

VantageCloud Lake

Product Overview and Customer Benefits

Robin Jessani
Technical Product Marketing



Table of Contents

- 3 Overcoming challenges with a cloud-native design
 - 3 Restrictive policies
 - 3 Disjointed tools and services
 - 4 Resource contention
 - 4 Isolated data silos
 - 4 Suboptimal governance

- 5 An architecture built for superior performance
 - 5 Point of deployment (POD)
 - 6 Primary cluster
 - 7 Compute clusters
 - 7 Lake File System (LFS)
 - 7 Open Data Formats
 - 7 Query fabric

- 8 Innovate faster, scale smarter, govern better

- 9 Empowering faster innovation
 - 9 Advanced analytics
 - 10 Open and connected ecosystem
 - 10 Business autonomy

- 11 Enabling smarter scaling
 - 11 Growth optimization
 - 11 Flexibility and sustainability
 - 11 Scalability

- 12 Supporting better governance
 - 12 Cost visibility
 - 12 Centralized management
 - 13 Lower total cost of ownership

- 14 Summary

Overcoming challenges with a cloud-native design

In recent years, the rapid growth of cloud technology has transformed the data and analytics landscape in profound ways. With cloud adoption becoming widespread, Teradata seized on this opportunity with the development of an offering that capitalizes on cloud-enabled frameworks to extend the capabilities of VantageCloud, Teradata's complete analytics and data platform for trusted AI.

This innovative deployment option, VantageCloud Lake, is built on a next-generation, cloud-native architecture, enabling businesses to meet the full spectrum of analytic enterprise workloads, including AI/ML, in the cloud. With its cloud-native design, VantageCloud Lake delivers advanced analytics to users of any skill level while also overcoming the main challenges that customers contend with in their existing data analytic ecosystems. These include:

- Restrictive policies
- Disjointed tools and services
- Resource contention
- Isolated data silos
- Suboptimal governance

Together, these factors can create significant obstacles that slow innovation and impede business growth. Below we'll examine each of these challenges and explore how VantageCloud Lake addresses them to help companies fast-track digital transformation and drive breakthrough business outcomes.

VantageCloud Lake helps companies fast-track digital transformation and drive breakthrough business outcomes

Restrictive policies

To protect SLAs, many companies adopt restrictive policies that require heavy reliance on IT for operational support, inadvertently introducing extra steps and bottlenecks that hinder innovation. But as the use of public clouds grows, stakeholders are increasingly demanding technologies that deliver greater autonomy and agility. Customers today require self-service independence that applies to all lines of the business, including Finance and Marketing.

VantageCloud Lake addresses this expectation by offering a self-service environment that enables business departments, teams, and users to provision, manage, and operationalize their own dedicated resources with little to no administrative involvement. Providing this level of unrestricted independence gives users the agility they need to quickly react to volatile business goals and accelerate innovation for the company.

Disjointed tools and services

As cloud computing adoption increases, so does software specialization. Many cloud service providers (CSPs) and third-party software companies offer an extensive suite of analytics tools and services that are attractive or even required for the business. Key stakeholders, like developers and data scientists, need access to these industry-leading technologies so they can innovate efficiently and effectively. While external tools and services can offer an intelligent way to enhance platform functionality, they can create management inefficiencies if they're not connected to the underlying platform.

VantageCloud Lake addresses this concern with an open and connected environment, leveraging open APIs to integrate with and manage a wide range of third-party tools and services on one platform. This significantly improves management efficiencies while enabling users to use their preferred tools and services, accelerating time to value.

Resource contention

There are often conflicting priorities when resources are distributed across the organization, and workloads with strict mission-critical SLAs risk jeopardy or disruption from growing demands. System workloads have to contend for shared resources, requiring IT to manually prioritize and allocate them based on workload importance. And when storage and compute resources are finite and coupled together, this negatively impacts business-critical workloads due to growing demands between teams. Ad hoc and exploratory workloads slow down performance, and complex analytics queries risk disrupting critical workloads.

Lines of business need their own independent compute so they can isolate their workloads from one another and accelerate innovation without this risk. VantageCloud Lake enables this by leveraging the cloud to separate compute and storage resources so that departments can have isolated and scalable multi-cluster compute. This eliminates resource contention by giving teams the flexibility to provision and regulate their own compute resources while not impacting workloads outside of their team or department.

