White paper

Teradata Cloud: Elastic Compute

Scale Smarter as Workloads Change

teradata.

Table of Contents

Executive summary

Enterprise workloads are changing. Scheduled pipelines, object-store analytics, AI preparation, ad hoc analysis, and agent-driven workloads create continuously shifting compute demand. Traditional cloud architectures — originally designed for web applications — force organizations to add capacity in large, coarse increments. Teradata Cloud Elastic Compute is different: it was purpose-built for the AI and agentic era, matching capacity precisely to the complex, variable demand patterns that define modern enterprise analytics.

Teradata Cloud Elastic Compute is designed for exactly this challenge. It delivers granular, isolated capacity that scales with workload demand — without permanently expanding the production core. Organizations can add focused compute in smaller increments, isolate variable workloads from production, access open data formats, and manage everything with intelligent workload controls.

The result: greater agility, closer alignment between cost and consumption, and the operational trust required to scale analytics and AI with confidence.

The Core Proposition

- Variable demand should not become permanent capacity.
- Elastic Compute adds focused, on-demand capacity around the workload — in smaller, more precise increments.
- Granular node-based scaling, isolated compute environments, scheduling, auto-suspend, and intelligent workload management align resources with actual demand.
- Enterprise-grade governance, security, and managed operations provide the control required to scale with confidence.

The Challenge: Variable Demand Should Not Become Permanent Capacity

Enterprise data platforms face a structural tension: workloads are increasingly variable, but traditional scaling adds capacity in large, coarse increments. The result is a predictable sequence of problems.

Scheduled pipelines run at off-peak hours and then sit idle. Object-store analytics spike around business events. AI preparation workloads consume bursts of compute during model training cycles. Ad hoc analysis from dozens of teams creates unpredictable demand peaks. Agent-driven workloads generate continuous, machine-scale consumption that no human team would ever produce.

When organizations respond by adding permanent capacity — provisioning for the peak — they pay for idle compute the rest of the time. When they underprovision, variable workloads compete with production, creating contention, delayed pipelines, and degraded performance.

The Scaling Challenge

- Adding capacity in large increments means paying for compute before it is actually needed — and continuing to pay when demand subsides.
- When variable workloads share compute with production, resource contention leads to missed SLAs and delayed pipelines.
- Moving to new data formats or platforms often requires another storage copy or a disruptive migration before any value can be realized.
- Without intelligent workload management, operational control becomes a burden — and cloud economics become harder to predict.

Teradata Cloud Elastic Compute: Scale Smarter as Workloads Change

Teradata Cloud Elastic Compute provides a new cloud architecture for the agentic era. Rather than forcing organizations to choose between overprovisioning and performance risk, Elastic Compute adds focused capacity around each workload — independently, in smaller increments, and only when needed.

Elastic Compute nodes are isolated by design. They scale variable workloads — ELT, experimentation, ad hoc analysis, demand peaks — in focused compute environments instead of competing for shared resources. Auto-scheduling and auto-suspend mean compute is consumed only during active workload execution, then released.

01

Granular, incremental scaling

Scale smarter as requirements change

Elastic Compute adds isolated, on-demand capacity in smaller node-based increments within configured limits for object-store analytics, ELT, batch, AI/ML, agentic, and team workloads — without permanently expanding the production core.

02

Workload isolation that protects production

Absorb variable demand without disrupting critical workloads

Variable ELT, experimentation, ad hoc analysis, and demand peaks run in focused compute environments instead of competing for shared resources — reducing risk to production performance and SLAs.

03

Open data access

Extend analytics and AI without disruptive re-platforming

Optimized for object storage and supported open table formats such as Iceberg and Delta. Modernize incrementally and activate open data within existing cloud architectures without another proprietary data copy.

04

Managed workload operations

Optimize changing workloads with operational and financial control

Uses policy, scheduling, thresholds, utilization, and workload signals — with governance, security, managed operations, and consumption visibility — for simpler operations and more predictable cloud economics.

The Four Pillars in Detail

01 — Granular, Incremental Scaling

Traditional scaling requires coarse jumps — warehouse-size increments that add capacity well ahead of actual demand. Elastic Compute changes that model. Capacity is added in smaller, node-based increments, within configured limits, for the specific workload type that needs it.

