White paper

Teradata Autonomous Knowledge Platform

From Insight to Autonomous Outcomes at Enterprise Scale

teradata.

Table of Contents

Executive summary

Enterprises are entering a new phase of the intelligence era: most organizations can build AI, but far fewer can operationalize it as a trusted, repeatable business capability. As AI agents move from isolated pilots into continuous, business-critical operations, they generate sustained high-concurrency demand, require reliable access to governed business context, and must operate across changing workload, deployment, security, residency, and economic requirements.

The Teradata Autonomous Knowledge Platform is designed for this reality. It connects governed business context, data and analytics, AI development, intelligent execution, workload-fit compute, governance, and deployment choice so enterprises can move from experimentation to trusted production AI. It is available through Teradata Cloud, Teradata Factory, hybrid architectures, and sovereign deployment patterns for organizations that require stronger residency and operational control.

This white paper introduces the autonomous knowledge category and explains how Teradata helps operationalize enterprise AI from trusted context to intelligent action. It details AI Studio as the unified development and management environment; Tera as the intelligent execution layer; Tera Context Engine as an open, neutral context and orchestration layer; Active and Elastic Compute for workload-fit cloud scale; a connected data foundation; Enterprise Vector Store and multi-modal knowledge; AgentStack; AI Services; and controlled deployment options including Teradata Factory and Teradata Sovereign Cloud.

Before exploring the architecture, deployment models, and capabilities that make this possible, it's important to clarify what Teradata means by autonomous knowledge—and how this concept differs from traditional analytics platforms and first-generation AI systems. This foundation provides the context for understanding why agentic workloads require a fundamentally different platform approach.

A practical lens for buyers

The way organizations evaluate and buy technology has changed. Rather than assessing tools in isolation, leading teams now focus on the measurable impact those tools deliver—and they want to see proof quickly, not at the end of a long pilot.

These decisions also rarely sit with IT alone. When a platform touches how work gets done across the organization, stakeholders across functions need to be part of the conversation from the start.

That's why it matters to begin with the outcome you're trying to achieve—whether that's accelerating a key process, improving decision quality, or reducing operational costs—and then evaluate whether the platform can reliably deliver it. The questions worth asking:

- How does this connect to what we're already running?
- How does it perform at our scale?
- What does it actually cost to get there?

The enterprise intelligence inflection point

From dashboards to continuously operating systems

The last two decades of enterprise analytics were primarily human-driven: business users and analysts queried data, reviewed dashboards, and acted through separate operational systems. The next decade will be increasingly machine-driven: agents will continuously monitor signals, generate hypotheses, run high volumes of queries, and take actions across systems.

Agents don't sleep: Why platform requirements change

Agentic workloads differ from human workloads in predictable ways: they're always on, API first, and context reasoning by default, and they produce sustained concurrency across mixed workload types. As agents evolve from basic Q&A to outcome-based execution, the workload impact grows sharply—driving higher query volumes, broader data access patterns, and stronger requirements for cost control and governance.

Why most agents fail in production

Many early agent deployments fail or stall when scaled because production AI requires more than model access. Agents need trusted business context, governed knowledge, lifecycle controls, workload-fit infrastructure, and an operating model that can sustain always-on performance, cost, security, and compliance. A platform designed for autonomous knowledge must address these requirements systematically—grounding agents in enterprise context, governing how knowledge is used, and adapting execution to workload demand.

What changes when agents become the primary data consumers

- Always-on operation (24/7) replaces episodic usage (8/5)
- Concurrency becomes a baseline requirement—not an edge case
- Mixed workloads (SQL + vector + ML + multi-modal) run side by side
- Context is required for every decision (semantics, lineage, policy, KPIs)
- Governance must be embedded, because agents can act—not just answer

The Teradata Autonomous Knowledge Platform

What the platform is

The platform is designed to activate autonomous enterprise intelligence and help organizations operationalize AI at scale. It brings together governed business context, data, analytics, AI workflows, intelligent execution, workload-fit compute, governance, and deployment choice so organizations can progress from fragmented experimentation to repeatable production outcomes.