Isolated data silos

Data movement costs the business time, manpower, and money. Many customers have spent millions of dollars moving their data into isolated pipelines and silos over the years, creating massive tech debt and management problems. These data silos are often created for short-term projects, only to end up underutilized or even forgotten—all the while incurring ongoing costs. Businesses need a way to consolidate and share their data against a single data store.

VantageCloud Lake addresses this issue by integrating with low-cost, third-party object storage to ingest and consolidate all sources of data. It also provides support for open table formats (OTFs) and open file formats, enabling users to effectively manage and query data within the object store. Data living in the object store can be accessed, shared, and queried across the ecosystem of connected systems. VantageCloud Lake's harmonized data approach significantly reduces data movement while minimizing costs accrued from poorly managed or forgotten data pipelines. The object store introduces additional cost savings by operating independently from compute. This separation of storage from compute allows customers to elastically grow or shrink their storage capacity without impacting their compute resources, providing them with the flexibility to efficiently manage each resource as required, ultimately optimizing resource allocation and reducing overall costs.

Suboptimal governance

Suboptimal governance and spend management practices make it challenging for businesses to effectively oversee their investments and finances. In the absence of sophisticated cloud technologies, organizations rely on manual governance processes, leading to errors and missed opportunities. As a result, they tend to overspend on their investments, underutilize their resources, and overlook optimization opportunities.

VantageCloud Lake solves for this by providing an extensive range of governance capabilities for both administrators and analysts, enabling customers to make the most of their investments. Administrators can utilize graphical resource consumption telemetry and query monitoring insights to make informed, agile decisions for the organization's performance tuning, forecasting, and budgeting. Analysts can utilize VantageCloud Lake's differentiated auto-scaling algorithm and policy-driven guardrails, such as the max number of auto-scalable nodes and on/off scheduling, to optimize resource usage. All of these governance capabilities work together to increase customer ROI while reducing total cost of ownership.



An architecture built for superior performance

The introduction of VantageCloud Lake advances Teradata as a leader in cloud analytics and data. With its reengineered architecture and industry-leading Analytics Database, VantageCloud Lake lends itself to modern cloud capabilities to bring superior performance across all enterprise workloads at scale. Analytics workloads of any requirement, whether business critical with tight SLAs, transactional, AI/ML, data engineering, or data science, can be executed to unlock profound discoveries that can transform the business.

To better understand how the new architecture resolves customer challenges, let's take a closer look at the components that comprise the VantageCloud Lake tech stack: point of deployment, Console, primary cluster, compute cluster, Lake File System, and query fabric.

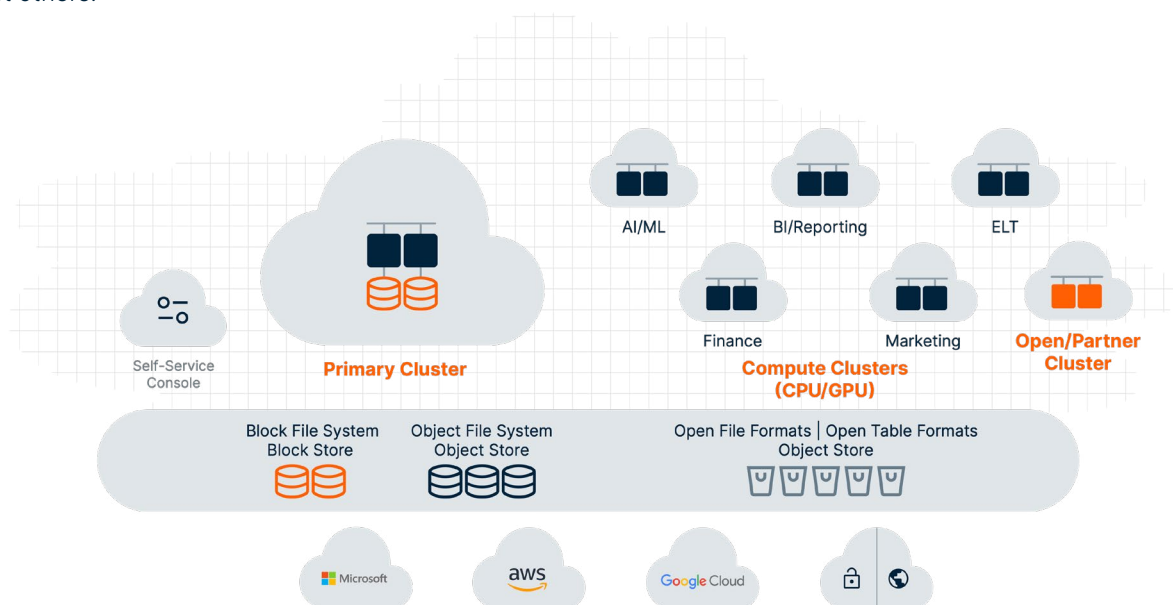
Point of deployment (POD)

