time Search (LogN)

VAST DataBase also delivers logarithmic-time search with sorted tables, which dramatically accelerates queries on massive datasets and removes the need for complex partitioning. Sorted tables minimize full scans by organizing data into efficient sequential I/O structures. The result is indexing-level performance with storage efficiency intact—benchmarks show up to 95x faster parallel point queries and 3x faster table load times compared to Iceberg.

### Native Ecosystem Integration with Trino and Spark

VAST DataBase runs Trino and Spark directly on its stateless CNodes, eliminating the need for separate clusters, storage management, and complex pipelines. Deployment and scaling are handled automatically. With engines co-located on VAST, queries gain direct NVMe access to all data—removing network overhead and delivering both operational simplicity and performance gains.

This combination allows enterprises to shift from batch-oriented analytics to truly interactive AI workloads at scale.

## Business and Operational Benefits

Beyond technical innovation, VAST DataBase delivers measurable business value:

**Operational simplicity**

Self-maintaining design eliminates manual tasks like partitioning, vacuuming, and compaction.

**Cost efficiency**

Flash-optimized architecture and similarity-based compression reduce storage by 3:1 or more, lowering both CapEx and OpEx.

**Future-proofing:**

Unifies structured, unstructured, and vector data, providing a single foundation for emerging multimodal AI.

By collapsing decades of architectural sprawl into a single platform, VAST DataBase enables organizations to innovate faster while reducing risk and cost.

## Conclusion

The VAST DataBase provides a unified, high-performance foundation for enterprise AI and analytics. It collapses complexity, removes silos, and delivers real-time performance at exabyte scale.

For data leaders, it represents an opportunity to replace fragmented, legacy systems with a single, future-ready architecture that finally makes enterprise AI operational at scale.

**Further Resources:**

- VAST DataBase Overview (Solution Brief)
- Why VAST DataBase is Faster and Easier Than Iceberg for AI & Analytics (White Paper)
- Beyond Iceberg: How VAST DataBase Delivers 20x Faster Analytics & AI Performance (Webinar)