

InsightEngine

Unlocking Real-Time Document Insights

NEW

VAST

InsightEngine

In today's data-driven world, extracting insights from vast unstructured document repositories is a challenge. VAST's Document Retrieval solution, powered by VAST InsightEngine, enables enterprises to efficiently search, retrieve, and summarize massive datasets in real time. By leveraging NVIDIA Inference Microservices (NIMs), organizations can quickly access, understand, and act on their data with precision.

Real-Time Document Search and Augmentation

The VAST Document Retrieval solution, built on VAST InsightEngine, is a customizable framework that leverages NVIDIA Inference Microservices (NIMs) to enable intelligent search, retrieval, and summarization. InsightEngine automates workflows to support real-time AI pipelines, orchestrating AI data processing to LLMs and NIMs. This enables enterprises to extract insights from complex, multi-format documents with speed and accuracy. It structures unstructured text into relevant, actionable summaries, making it ideal for industries like legal, finance, healthcare, and regulatory compliance.

Example Use Cases:

Legal, Compliance

Instantly find key executive statements or discussions across company-wide video archives.

IT & Engineering

Technical documentation, troubleshooting, best practices from scattered knowledge bases.

Customer Support

Provide instant, AI-powered responses from support documents, ticketing systems, and chat logs.

Finance & HR

Quickly locate relevant video evidence by searching for spoken phrases, objects, or detected events.

End-to-End AI Security & Governance with an Integrated DataBase and DataStore

VAST uniquely preserves access control throughout the entire document retrieval process by seamlessly integrating file system permissions with vector search capabilities. Because only VAST combines a DataStore and a DataBase in a single platform, it ensures that users can only retrieve and interact with vectors derived from data they are explicitly authorized to access—something no other solution can do. This real-time enforcement of identity-based policies maintains strict security and compliance, allowing organizations to confidently leverage AI and analytics without compromising data privacy or governance.

Why VAST?



Unified Data Management

Consolidate structured and unstructured data and all analytics workloads for seamless access and faster insights.



Scalability

Easily scale to exabyte-level datasets with high performance as data grows.



AI-Ready Infrastructure

Enable AI-driven workflows and real-time data analysis with integrated GPU-accelerated I/O.



Security and Compliance

Ensure compliance with AES-256 encryption, tenant-managed keys, and strong data security measures.



Collaboration

Support secure, multi-institution collaboration with a global namespace shared across locations.



Integrated AI Workflows

Streamline AI deployment with pre-packaged workflows and seamless integration of storage, compute, and AI pipelines.

Scalable and Secure AI-Driven Solutions

Designed to handle exabyte-scale datasets on VAST's Disaggregated Shared Everything (DASE) architecture, VAST InsightEngine ensures consistently high performance as data volumes grow. By unifying structured and unstructured data within a single platform, it streamlines analytics workflows and eliminates data silos.

AI Deployment for Document Intelligence

Deploying AI-powered document search and retrieval has traditionally been complex and resource-intensive. VAST InsightEngine simplifies this process by providing a fully integrated platform that combines high-performance storage, compute, and trigger and function automated AI pipelines. Leveraging NVIDIA's Nvidia Inference Microservices (NIMs), the platform automates data embedding and retrieval, eliminating manual intervention and accelerating the time-to-insight. This streamlined approach enables organizations to leverage real-time search and analysis capabilities effortlessly.

How It Works

VAST InsightEngine automates ingestion and retrieval of unstructured document data using an event-driven architecture powered by VAST DataEngine. Data flows into VAST DataStore, where DataEngine triggers initiate processing workflows. AI pipelines orchestrate vectorization, embedding, and chunking via NVIDIA NIMs, storing vector embeddings in VAST DataBase for scalable, high-speed retrieval. Fine-grained access control ensures unified governance across raw and vector data, delivering unprecedented AI security and compliance.

About VAST Insight Engine

VAST InsightEngine transforms unstructured data into real-time AI-driven decisions. It instantly generates AI-ready vector embeddings and automates complex AI data pipelines, managing exabyte-scale workloads without performance loss. A unified security model enforces fine-grained access controls across raw and vector data, securing every stage of the AI pipeline. Built for secure, scalable AI infrastructure, VAST InsightEngine turns unstructured data into actionable insights, driving efficiency, compliance, and competitive advantage.

Trusted by Industry Leaders

PIXAR

servicenow

Chan
Zuckerberg
Initiative

agoda

Key Benefits



Instant Answers from Any Document

No need to manually search across multiple repositories.



Real-Time, Up-to-Date Knowledge

Automatically processes new and updated documents for immediate availability.



Enterprise-Grade Security & Compliance

Ensures strict adherence to identity-based access controls and data governance policies.



Seamless AI Integration

Can embed into enterprise applications, intranet portals, and virtual assistants via APIs.

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 unlock insights from your data?

Contact us at

hello@vastdata.com