Not depicted in the graphic below is the POD mechanism that supports all activity taking place in the VantageCloud Lake platform. Each POD can support one to many customer accounts, known as tenants, which are separate database instances. Because tenant environments are logically isolated from one another, the failure of a single tenant will not affect others.

Console

The VantageCloud Lake Console is the web-based first point of entry for users. It's an easy-to-use, self-service browser with centralized navigation that contains an arsenal of capabilities that administrators and users will commonly interact with. These capabilities are accessible with single sign-on (SSO) login, providing both security and convenience as users access their data and initiate sessions to all connected applications.

Console functionality is enhanced with ask.ai, Teradata's generative AI-powered tool. Teradata ask.ai is a natural-language interface that enables users to "talk" to the platform to explore and make knowledge intelligence decisions instead of using complex tools and processing languages. This brings a new level of accessibility to wide range of users, not just those with advanced skill sets and knowledge of business intelligence and data science tools.



The Console serves three personas, each with their own pain points and requirements that the Console addresses:

- **Administrators:** The Console serves administrators by offering self-service provisioning, simplified user management, and performance monitoring to help identify and optimize slow-running queries. Console makes it is easy to ingest, copy, and access access all your data from across your organization with Flow, Data Copy, and QueryGrid. It provides graphical telemetry of the organization's compute and storage consumption that can be drilled down to the department level, and it includes filters for time, time ranges, and contract terms. And with the integration of ask.ai, administrators can retrieve system information effortlessly through natural-language queries and access to real-time support.
- **Business analysts and citizen data scientists:** The Console's self-service, user-friendly features enable users to perform analytics effectively and easily identify key insights. It offers various low-code/no-code data movement tools that simplify data access and collaboration across the organization to accelerate data exploration. The built-in data insights and point-and-click Query Builder give users access to quick metrics and visualizations to explore their data and identify trends. The Function Explorer facilitates direct access to ClearScape Analytics™, Teradata's advanced analytics capabilities within VantageCloud Lake, by offering clear instructions for building queries using the full catalog of in-database AI/ML functions. This allows users to efficiently construct complex queries like nPath analyses. When coupled with ask.ai, the Console flattens the learning curve, enabling users to generate data-driven insights without having to manually construct queries and scripts.
- **Technical users:** The Console empowers developers, data engineers, data scientists, and other technical users by simplifying advanced analytics processes. Users can access the end-to-end AI/ML lifecycle from data preparation to model training and deployment with ClearScape Analytics features, including SQL editor and ModelOps. Data sharing promotes cross-team collaboration by sharing up-to-date curated datasets. The Function Explorer offers a simplified user experience, allowing users to easily find in-database analytic functions that meet their needs. With ask.ai, users can ask about help features and available use cases or make requests like "Identify and resolve the syntax error in this line of code" to speed up data preparation or model creation steps and enhance productivity. The Console also integrates seamlessly with external partner tools and services



through an open ecosystem, enabling users to leverage their preferred technologies and streamline tasks such as model building and deployment for advanced analytics.

Primary cluster

Each environment has a primary cluster that's configured in the Console during the provisioning process. Users can either set up this pattern or leverage the primary cluster as an independent database system. This is very powerful, as customers can migrate to the primary cluster and then modernize by leveraging the compute clusters selectively by workload at their own pace.

There are several use cases that are uniquely addressed with the primary cluster because it leverages persistent, high-performant block storage associated with each access module processor (AMP). Customer lift-and-shift migrations to the primary cluster are the easiest, fastest, and lowest-risk way to migrate to VantageCloud Lake, with gradual workload modernization leveraging compute clusters and low-cost object storage occurring afterwards. Additionally, the primary cluster is optimized for enterprise transactional queries at scale, enabling strict mission-critical SLAs for workloads that require predictable sub-second performance.