Core architectural components

At a high level, the platform combines AI Studio as the unified development and management environment; Tera for intelligent execution; Tera Context Engine for governed enterprise context and context orchestration; Active and Elastic Compute for workload-fit execution; a connected data foundation across structured, unstructured, vector, and open-format data; and deployment models that span managed cloud, on-premises, hybrid, and sovereign requirements. These capabilities are complementary. For instance,Tera Context Engine can serve multiple AI applications and agents across heterogeneous environments, while Tera acts as the platform's intelligent execution system, translating governed context and business intent into optimized execution decisions across workloads, agents, and compute resources. Together they create a stronger path from enterprise understanding to trusted action.

Defining autonomous knowledge

What autonomous knowledge means

Autonomous knowledge is the ability of an enterprise platform to transform raw structured, unstructured, and multi-modal data into trusted, governed understanding. It captures meaning (semantics), relationships, lineage, and constraints so that knowledge can be reused safely across people, applications, and agents.

Autonomy at two levels: System and knowledge

In the Autonomous Knowledge Platform, autonomy operates across two complementary layers. System-level autonomy manages execution—predicting workload behavior, placing and sizing compute, and optimizing performance, cost, and data locality. Knowledge-level autonomy enables continuous creation and evolution of enterprise knowledge—leveraging agents to curate and refine data products and context over time.

Autonomous by design

The Teradata Autonomous Knowledge Platform unifies system-level autonomy with enterprise-grade knowledge so AI agents can operate continuously, make trusted decisions, and turn insight into governed action at scale—with minimal manual intervention.

AI Studio:
The unified platform interface

The interface for every platform and every persona

AI Studio is the place where creators build, activate, and manage AI outcomes at scale. It provides a single, consistent interface across cloud, on-premises, and hybrid deployments—so teams do not need separate toolchains for different environments.

What AI Studio provides

AI Studio brings together the capabilities required to create AI outcomes, including agent development, model operations, analytics and machine learning (ML), Python compute, and accelerated compute. It's designed to unify exploration, development, deployment, and governance so teams can move from prototypes to production with repeatability.

These capabilities are deliberately sequenced to build on one another, forming a complete enterprise AI platform that delivers tangible customer value at every layer. The foundation is enhanced Python compute, providing an industry-standard Jupyter Notebook experience with native Teradata machine learning connectivity. This allows data engineers and data scientists to work directly with governed Teradata data using Python—without exporting data to external platforms that may not scale, perform, or meet enterprise security and privacy requirements. Building on this foundation, conversational AI democratizes access to data and insights through a natural language interface, enabling self-service analytics for nontechnical users while ensuring the accuracy, repeatability, and governance required for enterprise grade business intelligence (BI).

Through Teradata AgentStack, agent capabilities introduce autonomy through procode Python orchestration, leveraging frameworks such as LangChain and LlamaIndex alongside ModelOps to industrialize the full AI lifecycle, including deployment, monitoring, and retraining. Partner integrations such as WisdomAI and Karini further enrich AI Studio—WisdomAI powers agentic BI experiences while Karini accelerates agentic AI use cases, all fully grounded in Teradata's governed data foundation.

Finally, Enterprise Vector Store connects unstructured knowledge with structured enterprise data, enabling retrieval augmented generation and agentic workflows at scale. Taken together, these capabilities position AI Studio as a unified, governed AI platform—rather than a collection of disconnected tools.

Example workflows enabled by AI Studio

A data scientist prototypes a retrieval-augmented agent, connects it to governed enterprise knowledge, tests it against a repeatable evaluation set, then deploys and monitors it—using one workspace.

An analyst uses natural language to explore governed data products and publish an outcome to an operational workflow, while platform teams preserve policy enforcement and lineage.

A unified workspace

AI Studio brings together:

- Humans
- Apps and agents
- AI marketplace
- ModelOps
- AgentStack
- Generative AI and model services
- Analytics and ML
- Python compute (scale-up and scale-out)
- CPU/GPU accelerated compute

Tera: The agentic coworker for enterprise data work

Intelligent execution for enterprise AI

Tera is the agentic coworker for enterprise data work. It helps teams move from AI intent to business outcome faster by combining enterprise context, specialized skills, tools, data, models, and intelligent execution in a single governed experience. Natural-language access and Teradata expertise make sophisticated data and AI work accessible to more users while improving the value of each model interaction.

