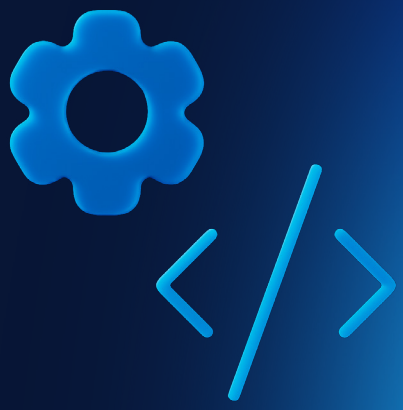
VAST DataEngine

Real-Time Compute Fabric for Data & AI



Modern AI and analytics pipelines break under the weight of slow ETL, fragmented orchestration, and constant data movement. VAST DataEngine changes the model by bringing compute to the data. As part of the VAST AI Operating System, it makes it possible to create and execute event-driven workflows using triggers, serverless functions, and containerized compute engines directly on VAST—powering real-time, governed pipelines without operational complexity.

How It Works

Programmable Compute Where Data Lives

The VAST DataEngine provides an elastic compute fabric that facilitates execution directly on VAST. This allows you to execute integrated engines like Trino and Spark on CNodes, and orchestrate custom Python functions and containerized services for in-place processing—all without moving data or managing external clusters. This consolidated environment simplifies operations, accelerates analytics, and gives teams a flexible foundation for everything from log enrichment to SQL dashboards and AI pipelines. With compute and data on the same platform, you get faster results and less operational overhead.

Automated, Event-Driven Pipelines

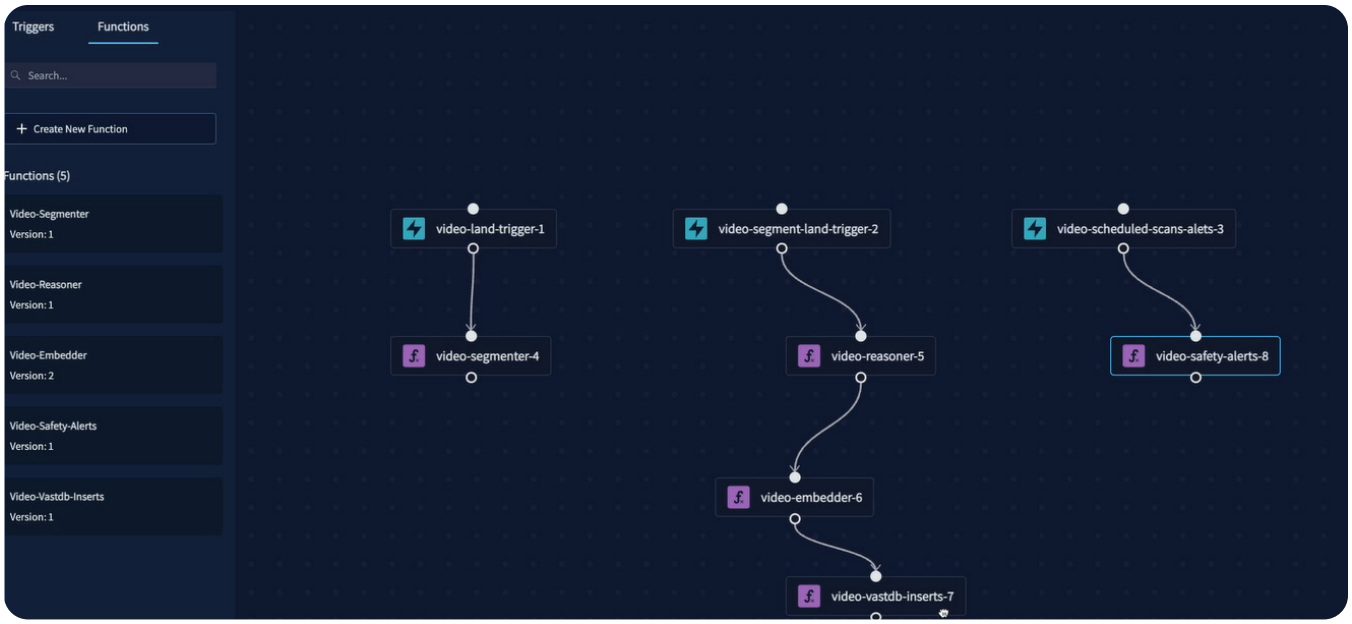
With DataEngine, data becomes active the moment it arrives. The VAST Event Broker can detect file, object, and streaming events and triggers downstream workflows automatically—no job schedulers, cron management, or fragile ETL required. Whether enriching metadata, loading new records into VAST DataBase, or initiating vectorization and RAG pipelines, DataEngine keeps intelligence flowing continuously and in real time.

About Us

VAST is the AI Operating System for simplicity at scale. VAST engineered the first new scale-out architecture in 20 years to take you beyond the conventional limits of scale, resilience and cost.

Looking to simplify your data center and unlock insights from all of your data? Contact us at

hello@vastdata.com



Flexible Serverless Functions for In-Place Python Execution

Developers can author lightweight Python functions that execute in containerized environments, orchestrated by the DataEngine. These functions can route, parse, transform, or enrich data, or initiate inference workflows—without complex scheduling tools or the need to move data into external compute clusters. Logic is versioned, governed, and reusable, enabling teams to standardize processing and accelerate development. While engineers can work in a code-first model using their preferred IDE, DataEngine also provides an intuitive low-code interface that allows users to visually build, manage, and update pipelines. This blend of programmable flexibility and visual simplicity makes it easy to deliver production-grade automation across streaming, batch, and AI workflows.

Built-In Workflow Observability

The VAST DataEngine provides full observability and auditing across every pipeline, giving teams complete insight into how data is triggered, processed, and transformed. Each workflow step is tracked, making it simple to see who accessed data, which functions operated on it, and how it progressed through the system. Execution metadata and performance telemetry are surfaced directly within the platform for rapid diagnostics and optimization. This unified operational view eliminates the need to stitch together logs from multiple services and ensures that data processing remains trustworthy, compliant, and easy to manage throughout its lifecycle.

Unified Pipeline Automation: One Platform for Streaming, Batch, SQL & Vector

DataEngine orchestrates multi-modal pipelines—spanning streaming ingestion, metadata enrichment, batch transformations, SQL, and vector search—without external tools or workflow engines. With VAST InsightEngine on the VAST DataEngine, teams can automate end-to-end RAG and vectorization pipelines. Using Trino and Spark on CNodes, analysts can execute real-time SQL without copying data.

Toolkits to Operationalize AI: Engines to Activate Data, Vector Search, and Agents

DataEngine powers specialized VAST Engines:

- **InsightEngine** to accelerate AI deployment by eliminating integration steps & data movement, while enabling automated and governed AI pipelines
- **AgentEngine** for creation, management and orchestration of AI agents
- **SyncEngine** to discover, onboard, and prepare diverse data from virtually any source

Together, they unify ingest, enrichment, vector search, and agent execution into a single governed platform, enabling faster AI deployment without fragile integration.

Why VAST

Unified Data + Compute

Bring processing to the data without clusters or copies.

Real-Time Workflows

Trigger pipelines instantly as data changes.

Programmable + Flexible

Python functions + containers run anywhere.

Exabyte Scale

DASE architecture delivers predictable scale as data grows.

Governed + Observable

Identity-aware access + full pipeline visibility.

AI-Ready Foundation

InsightEngine framework for ingest, vectorization, and RAG.

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

11/2025