The primary cluster architecture is the same as an analytics database, consisting of interconnected nodes with AMPs that handle data processing, movement, and management from disk drives. This design allows the primary cluster to efficiently distribute workloads across the user environment for optimized query planning.

Compute clusters

Compute clusters operate on nodes independent from the primary cluster to serve unique functions. They are highly elastic by design and can automatically scale on/off and up/down in accordance with the user's policy-driven configuration, enabling the system to automatically optimize compute resource allocations based on various parameters like resource utilization and workload priority. Compute clusters leverage object storage to store large volumes of data in a cost-effective manner, while the primary cluster stores the required metadata. An environment can have zero to many compute clusters; it depends entirely on how much compute is needed to complete the query queue within the expected time frame.

Compute clusters are used to execute a broad range of enterprise workloads, including analytics, reporting, ELT, and departmental workloads. Teams like Finance or Marketing can dynamically provision multiple compute clusters against a single data store and will have their query work executed on their own dedicated resources.

Analytics compute clusters are also offered to support memory-intensive analytic applications that rely on in-database and/or imported analytic functions to run in a containerized, open analytics framework. They provide twice as much memory as standard compute clusters and therefore are ideal for advanced AI/ML workloads like data preparation and model building/training/operationalization. Analytics compute clusters provide the flexibility to use either GPUs or CPUs, depending on the complexity and size of the workload.

Since compute clusters are independent of one another and rely on the primary cluster to send them work, they can be added or removed as needed. More importantly, the failure of a compute cluster will not disrupt work being processed elsewhere.

Lake File System (LFS)

The LFS is VantageCloud Lake's unified data storage system that supports block storage via the Block File System (BFS) and object storage via the Object File System (OFS). While customer-owned data stored in object storage can be accessed outside of VantageCloud Lake using Teradata's Native Object Store (NOS), many attractive benefits are gained by ingesting data into the LFS, such as access to Teradata's Analytics Database capabilities, optimized performance, and data sharing between connected Teradata systems. The BFS manages tables stored in block storage available only to the primary cluster. Block storage uses AMP storage to offer persistence for business-critical workloads that require high reliability and low latency for tactical workloads with strict SLAs.

The OFS manages tables stored in object storage available to both compute clusters and the primary cluster. Object storage is low cost and therefore a preferable solution for reducing storage costs tied to large volumes of data. It also takes advantage of the compute cluster architecture to scale and isolate workloads. From a workload perspective, object storage is a preferable storage solution for analytics workloads that are not time sensitive or are generally long running.

The combination of the OFS and compute clusters enables complete separation of compute and storage: data stored within VantageCloud Lake is accessible to compute clusters independently managed by different teams. This architecture makes the system fully scalable, with each team managing its own compute resources without needing to move or copy shared data.

Open Data Formats

VantageCloud Lake further enhances data management by supporting low-cost object storage using open file formats such as Parquet, Avro, and CSV and **open table formats (OTF)** such as Delta Lake and Iceberg. Integration with OTFs enables data living in customer-owned object stores to be centrally managed and optimized for performance on one OTF storage layer that sits on top of the object store. The OTF storage layer brings fundamental database capabilities to customers' data lakes with interoperability across a broad range of open file formats and storage types. This enables seamless data preparation for any language or framework and significantly reduces data management complexities and replication across the VantageCloud ecosystem.

Query fabric

Query fabric is powered by Teradata's QueryGrid software, allowing VantageCloud Lake systems to connect to data stores and processing engines between multiple systems to enable a hybrid, multi-cloud ecosystem. Query fabric facilitates data connectivity within the ecosystem while also enabling push-down processing to send query instructions to the data at its source, creating a harmonized data approach that minimizes data movement. Query fabric also significantly improves query performance through the strategic use of bidirectional connectors for workload parallelization and Teradata's Query Optimizer tool for planning query paths across systems. This results in a seamless and high-performing query federation, bridging the gap between both Teradata sources and external data sources for comprehensive analytics while eliminating the constraints of data silos.