While many AI systems focus primarily on generating responses, enterprise work requires execution. Tera goes beyond chat by interpreting intent and coordinating the appropriate skills, tools, data, context, and models needed to complete complex, multi-step work across analytics, AI development, and platform operations.

Whether exploring data, building models and AI applications, generating code, or operating the platform, users can work through purpose-built experiences such as Analyze, Code, and Console instead of navigating disconnected tools. An intelligent agent harness coordinates execution, specialized Teradata skills package reusable expertise, MCP connectivity extends workflows across enterprise tools, and model choice gives organizations flexibility as their AI architecture evolves.

Why Tera matters

- **More value per model call:** Generic prompting burns tokens; smart routing spends them where they pay off.
- **More trustworthy outcomes:** Generic assistants guess without enterprise context; Tera is grounded in it.
- **AI where data work happens**: Others bolt AI onto a separate stack; Tera runs where the data already lives.
- **More people get value from AI:** No deep Teradata specialization required; Tera democratizes data access.
- **Faster AI outcomes:** Others help you generate answers; Tera helps you complete work.
- **Freedom of choice:** Closed platforms lock you to one model or one deployment mode; Tera lets you choose.

Tera Context Engine: Open, neutral context and orchestration for trusted enterprise AI

Enterprise AI cannot produce trustworthy outcomes from data access alone. It needs business meaning, relationships, policies, operating context, and a governed way to make that knowledge available during AI reasoning and execution. Tera Context Engine addresses this requirement by transforming distributed enterprise and industry information into reusable, governed context that applications and agents can use consistently across use cases.

Tera Context Engine is an open, neutral context and orchestration layer for enterprise AI. It is designed to operate above and across databases, data platforms, pipeline engines, catalogs, knowledge repositories, models, applications, and AI agents—including Teradata and third-party environments. It combines Industry Knowledge Models, semantic relationships, governance, lineage, quality, access controls, policy, auditability, and deterministic grounding so AI can work from governed business meaning rather than unsupported assumptions.

Tera Context Engine also supports a continuous knowledge and execution loop. Enterprise metadata, approved operational signals, AI responses, corrections, and outcomes can enrich reusable context over time while human governance and expert oversight remain in place. This helps organizations build context once, govern it consistently, and reuse it across platforms, models, teams, applications, and agents—reducing repeated context assembly, retrieval overhead, ambiguity, and unnecessary model interactions.

Within the Autonomous Knowledge Platform, Tera Context Engine helps create AI-ready data products by packaging enterprise data with the context AI needs to understand and act on it: domain meaning, semantic relationships, policy, lineage, memory, observability signals, search, and prediction. Industry Knowledge Models accelerate this process by contributing expert-authored terminology, relationships, policies, processes, metrics, and operating conditions for complex industries.

Just as important, Tera Context Engine reduces the context fragmentation that emerges when semantics, vectors, features, memory, governance, and observability are rebuilt independently for every AI use case. Its neutral approach complements existing enterprise investments rather than requiring Teradata to be the underlying data platform. When used with Tera, governed enterprise knowledge can inform intelligent execution; neither product is dependent on the other, and each provides independent value.

Why Tera Context Engine matters

- **Turns data into AI-ready knowledge:** Packages enterprise data with business meaning, lineage, policy, and relationships so agents can use it reliably.
- **Improves trust and explainability:** Grounds AI outputs in governed context so results are more consistent, traceable, and defensible.
- **Reduces repeated work:** Creates reusable AI-ready data products so teams do not need to rebuild context for every use case.
- **Supports continuous learning:** Uses real enterprise query and workflow signals to enrich the knowledge layer over time.
- **Connects context to action:** Works with Tera and the broader platform to move from governed knowledge to trusted execution, while keeping Tera Context Engine positioned as platform-wide.