This applies across the full range of enterprise workload types:

- Object-store SQL analytics: Query one governed data domain without creating another platform or data copy.
- Scheduled ELT and batch: Run nightly, month-end, campaign, fraud, risk, or supply-chain processing when demand spikes — then release the compute when the job is done.
- AI/ML preparation: Scale feature engineering and model-preparation SQL against approved data paths, on demand.
- Agentic workloads: Give agents and applications a controlled execution path for variable, continuous requests.
- Departmental analytics: Support finance, marketing, risk, operations, or supply chain teams in a governed, isolated environment.

Competitive advantage

Traditional elastic architectures may require coarse warehouse-size increases or full-cluster duplication that add capacity ahead of actual demand. Teradata Elastic Compute scales in precise, node-based increments — only what each workload needs, when it needs it.

Business impact

Greater agility, improved resource utilization, and closer alignment between capacity, workload demand, and consumption cost.

02 — Workload Isolation That Protects Production

When variable workloads run on shared compute, everyone competes for the same resources. Analytics pipelines slow. SLAs slip. Teams work around the problem by duplicating environments — which creates its own cost and governance burden.

Elastic Compute eliminates this tradeoff. Each Elastic Compute environment is isolated — variable ELT, experimentation, ad hoc analysis, and demand peaks run in their own focused compute environments without touching production resources.

Competitive advantage

Shared compute models can create contention when variable workloads and SLA-bound production services compete for the same resources. Elastic Compute isolates each workload environment so production performance is never at risk.

Business impact

Scale new analytics and AI use cases while reducing risk to production performance and user experience.

03 — Open Data Access Without Disruptive Re-Platforming

Many enterprise organizations have data distributed across object storage in open table formats — Iceberg, Delta, and others. Accessing that data with traditional tools often requires another storage copy, a platform migration, or both.

Elastic Compute is optimized for object storage and supported open table formats. Organizations can work with governed data in place — extending analytics and AI to open data without disruptive re-platforming or proprietary data copies. Modernization happens incrementally, at workload level. Elastic Compute can also access multiple open table formats simultaneously — no rearchitecting, no forced platform migration. Organizations can additionally use Elastic Compute to vectorize unstructured data and write vectors directly to object store, enabling low-cost storage for AI and search use cases.

Competitive advantage

Proprietary architectures can require additional data copies or significant migration before customers can activate data across workloads. Elastic Compute accesses multiple open table formats natively — no rearchitecting, no additional copies.

Business impact

Reduce duplication and extend analytics and AI across existing cloud data architectures.

04 — Managed Workload Operations

Elasticity without management is just uncontrolled spend. Elastic Compute includes intelligent workload management — policy-based scheduling, utilization thresholds, workload prioritization, and auto-suspend — so organizations can control how and when compute is consumed.

Governance, security, managed operations, and consumption visibility are built in. Unlike point compute engines that delegate operational control to customers, Teradata Cloud manages mission-critical workload placement, prioritization, and economics by default.

Competitive advantage

Point solutions can leave workload prioritization, operational control, and consumption management to customers. Teradata Cloud manages these by default — with governance, security, and managed operations built in.

Business impact

Simpler operations, stronger workload control, and more predictable cloud economics.

Elastic Compute in Practice

Use Case: Self-Service ELT on the Data Lake

Consider a financial services organization managing petabyte-scale data lakes built on open table formats. Their data engineering teams run hundreds of pipeline jobs daily — transforming raw ingestion into curated, analytics-ready datasets. These jobs spike around market events, quarter-end closes, and regulatory reporting cycles.

With fixed infrastructure, they face a familiar choice: provision for peak and pay for idle compute 70% of the time, or accept job queuing and delays during critical periods.

With Elastic Compute, compute spins up when the pipeline runs and shuts down when it finishes — consuming Teradata's full function library, query optimization, and governance on-demand, at the cost profile of cloud-native processing. The outcome is faster pipeline delivery, lower infrastructure cost, and engineering teams freed to build rather than manage capacity.

What faster, more reliable pipelines unlock:

- **AI readiness:** Downstream AI models are only as current as the data that feeds them. More frequent pipeline execution means fresher features and better model performance.
- **Agent enablement:** AI agents operating continuously require data that is continuously refreshed. Elastic pipelines keep agents current without continuous compute cost.
- **Business agility:** Teams that can spin up a new ELT pipeline in hours respond faster to new data sources, regulatory requirements, and business opportunities.
- **Cost accountability:** With consumption-based pricing, business units see exactly what their workloads cost — creating natural incentives for efficient pipeline design.