Innovate faster, scale smarter, govern better

The architectural components of VantageCloud Lake enable customers to achieve three key goals:

- **Innovate faster:** Customers can capture a greater ROI and achieve a faster time to value. VantageCloud Lake's self-service, user-friendly interface flattens the learning curve so that users of any skill level are empowered to drive data-driven insights with agility. The environment is open and connected to external technologies and includes over one hundred in-database functions to speed up data preparation and model operationalization workflows while enhancing the ability to extract value from data.
- **Scale smarter:** VantageCloud Lake is an environmentally sustainable and economical platform, optimizing resource utilization to ensure customers only pay for the resources they use. Teradata's smart-scaling technology provides dynamic resource allocation, delivering cost-effective and automated scaling as business needs evolve.

- **Govern better:** VantageCloud Lake equips customers with centralized management over the entire organization's financial governance. Through consumption telemetry with visualizations and data filtering, administrators gain tighter control. This enables streamlined resource management, eliminates overruns, and empowers user flexibility without disrupting core operations.

In the next section, we'll examine the key drivers behind each of these benefits.

VantageCloud Lake's self-service, user-friendly interface flattens the learning curve so that users of any skill level are empowered to drive data-driven insights with agility

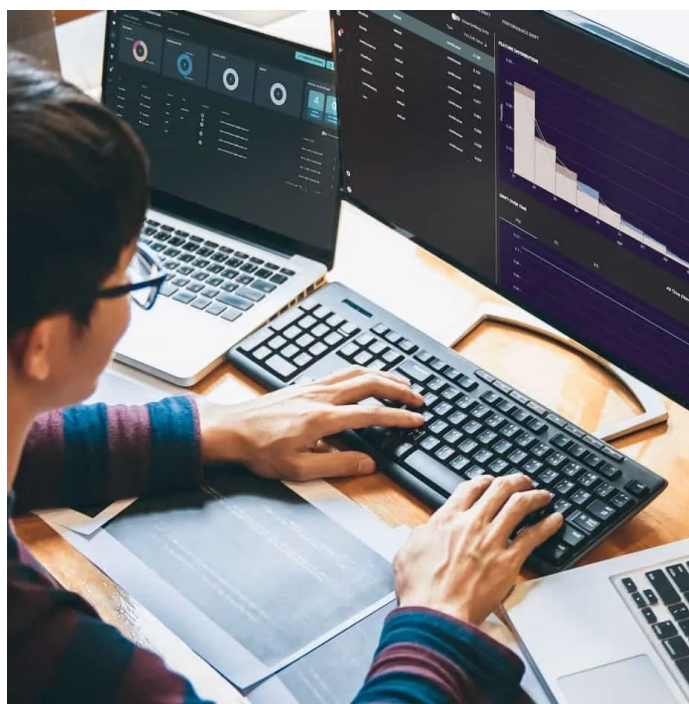


Empowering faster innovation

Advanced analytics

ClearScape Analytics is a comprehensive suite of powerful analytical capabilities seamlessly integrated into VantageCloud Lake. This integration is a standout feature of VantageCloud Lake, setting it apart from other data analytics platforms. ClearScape Analytics serves as an industry-leading advanced analytics solution that addresses every aspect of the analytics lifecycle to deliver accelerated value at scale. From streamlining data preparation to enabling efficient model deployment, ClearScape Analytics, in conjunction with VantageCloud Lake, empowers organizations to:

- Accelerate data preparation time by up to 70%
- Go from a few models to millions of models at one-seventh the cost
- Deploy models in minutes or days and operationalize them at scale



ClearScape Analytics makes this possible with features like:

- The most extensive, versatile in-database functions for better data prep, model training, and deployment to solve complex analytic problems, minimize risk, and deliver insights based on a complete view of your data—all with zero data movement
- Frictionless connectivity to a variety of open-source and partner AI/ML tools to make it easier to access insights and maintain the flexibility to make the right choices for your business, resulting in increased efficiency and reduced time to value
- Integrated ModelOps, with capabilities including standardized lifecycle management, automated monitoring, simplified deployment, and advanced governance, which removes the friction of getting AI/ML models into production and cuts deployment time from months to days
- Bring Your Own Model (BYOM) and Bring Your Own Analytics (BYOA) capabilities that make it easy for data science teams to operationalize externally developed models, accelerating time to value and delivering improved growth and performance for all models, even those developed in other tools

Users of any skill level have direct access to ClearScape Analytics via the ask.ai tool. The richness of ClearScape Analytics' in-database functions, combined with the natural-language interface of ask.ai, flattens the learning curve for nontechnical users and accelerates time to value for technical users so that the entire business workforce can innovate faster with their data.



