



News Release

# Teradata Delivers Private AI Innovation in New Offering: Teradata AI Factory

2025-06-24

Purpose-built, on-premises product with NVIDIA comes ready-to-run and delivers trustworthy AI with predictable pricing, security, governance, and compliance

SAN DIEGO--(BUSINESS WIRE)-- Teradata (NYSE: **TDC**) today announced the launch of Teradata AI Factory, an integrated solution that brings Teradata's cloud-based AI/ML capabilities - including predictive, generative, and agentic AI - to secure, on-premises deployments.

AI Factory unifies critical components like data pipelines, algorithm execution, and software infrastructure into a single, scalable system, designed for rapid innovation through private, trusted AI. Its seamless integration of software, hardware, and pre-packaged Teradata and modern third-party tools reduces compliance risks and cost overruns, while simplifying AI operations. When combined with Teradata AI Microservices with NVIDIA and customer-provided NVIDIA GPUs, AI development, including native RAG pipelines, is accelerated. Teradata AI Factory, built with the **NVIDIA Enterprise AI Factory** validated design, is ideal for highly regulated industries like healthcare, finance, and government, as well as any enterprise seeking greater autonomy over its AI strategy.

The AI landscape is being reshaped by global instability and increasing data sovereignty requirements, driving organizations to seek better control over their infrastructure. This need coincides with financial challenges that can be caused by both underutilized GPU investments and unpredictable cloud costs, particularly in enterprise-scale hybrid environments. The growing complexity of AI ecosystems compounds these issues, fueling demand for turnkey solutions that address both governance and cost concerns. On-premises, private AI solutions meet these challenges by offering risk mitigation, improved resource management, and enhanced oversight - enabling trusted business outcomes in uncertain times.

"Market dynamics are increasing buyer interest in on-premises solutions," said Teradata's Chief Product Officer, Sumeet Arora. "Teradata remains the clear leader in this environment, with proven foundations in what makes AI



meaningful and trustworthy: Top-notch speed (performance), predictable cost (resource efficiency), and integration with the golden data record (which may already live on Teradata). Teradata AI Factory builds on these strengths in a single solution for organizations using on-prem infrastructure to gain control, meet sovereignty needs, and accelerate AI ROI.”

According to a Gartner® report from March, “By 2028, more than 20% of enterprises will run AI workloads (training or inference) locally in their data centers, an increase from approximately 2% as of early 2025.” (“How to Determine Infrastructure Requirements for On-Premises Generation AI” by Chandra Mukhyala, Jonathan Forest, Tony Harvey from March 5, 2025)

## Key Benefits

Teradata AI Factory is designed to deliver the complete package for private, trusted AI at scale: the security and control of on-premises infrastructure, cost-effective analytic performance, and seamless integration of hardware and software, including **Teradata’s Enterprise Vector Store** and Teradata AI Microservices, which leverages NVIDIA NeMo microservices for native RAG pipeline capabilities.

- Sensitive data stays within the organization’s infrastructure to reduce risks and compliance issues. Teradata AI Factory reduces the vulnerabilities of public or shared AI platforms, such as data exposure, intellectual property loss, and regulatory non-compliance. It aligns with compliance standards like GDPR and HIPAA, making it ideal for industries where data sovereignty and privacy are essential, and by processing data locally, it enhances security and operational integrity for any organization seeking greater control.
- Manage resources with certainty using the most efficient AI performance. AI Factory maintains Teradata’s leadership in delivering the best performance for the price, thanks to its localized architecture that reduces data movement, lowers latency, and improves efficiency. It enables customers to run AI models on CPUs or accelerate them using customer-provided GPU infrastructure, avoiding unpredictable cloud pricing while maximizing hardware investments. This approach ensures consistent operational costs and scalable private AI innovation with long-term financial sustainability.
- Ease-of-use in a ready-to-run, integrated stack that maximizes AI investments. Seamless integration of hardware and software removes the need for custom setups, enabling organizations to scale AI initiatives faster and reduce operational complexity. Teradata AI Microservices with NVIDIA NeMo delivers native RAG pipeline capabilities for ingestion and retrieval – integrating embedding, retrieval, reranking, and guardrails. This enables natural language querying across all data, delivering faster, more intelligent insights at scale.

## What’s Included

Teradata AI Factory components include:

- AI Platform for Rapid Innovation: Teradata AI Factory, built on the company's IntelliFlex platform, is an on-premises AI solution configured for AI workloads. It includes Teradata Enterprise Vector Store for integrating structured and unstructured data in generative AI applications.
- Software Infrastructure: The new AI Workbench provides a self-service workspace with access to libraries and functions, such as those in **ClearScape Analytics**. Built-in support for model lifecycle management, compliance, and one-click LLM deployment, as well as tools like JupyterHub, ModelOps, Airflow, Gitea, and Devpi, further streamline deployment and reduce friction for developers.
- Algorithm Execution: AI Factory enables predictive and generative algorithm execution at scale with high performance in platform analytics. It connects to customer GPUs via Teradata AI Microservices with NVIDIA, enabling high-performance AI workloads, faster model development, and native RAG processing within Teradata's open ecosystem.
- Data pipelines: Modern data ingestion tools, along with Teradata's internal solutions like QueryGrid, efficiently move and connect data across systems. Support for Open Table Formats (OTFs), object stores, and NVIDIA tools for complex formats like PDFs helps build robust, end-to-end data pipelines.

## Availability

Teradata AI Factory is available now.

## About Teradata

At Teradata, we believe that people thrive when empowered with trusted information. We offer the most complete cloud analytics and data platform for AI. By delivering harmonized data and trusted AI, we enable more confident decision-making, unlock faster innovation, and drive the impactful business results organizations need most. See how at **Teradata.com**.

The Teradata logo is a trademark, and Teradata is a registered trademark of Teradata Corporation and/or its affiliates in the U.S. and worldwide.

## MEDIA CONTACT

Jennifer Donahue

**[jennifer.donahue@teradata.com](mailto:jennifer.donahue@teradata.com)**

Source: Teradata