Workload-fit cloud compute: Active + Elastic Compute

Active Compute for sustained, predictable workloads

Active Compute is designed for sustained, predictable, mission-critical workloads that require persistent capacity, high concurrency, stable service levels, and continuous operation. It provides the always-on foundation for production analytics and agentic workloads where availability and predictable execution are paramount.

Elastic Compute: Scale smarter as workloads change

Variable demand should not become permanent capacity. Elastic Compute adds isolated, on-demand capacity around dynamic workloads in smaller, node-based increments within configured limits. It is designed for object-store analytics, ELT and batch, AI/ML preparation, agentic workloads, experimentation, ad hoc analysis, and other demand patterns that should scale independently without permanently expanding the production core.

Elastic compute with open table formats (OTFs)

Elastic Compute combines four capabilities: granular incremental scaling; workload isolation that protects SLA-bound production services; open data access to supported table formats such as Iceberg and Delta in object storage without unnecessary data movement or disruptive re-platforming; and managed workload operations using policy, scheduling, thresholds, utilization signals, auto-suspend, governance, security, and consumption visibility. The result is closer alignment between workload demand, performance, and consumption economics.

Example: An organization can run continuous fraud scoring on Active Compute while using isolated Elastic Compute for periodic feature engineering against Iceberg data in object storage. Production remains protected, variable capacity is consumed only when needed, and both workload patterns operate within a governed Teradata Cloud environment.

Workload placement guidance

Use always-on compute for:

- Production analytics
- SLAs
- Continuously operating agents
- Latency-sensitive workloads

Use elastic compute for:

- Exploration
- Feature engineering
- Large-scale transformation
- Background training
- Ad-hoc workloads

It's one platform with two compute modes —no duplicated data or governance gaps.

Connected data foundation

What it is

The connected data foundation is the platform's unified data layer for accessing governed structured, unstructured, vector, semantic, metadata, and open-format data across environments. It supports data on block storage, object storage, and supported open table formats such as Iceberg and Delta, as well as Enterprise Vector Store, helping teams extend analytics and AI without creating unnecessary copies or separate governance silos.

Why it matters

A connected foundation matters because agents require complete enterprise context. If structured, unstructured, and open-format data live in separate systems, organizations pay a tax in data movement, duplicated governance, and operational inconsistency. The connected data foundation is designed to reduce that fragmentation while preserving customer ownership of data and metadata.

To ensure access to both structured and unstructured data in a single system, the connected foundation includes Enterprise Vector Store with vector capabilities designed to operate within the same enterprise governance environment as core data. This supports retrieval augmented generation (RAG), hybrid retrieval approaches, and agent grounding in enterprise truth.

Plus, through Teradata and Pinecone's partnership, users can enable a variety of low-latency use cases for real-time results. Tera Context Engine builds on this connected foundation by turning distributed data, metadata, semantics, and usage signals into governed context that agents can use consistently across workflows.

Example: One query surface across data types

The connected data foundation, along with the integration of Unstructured, is designed to ingest and reason over documents, images, audio, and other content types so agents can operate on the full enterprise information estate. With this knowledge available, a team can run analytics over mission-critical structured data while also accessing OTF data in object storage and unstructured content represented through vectors—without shifting between multiple governance models.

Open by design. Trusted in production.

Organizations increasingly adopt open table formats to avoid lock-in. The connected data foundation pairs openness with enterprise-grade governance (identity, access control, lineage, and policy enforcement) so open data remains usable for production AI and analytics.

Autonomous execution and platform intelligence

Autonomous execution agents

At platform scale, manual tuning becomes a bottleneck. The platform is designed to apply autonomy to execution: observing workload behavior, optimizing placement and sizing, and continuously balancing performance, cost, governance, and data locality.

Why autonomy matters

Autonomy enables continuous operation. It reduces operational overhead and allows teams to focus on outcomes rather than infrastructure decisions—especially as agent workloads grow in volume and complexity.

Deployment choice: Cloud, Factory, Sovereign Cloud, and hybrid

Teradata Cloud

Teradata Cloud is a fit-for-purpose cloud architecture for the agentic era. It combines persistent Active Compute for sustained, predictable workloads with on-demand Elastic Compute for dynamic, variable workloads, allowing each workload to use the compute model that best fits its performance, duration, service-level, and economic requirements. Managed security, governance, availability, workload controls, and consumption visibility provide the operational trust required for production AI.

