

Teradata × Unstructured

Unlock Scalable, Multi-Modal Intelligence Across Your Enterprise



The challenge

Enterprises are rapidly adopting generative and agentic AI, but most struggle to operationalize the data that matters most. Text, documents, images, audio, video, and other unstructured content now account for most enterprise information—yet this data remains siloed in disconnected systems, invisible to the AI models built to drive business value.

Standalone vector databases introduce complexity, data movement, and rising operational costs, while cutting off access to the governed structured data and enterprise controls organizations depend on. The result: AI pilots that stall before reaching production.



The partnership

Teradata and Unstructured have joined forces to solve this problem end to end. Unstructured—ranked #1 on the SCORE Benchmark for real-world enterprise document parsing—automatically extracts, enriches, chunks, and embeds content across 60+ file types from 20+ enterprise sources, including SharePoint, Confluence, S3, and Slack.

Those embeddings are stored, governed, and queried natively inside Teradata alongside structured data, under the same security and compliance model your organization already runs. Together, the two platforms deliver a complete AI data foundation—from raw enterprise content to production-ready AI—without duplicating infrastructure, moving data outside your environment, or introducing a new vendor relationship.

Organizations aren't missing a vector database.

They're missing a complete data foundation for AI.



The organizational impact you can't afford to ignore

Most enterprise AI projects stall not because the models are wrong, but because the data foundation is incomplete. Teradata Enterprise Vector Store and Unstructured together eliminate that gap—connecting the 80% of enterprise knowledge that has been invisible to AI with the governed structured data your organization already trusts.



Capabilities designed for enterprise AI

Teradata and Unstructured deliver a seamless pipeline from raw enterprise content to production-ready AI. Here's how the partnership works, end to end.

Step 1: Ingest—Unstructured connects to your content, wherever it lives

Unstructured connects to 20+ enterprise sources and handles 60+ file types, including PDFs, images, audio, video, and spreadsheets. Unstructured automatically extracts, enriches, chunks, and generates embeddings from your content. Processing runs inside your network so no data leaves your environment, and the output lands directly in Teradata with no custom code or pipeline engineering required.

Step 2: Store and search—Teradata governs embeddings alongside your structured data

Embeddings generated by Unstructured land natively in Teradata Enterprise Vector Store, sitting alongside your structured data in a single governed platform. Teradata then delivers hybrid search—combining vector similarity, lexical search, metadata filtering, and structured SQL in one query—so AI agents can simultaneously retrieve insights from documents and data tables without any synchronization layer. Built on Teradata's proven MPP architecture, the system scales linearly to billions of vectors with consistent performance and predictable cost.

Step 3: Query - Fusion or hybrid search, combining lexical, semantic, and metadata search

Because Unstructured's document embeddings land directly inside Teradata, AI agents can run a single governed query that joins structured enterprise data with embedded document content at the same time. An agent can reason over a customer record and a policy document together, or a transaction history and a compliance filing—in one call, with full auditability. No pipeline to maintain. No separate integration layer.

Step 4: Govern—One security and compliance model, deployed anywhere

Unstructured processes content in VPC or on premises so sensitive data never leaves your environment. Once in Teradata, vectors, structured data, and document embeddings all share the same RBAC, audit trails, encryption, and compliance frameworks your organization already runs. Teradata deploys identically across cloud, on-premises, and air-gapped environments—critical for regulated industries where data sovereignty is nonnegotiable. Every AI response is fully traceable back to its source document.

Step 5: Build and deploy—AI Studio powers agents and RAG workflows on top

The governed, search-ready data foundation created by Unstructured and Teradata feeds directly into Teradata AI Studio, where teams build and deploy RAG pipelines, agentic workflows, and intelligent applications. Through the Autonomous Knowledge Platform, Teradata orchestrates the entire experience—from ingestion through retrieval to agent response—so developers focus on building AI value rather than managing infrastructure. Native LangChain integration, SQL and Python SDKs, and support for models from NVIDIA, OpenAI, Hugging Face, and BYOM ensure the platform meets teams where they are.

Use cases across industries

INSURANCE

Claims adjudication

AI can access structured claims history but is blind to the damage photos, policy PDFs, and correspondence that contain the actual evidence. Adjusters continue to work across disconnected systems.

Outcome: Adjudication time reduced; full compliance audit trail; no data movement outside the platform.

FINANCIAL SERVICES

KYC (Know Your Customer) and AML (Anti-Money Laundering)

Analysts manually parse identity documents, filings, and foreign-language correspondence to cross-reference against transaction records. Reviews take weeks and create compliance exposure.

Outcome: Document review from weeks to hours; full document-to-decision audit trail for regulators.

HEALTHCARE

Clinical decision support

Clinical notes, radiology images, and audio dictations are inaccessible to AI. Compliance requirements mean data cannot move to a cloud-only vector store, stalling AI initiatives entirely.

Outcome: Clinical AI unblocked; HIPAA compliance maintained; on-premises deployment; multi-modal in one platform.

TELECOMMUNICATIONS

Network operations

Engineers troubleshoot outages by manually searching runbooks, incident reports, and vendor manuals. Resolution knowledge lives in documents, separated from the structured telemetry data in Teradata.

Outcome: Faster mean time to resolution; institutional knowledge preserved and queryable; reduced reliance on senior engineer availability.