Use Case: AI-Driven Demand Forecasting in Retail

Elastic Compute applies equally across industries wherever variable, event-driven compute demand meets governed enterprise data. Consider a global retailer running AI-powered demand forecasting across thousands of SKUs and distribution centers. Forecasting models require periodic retraining against large historical datasets — a compute-intensive process that runs on a weekly or event-triggered schedule, not continuously.

With Elastic Compute, feature engineering and model-preparation SQL run in isolated compute environments — spinning up when needed, releasing when done — while production analytics remain unaffected. Vectors generated from unstructured product data (imagery, descriptions, reviews) can be written directly to object store, enabling low-cost AI retrieval without duplicating data into a separate platform. The pattern extends naturally to manufacturing, logistics, and any industry where workloads are scheduled, intensive, and predictable in timing but variable in scale.

Why Teradata's Approach Is Different

Most cloud platforms can add compute on demand. The harder problem is making that elasticity work across governed enterprise data, mixed workload types — now including AI preparation and agentic workloads —, and mission-critical SLAs — without creating contention, duplicating data, or fragmenting operational control.

That is where workload-aware elasticity matters. When compute scales in isolation, variable pipelines and AI workloads absorb demand without touching production. When open data formats are supported natively, teams extend analytics without disruptive migrations. When scheduling, prioritization, and governance travel with the compute, organizations maintain control over cost and performance even as workload patterns change.

The result is elasticity that actually fits enterprise scale — not just infrastructure that grows and shrinks, but compute that adapts intelligently to what each workload needs, when it needs it, without compromise to the data environment it runs against.

What Sets Teradata Elastic Compute Apart

- **Scale against governed data in place:** No ETL to a separate compute layer, no diverging schemas, no additional security perimeter to manage.
- **Intelligent workload management built in:** Scheduling, prioritization, and operational controls ensure variable workloads never disrupt production performance or SLAs.
- **Native open table format support:** Work with Iceberg, Delta, and OTF-based data lakes without data movement or re-platforming.
- **Granular node-based scaling:** Add capacity in precise, workload-sized increments — so infrastructure cost stays aligned with actual demand.
- **Enterprise governance as a standard:** Security, managed operations, and consumption visibility

Not Every Workload Benefits from Elastic Compute — That's By Design

That is why Teradata offers an Intelligent Workload Fitness Assessment — to identify the specific pipelines where Elastic Compute would deliver measurable improvement before any commitment is made. The assessment evaluates workload patterns, scheduling, concurrency, and data locality to determine where isolated, on-demand compute creates the most value. Organizations get a clear picture of expected outcomes before they scale.

Ask your Teradata team about the Intelligent Workload Fitness Assessment to see which of your workloads

The Case for Acting Now

Variable demand is not going away. Scheduled pipelines, AI preparation, agentic workloads, and ad hoc analytics will continue to create shifting, unpredictable compute requirements. Every day that organizations respond to that demand by adding permanent capacity, they pay the cost in overprovisioned infrastructure, resource contention, and operational complexity.

Teradata Cloud with Elastic Compute is a starting point that pays dividends immediately: lower pipeline costs, closer alignment between compute and consumption, and a platform that scales with changing workloads without a corresponding increase in operational burden.

The question is not whether to adopt elastic infrastructure. It is whether to do it with the workload isolation, open data access, and intelligent management that enterprise scale requires — or in a way that trades those properties for short-term simplicity.

Scale smarter → isolate variable demand → activate open data → operate with control

Teradata Cloud Elastic Compute:
Scale with Precision. Operate with Confidence.

- **Granular, node-based Elastic Compute** — scale in precise increments without permanently expanding the production core.
- **Workload isolation by design** — variable workloads scale independently without disrupting production SLAs.
- **Open table format support** — Iceberg, Delta, and OTF-based data lakes, no re-platforming required.
- **Intelligent workload management** — policy-based scheduling, auto-suspend, governance, and consumption visibility built in.
- **Consumption-based economics** — infrastructure cost aligns with actual workload demand and business value delivered.
- **Proven enterprise trust** — decades of workload management, security, and mission-critical scale.

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

teradata.