



News Release

# Teradata Makes Real-World GenAI Easier, Speeds Business Value

2024-10-08

New bring-your-own LLM capability enables Teradata customers to simply and cost-effectively deploy everyday GenAI use cases with NVIDIA AI to deliver flexibility, security, trust and ROI. New integration with the full-stack NVIDIA AI platform delivers accelerated computing.

Teradata (NYSE: **TDC**) today announced new capabilities for VantageCloud Lake and **ClearScape Analytics** that make it possible for enterprises to easily implement and see immediate ROI from generative AI (GenAI) use cases.

As GenAI moves from idea to reality, enterprises are increasingly interested in a more comprehensive AI strategy that prioritizes practical use cases known for delivering more immediate business value – a critical benefit when **84 percent of executives expect ROI from AI initiatives within a year**. With the advances in the innovation of large language models (LLMs), and the emergence of small and medium models, AI providers can offer fit-for-purpose open-source models to provide significant versatility across a broad spectrum of use cases, but without the high cost and complexity of large models.

By adding **bring-your-own LLM (BYO-LLM)**, Teradata customers can take advantage of small or mid-sized open LLMs, including domain-specific models. In addition to these models being easier to deploy and more cost-effective overall, Teradata's new features bring the LLMs to the data (versus the other way around) so that organizations can also minimize expensive data movement and maximize security, privacy and trust.

Teradata also now provides customers with the flexibility to strategically leverage either GPUs or CPUs, depending on the complexity and size of the LLM. If required, GPUs can be used to offer speed and performance at scale for tasks like inferencing and model fine-tuning, both of which will be available on VantageCloud Lake. **Teradata's collaboration with NVIDIA**, also announced today, includes the integration of the NVIDIA AI full-stack accelerated

computing platform, which includes **NVIDIA NIM**, part of the **NVIDIA AI Enterprise** for the development and deployment of GenAI applications, into the Vantage platform to accelerate trusted AI workloads large and small.

“Teradata customers want to swiftly move from exploration to meaningful application of generative AI,” said Hillary Ashton, Chief Product Officer at Teradata. “ClearScape Analytics’ new BYO-LLM capability, combined with VantageCloud Lake’s integration with the full stack NVIDIA AI accelerated computing platform, means enterprises can harness the full potential of GenAI more effectively, affordably and in a trusted way. With Teradata, organizations can make the most of their AI investments and drive real, immediate business value.”

### Real-world GenAI with Open-source LLMs

Organizations have come to recognize that larger LLMs aren’t suited for every use case and can be cost-prohibitive. BYO-LLM allows users to choose the best model for their specific business needs, and according to Forrester, **forty-six percent of AI leaders plan to leverage existing open-source LLMs in their generative AI strategy**. With Teradata’s implementation of BYO-LLM, VantageCloud Lake and ClearScape customers can easily leverage small or mid-sized LLMs from open-source AI providers like Hugging Face, which has over 350,000 LLMs.

Smaller LLMs are typically domain-specific and tailored for valuable, real-world use cases, such as:

- Regulatory compliance: Banks use specialized open LLMs to identify emails with potential regulatory implications, reducing the need for expensive GPU infrastructure.
- Healthcare note analysis: Open LLMs can analyze doctor’s notes to automate information extraction, enhancing patient care without moving sensitive data.
- Product recommendations: Utilizing LLM embeddings combined with in-database analytics from Teradata ClearScape Analytics, businesses can optimize their recommendation systems.
- Customer complaint analysis: Open LLMs help analyze complaint topics, sentiments, and summaries, integrating insights into a 360° view of the customer for improved resolution strategies.

Teradata’s commitment to an open and connected ecosystem means that as more open LLMs come to market, Teradata’s customers will be able to keep pace with innovation and use BYO-LLM to switch to models with less vendor lock-in.

### GPU Analytic Clusters for Inferencing and Fine-tuning

By adding full stack NVIDIA accelerated computing support to VantageCloud Lake, Teradata will provide customers with LLM inferencing that is expected to offer better value and be more cost-effective for large or highly complex models. NVIDIA accelerated computing is designed to handle massive amounts of data and perform calculations quickly, which is critical for inference - where a trained machine learning, deep learning or language model is used to make predictions or decisions based on new data. An example of this in healthcare is the reviewing and



summarizing of doctor's notes. By automating the extraction and interpretation of information, they allow healthcare providers to focus more on direct patient care.

VantageCloud Lake will also support model fine-tuning via GPUs, giving customers the ability to customize pre-trained language models to their own organization's dataset. This tailoring improves model accuracy and efficiency, without needing to start the training process from scratch. For example, a mortgage advisor chatbot must be trained to respond in financial language, augmenting the natural language that most foundational models are trained on. Fine-tuning the model with banking terminology tailors its responses, making it more applicable to the situation. In this way, Teradata customers could see increased adaptability of their models and an improved ability to reuse models by leveraging accelerated computing.

#### Availability

ClearScope Analytics BYO-LLM for Teradata VantageCloud Lake will be generally available on AWS in October, and on Azure and Google Cloud in 1H 2025.

Teradata VantageCloud Lake with NVIDIA AI accelerated compute will be generally available first on AWS in November, with inference capabilities being added in Q4 and fine-tuning availability in 1H 2025.

#### 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](https://www.teradata.com)**.

#### Media Contact

Jennifer Donahue

Teradata

**[Jennifer.Donahue@Teradata.com](mailto:Jennifer.Donahue@Teradata.com)**