Teradata Factory

Teradata Factory brings regulated enterprise AI into customer control through a ready-to-run, integrated data warehouse, lakehouse, analytics, AI, and agentic stack on configurable Dell and NVIDIA infrastructure. It is designed for organizations that require private AI, data sovereignty, predictable economics, and control over infrastructure, access, security, governance, and recovery, while bringing AI Studio, models, analytics, and accelerated computing close to governed enterprise data.

Factory is designed so compute, GPUs, storage, networking, capacity, and service levels can be scaled independently by workload. Its deployment model differs from Teradata Cloud in one important respect: Elastic Compute, which provides cloud-style on-demand scaling, applies to Teradata Cloud deployments. Organizations evaluating Factory for on-premises workloads should plan capacity accordingly.

Teradata Sovereign Cloud

Teradata Sovereign Cloud delivers the cloud operating experience within customer-controlled sovereign boundaries. It is designed for governments and regulated enterprises that must determine where data and workloads reside, how they are accessed, and which authorized operators can manage them while maintaining identity, access, privacy, lineage, auditing, policy enforcement, security, continuity, and resilience requirements.

True hybrid operation

Teradata supports hybrid operating models so organizations can place workloads according to business, performance, regulatory, residency, and control requirements. Mission-critical or sensitive workloads can remain on customer-controlled infrastructure while managed cloud resources support global analytics, experimentation, and variable demand—using a consistent platform approach and governance model where supported.

Deployment decision guide

- Choose Teradata Cloud when you need managed operations, elastic innovation, and global scale.
- Choose Teradata Factory when you need predictable economics, sovereignty, private AI, or on-premises modernization.
- Choose hybrid when you need both—one platform operating model across environments.
- Choose Teradata Sovereign Cloud when you need a cloud operating experience with stronger control over data residency, access, operations, and regulatory requirements.

AI Services: Accelerating business outcomes

Why services matter in the agentic era

Many organizations can build AI but struggle to operationalize it in production. AI Services helps customers identify high-value use cases, establish the required context and governance foundation, design and implement AI workflows, and accelerate adoption using platform-native capabilities and proven delivery patterns.

How AI Services complement the platform

AI Services are designed to accelerate time to value and reduce operational and governance risk. They don't replace platform autonomy, but help customers operationalize it through strong solution design, implementation rigor, and adoption enablement.

Possible outcomes

AI Services engages across outcome categories including: operational intelligence (computer vision, process monitoring, anomaly detection); customer and commercial outcomes (promotion analytics, churn reduction, onboarding optimization); and public-sector transformation (large-scale text intelligence, data quality programs, decision support). Engagements are scoped to deliver measurable business impact within defined timelines, using platform-native capabilities and proven delivery patterns.

Partner ecosystem: Extending without fragmentation

Partner-first ecosystem

The platform is designed to work within connected ecosystems. Partnerships help organizations bring flexible choice in model providers, ingestion and processing technologies, and infrastructure acceleration while preserving enterprise governance.

Partner roles and examples

Partners can support a range of capabilities—from processing unstructured documents and integrating agentic frameworks to accelerating GPU-based training and inference. The platform is built to flex across hybrid and multi-cloud environments, fitting naturally within the major cloud ecosystems.

Partner positioning

Partners are integrated within the platform so organizations can work from a consistent governance model and reduce the overhead of managing disconnected tools across vendors.

Trust, governance, and enterprise readiness

Trust built in—not bolted on

Autonomous intelligence is only as valuable as its trustworthiness. The platform is designed to apply security, access control, lineage, and policy enforcement consistently across data, AI workflows, and agentic execution.

Operating autonomous intelligence safely

As agents scale, auditability and compliance become central requirements. A governed foundation ensures decisions and actions remain explainable, traceable, and aligned with policy constraints.

