White paper

How to advance your analytics

teradata.

Version 2.0

Table of contents

Introduction

Machine learning and artificial intelligence are rapidly becoming foundational technologies across industries, driving innovation and shaping competitive advantage. As a result, organizations are investing heavily in AI/ML capabilities, with global spending continuing to accelerate. Yet, despite this momentum, many companies still face a gap between experimentation and impact. Experts have observed a persistent "productivity paradox," where the transformative potential of AI/ML has yet to be fully realized at scale. This is often due to uneven access to the specialized talent and infrastructure required for successful deployment. While proof-of-concept projects are increasingly common, relatively few organizations have demonstrated the ability to operationalize and scale AI/ML solutions across their enterprise.

The need to standardize the structure and processing of analytic datasets has long been acknowledged, but progress has been slowed by a fragmented landscape of tools, technologies, and siloed data environments. Traditional approaches—often built around bespoke pipelines for individual machine learning projects—struggle to keep pace with today's demands. These challenges are now amplified by the increasing complexity of accessing and integrating both structured and unstructured data, which is frequently distributed across diverse systems, formats, and cloud platforms. At scale, this fragmentation creates bottlenecks in data discovery, preparation, and governance. The result is diminished productivity, with data scientists spending the majority of their time wrangling data instead of building models, along with delayed time-to-market and high failure rates for analytic initiatives.

The advanced analytics process can be thought of as consisting of three main components:

1. **Feature engineering:** This is a data management, integration, and manipulation task
2. **Model training:** This task—often referred to as model creation or modeling—involves using existing features and designing new ones to build mathematical models that can accurately predict future outcomes and support business decision-making
3. **Deployment:** The final step is to use the selected features and trained model to apply a scoring function to production data, to generate predictions

Teradata's analytics strategy is built on the belief that scaling machine learning and AI initiatives requires a strong focus on feature reuse, data harmonization, and model deployment. Feature engineering and model scoring align closely with Teradata's core value propositions, making them an ideal fit for Teradata Vantage® and ClearScape Analytics®.

To support scalable and efficient analytics, data science pipelines should be redesigned to populate and maintain an enterprise feature store—implemented as tables within an analytic RDBMS. This enables feature reuse across multiple models for both training and scoring. These curated sets of predictive variables are already accelerating data scientist productivity and reducing time to value for analytics in leading organizations. By deploying Teradata's Enterprise Feature Store on existing infrastructure, customers can minimize data movement and duplication, lowering total cost of ownership and latency. At the same time, they benefit from the high performance and scalability of Teradata Vantage® for processing large analytic datasets.

Model training typically uses carefully selected samples of historical data, while model scoring often requires access to complete, up-to-date analytic datasets from the feature store. Scoring is a mission-critical task that must deliver predictions to operational endpoints—often in near real time. Teradata systems are designed for high availability and industry-leading mixed-workload management. They are commonly integrated with operational endpoints across multiple channels and support "tactical" queries with response times in the tens of milliseconds—ideal for real-time scoring. With ClearScape Analytics®, Teradata customers can train predictive models using in-database functions or external tools via Bring Your Own Model (BYOM) and score those models directly against production data at scale. This is enabled by using curated subsets from the Enterprise Feature Store for training, whether in database or externally. Together, these capabilities make Teradata Vantage® and ClearScape Analytics® a powerful platform for operationalizing AI/ML at enterprise scale.

The Challenge

Data scientists and engineers often use pipelines—end-to-end processes built to solve specific problems on a project-by-project basis. These typically begin with feature engineering (also known as data wrangling). For small-scale development, testing, or exploratory work, this approach is efficient and repeatable, which is essential for research and regulated industries. Repeatability is usually achieved by storing pipeline code in version control systems like Git or SVN.

However, scaling this method across an enterprise can quickly become inefficient. It often leads to data silos and code-based processes that are difficult to understand—sometimes even by the original author. This phenomenon has been described as "pipeline jungles," where tangled data dependencies contribute significantly to technical debt. In fact, data dependencies tend to be more costly than code dependencies.

While data scientists are often seen as curious and eager to explore new tools, the reality is more nuanced. The community is diverse, and many professionals prefer to deepen their expertise in familiar tools and languages. For example:

- Python users tend to stick with Pythonic programming practices
- R users often rely on custom libraries they've built over time
- SAS remains a trusted tool in highly regulated industries

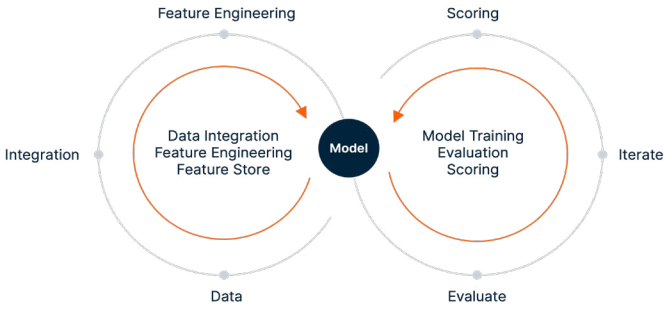
No single language or framework dominates the analytics landscape. For every advocate of R or SAS, there's a counterpart devoted to Python. This loyalty can make it challenging to introduce new tools or languages into established teams.

There is no universally "best" technology for model training across large, diverse organizations. Many different tools and methods can produce good outcomes. Data scientists are typically more productive and creative when using tools they know well, so enforcing a single technology can be counterproductive.

The rapid growth of machine learning and AI has led to a flood of new tools and frameworks. As the market matures, some will fade, others will specialize, and a few may consolidate into dominant platforms. But predicting which will prevail is uncertain and risky.

Beyond technical challenges, organizations face practical hurdles in adopting AI and AI agents:

- Limited access to skilled talent
- Rising infrastructure and compute costs
- Increasing concerns around data privacy and governance
- Steep learning curves that slow adoption and reduce ROI



Despite growing investment in advanced analytics, many projects never make it to production—and therefore fail to deliver value to the business. While "production" can mean different things to different roles, for business stakeholders it means that model outputs are accessible, trusted, and actively used to inform decisions—such as recommending the next best offer, reducing customer churn, or adjusting retail pricing.

To achieve this, production systems must be scalable, high-performing, robust, maintainable, and secure. Crucially, especially in regulated industries, organizations must be able to explain how and why a model made a specific prediction—even long after the fact. This level of transparency is essential for building trust and ensuring compliance.

The solution: Teradata's analytics and AI suite of capabilities

Many organizations struggle to scale analytics because they attempt to standardize on a single tool, language, or framework. However, the most successful organizations focus less on standardization and more on optimizing the entire analytics lifecycle—from data to activation. A flexible approach allows teams to quickly adopt new deep learning libraries or analytic tools that best fit the task at hand.

The key is to decouple the analytics process into distinct components, each running on the most suitable technology. Just as modern compute architectures separate storage from compute, analytics workflows should separate data preparation, model training, and deployment. This enables polyglot programming, where the best tools, languages, and frameworks are chosen based on the task or the user's expertise.

ClearScape Analytics® and AI capabilities support this modular approach, enabling users to build scalable, efficient, and adaptable analytics pipelines.

1. Data management and feature engineering

Data preparation still consumes up to 80% of the effort in analytics projects. To make AI and machine learning truly scalable, organizations must move away from building one-off data pipelines and instead invest in curated feature stores—centralized repositories of reusable, high-value features. Solutions like Teradata's Enterprise Vector Store unify access to structured and unstructured data, enabling faster discovery of relevant features, improved model accuracy, and reduced time spent on data wrangling. Embedded vector search capabilities allow teams to apply semantic understanding across diverse data types, driving smarter AI outcomes.

2. Model training

Model development is the core domain of data scientists. Within an enterprise framework, they need the freedom to explore data and algorithms using the tools they know best. ClearScape Analytics® supports this flexibility, allowing integration with external tools and languages through Bring Your Own Model (BYOM) and Bring Your Own LLM (BYO-LLM) capabilities.

However, the industry has overemphasized model creation. Production analytics at scale require more than just a predictive model—they demand a full lifecycle approach. To improve productivity and time to market, organizations should reuse features from the enterprise feature store and ensure any new features are added back for future use.

While feature engineering and model training are tightly linked during discovery, they should be treated as distinct activities once a model is finalized. This separation enables ClearScape Analytics® to seamlessly incorporate externally trained models alongside in-database models.

3. Model deployment

Once a model is trained, it must be deployed to generate predictions using live production data. With both the model and feature store residing in database, Teradata enables robust, scalable, and low-latency scoring without data movement.