Why Teradata

Teradata's Autonomous Knowledge Platform is designed to operationalize enterprise AI from governed context to trusted action. It connects the workspace, knowledge, data, AI, agent, compute, governance, and deployment layers required to run production AI across managed cloud, customer-controlled, hybrid, and sovereign operating models. The objective is not simply to run models close to data, but to create trusted, reusable knowledge and apply it through intelligent, governed execution with workload-fit economics.

Most enterprises already possess vast amounts of data and analytics capability, yet struggle to convert fragmented pilots into repeatable, governed business capabilities. The challenge is not simply connecting AI to data. It is connecting AI to the governed business context that gives data meaning, then providing the compute, execution, deployment, and operating controls required to act on that context reliably at scale.

In an AI-native enterprise, knowledge doesn't arise spontaneously from queries or dashboards. It's produced through an intentional data model design that encodes meaning, relationships, memory, and learning directly into the platform. Teradata enables this through AI-native data products—a standardized, modular approach for turning enterprise data into trusted, reusable knowledge that agents can discover, understand, and act on within governed enterprise controls.

AI-native data products bring together six tightly coordinated capabilities—each one necessary to produce knowledge at scale:

- **Domain** models core business entities as authoritative sources of truth, with temporal consistency and reuse
- **Semantic** makes meaning and relationships machine discoverable, enabling autonomous navigation and correct joins
- **Memory** captures architectural decisions, prior agent actions, and learned strategies so intelligence persists over time
- **Observability** records outcomes and quality signals, closing the loop between action and learning
- **Search** enables semantic similarity and retrieval across enterprise data
- **Prediction** delivers engineered features and model outputs with point-in-time correctness for trustworthy decisions

Together, these capabilities transform static data into a living knowledge system—one that compounds in value as agents interact with it, learn from outcomes, and share that learning across use cases and domains.

Why fragmented stacks fall short

While many architectures attempt to assemble these capabilities using lake houses, vector databases, feature stores, catalogs, and external memory systems, fragmentation introduces structural limits. Semantic context drifts across systems, features lose temporal integrity, and learning loops break when outcomes cannot reliably feed back into agent memory. Knowledge becomes duplicated, brittle, or ephemeral.

Autonomous decision-making requires cross capability joins at decision time—for example, combining core business entities with semantic context, vector similarity results, engineered features, prior decisions, and outcome signals in a single execution path. This is not feasible when these elements live in separate systems with different governance models, performance characteristics, and latency profiles.

Within Teradata-managed data environments, AI-native data product capabilities can operate together on a single enterprise-grade platform, enabling efficient joins across domain data, semantics, vectors, features, memory, and observability without unnecessary data movement or duplication. Tera Context Engine extends this approach by connecting and governing context across broader third-party enterprise environments as well.

This colocation preserves semantic consistency, governance, and point-in-time correctness while supporting the complex access patterns required by autonomous agents. Teradata's shared-nothing architecture, mature workload management, and analytic performance make it possible to execute these AI-driven joins efficiently —even as data volumes, agent activity, and concurrent workloads scale.

Teradata doesn't merely integrate AI components—it structurally enables the manufacturing of enterprise knowledge at scale.

Critically, this approach builds on what businesses already have. Your existing data, on and accessible to Teradata, is the asset. Organizations can progressively uplift existing data models by introducing semantic discovery and memory alongside current workloads, immediately making data usable by agents and preserving institutional knowledge without disruption.

Over time, as observability, search, and prediction capabilities are added, enterprises evolve toward a true knowledge platform: one where every new AI initiative benefits from prior learning, shared context, and governed reuse.

Conclusion

The future of enterprise intelligence is increasingly autonomous—but the near-term enterprise priority is operationalization. As AI and agents move into continuous production work, organizations need trusted business context, intelligent execution, workload-fit scale, governance, deployment choice, and predictable economics. **Autonomous knowledge brings these capabilities together by turning enterprise data and context into a trusted, reusable foundation that AI can understand, act on, and continuously improve.** The Teradata Autonomous Knowledge Platform helps enterprises move from experimentation to governed production AI—and from insight to autonomous outcomes at scale.

17095 Via Del Campo, San Diego, CA 92127 Teradata.com

teradata.