The parallel architecture of Teradata Vantage® supports high-performance batch scoring and near-real-time predictions. Its hash-based file system allows scoring operations to execute in milliseconds with minimal resource usage.

Using Teradata ModelOps, organizations can automate model deployment, monitor for drift, retrain models, and manage champion/challenger workflows. ModelOps provides a comprehensive framework for governance, lifecycle management, and responsible AI, ensuring compliance and transparency across the entire pipeline.

4. AI-driven tools

As generative and agentic AI evolve, enterprises are seeking ways to unlock value from unstructured data. Teradata's Enterprise Vector Store, combined with retrieval-augmented generation (RAG) models, enables fast access to both structured and unstructured data.

Tools like AgentBuilder allow organizations to create autonomous AI agents that leverage trusted data and domain expertise. Meanwhile, AI Workbench offers a unified environment for development, operationalization, and collaboration—complete with notebooks, ModelOps, and third-party tool integrations.

Teradata strengths

The unique implementation of massively parallel processing (MPP) in Teradata Vantage® offers significant advantages over competing platforms. It provides a flexible, high-performance foundation for innovation and enterprise-scale analytics.

As machine learning becomes more ubiquitous, organizations will need to deploy tens to hundreds of millions of predictive models in production. Teradata Vantage® and ClearScape Analytics® have already proven their ability to scale:

- **Vertically:** Training models on millions of observations and scoring against hundreds of millions multiple times per day
- **Horizontally:** Supporting hyper-segmentation use cases by training and scoring millions of models daily

Key capabilities that power enterprise-grade AI include:

- **Always-parallel architecture:** Teradata's architecture combines parallel processing with a cost-based optimizer, enabling fast, efficient handling of large and complex datasets for data prep, model training, and scoring
- **Low-latency access:** Built on a logical, hash-based filesystem, Teradata Vantage® delivers near-instant access to localized data—ideal for tactical queries like real-time model scoring
- **Mixed-workload management:** Teradata Vantage® supports concurrent operational and analytical workloads without duplicating data across silos, ensuring efficiency and consistency
- **In-database analytics:** Native support for R and Python allows data scientists to prepare data, train models, and score predictions directly in-database—eliminating the need to rework models built in external tools
- **External tool integration**: Tight integration with leading analytics platforms (e.g., SAS scoring accelerator) enables seamless deployment of externally trained models at scale
- **QueryGrid® and IPE**: Teradata's virtualization and planning technologies allow fast, transparent access to data across lakes and ecosystems, enabling flexible exploration and integration
- **Auditability and explainability:** Temporal tables and full query logging provide precise visibility into model behavior—allowing organizations to understand why a prediction was made at any point in time
- **Bring Your Own Model (BYOM):** Teradata supports a wide range of model formats and tools—including PMML, ONNX, H2O MOJO, Dataiku, DataRobot, Hugging Face, SQL conversion, and native code—making it easy to operationalize models built anywhere
- **ModelOps framework:** Teradata ModelOps automates deployment, monitoring, and governance of AI/ML models, ensuring compliance, traceability, and faster time to value
- **Hybrid cloud flexibility:** Teradata Vantage® offers consistent functionality across on-premises, cloud, and hybrid environments—eliminating the need to reengineer applications and simplifying platform adoption
- **AI-driven innovation:** Teradata empowers enterprises to unlock insights from structured and unstructured data with tools like the Enterprise Vector Store, AgentBuilder for autonomous AI agents, and AI Workbench for streamlined development and deployment.

Conclusion

As technical debt in machine learning deployments grows and market dynamics shift, organizations must take a more comprehensive view of the analytics lifecycle. While competing for model training workloads remains important, there's a significant opportunity to support customers by migrating their advanced analytics workloads to Teradata—particularly data management and production deployment.

This shift in focus can unlock substantial value by improving productivity, reducing time-to-market, and lowering failure rates. Teradata Vantage® provides a robust platform for operationalizing analytics, helping organizations move beyond fragmented pipelines and toward scalable, integrated solutions.

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.**

About the authors

Chris Hillman is global AI lead, sales analytics, at Teradata. He has more than 20 years of experience working with analytics across many industries.

Martin Willcox is senior vice president of sales analytics at Teradata. He drives sales and consumption of Teradata solutions and services throughout Europe, the Middle East, and Africa.

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

teradata.