Open and connected ecosystem

By offering an API-driven, open, and connected ecosystem, VantageCloud Lake's analytic capabilities are extended to a broader range of technologies. Customers gain a competitive advantage by integrating their environment with external tools and services that boost efficiency and serve unique use cases.

This interconnection with the public cloud infrastructure extends to every facet of the data analytics journey. Data stored in other Teradata systems or external locations is easily accessed through NOS, fostering data consolidation and the elimination of data silos. Partner integrations with modern data stack tools like dbt and Airbyte simplify ELT and data preparation tasks. For analytics, users can access VantageCloud Lake through Jupyter Notebook to write in Python or R languages and create detailed visualized reports that effectively convey insights. They can also leverage BYOM or Bring BYOA to optimize model building and deployment workflows, accelerating the operationalization of advanced analytics. This level of flexibility allows customers to use cloud technologies that best align with their business needs.

Business autonomy

VantageCloud Lake provides businesses with greater autonomy, empowering every facet of the organization to drive transformative innovation. Departments and teams are equipped with dedicated yet independent resources that ensure workloads do not disrupt one another, thus enabling workload isolation throughout the organization. Self-service provisioning of scalable resources eliminates reliance on IT and associated waiting times. Collaboration is fostered through services like in-place data sharing that facilitate the secure sharing of private data between teams.

Autonomy is further enhanced by Console features that are self-service and easy to use. These features empower users of any skill level to independently perform analytics, freeing the organization from technical barriers and bottlenecks that slow down innovation. VantageCloud Lake offers a wide array of user-friendly data movement tools that streamline data accessibility and collaboration between teams for accelerated data exploration efforts. The Function Explorer and Query Builder tools provide clear, step-by-step instructions to aid users in easily crafting queries using ClearScope Analytics' extensive library of in-database functions. If that isn't enough, the generative AI-powered ask.ai tool paves the way for a simplified analytics journey from start to finish, fostering innovation at warp speed for users of any skill level.

Enabling smarter scaling

Growth optimization

Compute clusters are highly elastic and automatically scale using Teradata's "scale smarter" algorithm, which allocates additional compute resources only once existing resources have been fully utilized. While other platforms may trigger an auto-scale after an ambiguous limit has been surpassed, such as query volume or user concurrency, VantageCloud Lake optimizes growth by making sure that a compute cluster is fully used before auto-scaling another one, ensuring that customers only pay for the resources they use. Policy-driven guardrails are also included to offer tighter controls for optimizing investment efficiency and staying within budget as teams expand and grow. When combined with Teradata's workload management expertise and its Query Optimizer tool, which optimizes query prioritization and resource distribution, customers know that their fixed resources are being optimally utilized.

Flexibility and sustainability

VantageCloud Lake's cloud-native architecture supports multi-cluster auto-scaling. This capability enables each department, team, and individual to dynamically and automatically grow and shrink their specific environment without disrupting or impacting other environments, creating a seamless coexistence between the organization's resources and operations. This empowers customers to stay agile in the market with flexible and sustainable scaling across the entire organization.

Scalability

Teradata's scope of scalability is proven to be industry leading. With over 1,000 nodes successfully scaled and benchmarked, without any outages or downtime, VantageCloud Lake demonstrates that it can handle workloads of the largest and most complex customer environments.

Teradata's smart-scaling technology provides dynamic resource allocation, delivering cost-effective and automated scaling as business needs evolve



Supporting better governance

Cost visibility

The VantageCloud Lake Console provides detailed telemetry data of the environment to facilitate financial transparency and chargeback to customers. Graphical telemetry of storage and compute consumption, which can be drilled down to the department level, is provided with time, time range, and contract term filters. This makes it easier for administrators to navigate and comprehend financial telemetry data, enabling greater agility and accuracy in their decision-making. Additionally, the Console includes cost avoidance metrics like cost-performance ratio to help manage budgeting and unit allocation.

The generative AI-powered ask.ai tool also facilitates administration. Using natural language, administrators can easily retrieve system information related to VantageCloud Lake, such as the status of the environment and compute groups. With a question like, “Which compute groups are active and what’s the consumption of each?” administrators can identify compute groups that are active but not in use, presenting an opportunity to hibernate them and save costs. Additionally, ask.ai has access to general documentation, so users can get immediate assistance when they encounter problems or have questions.

Centralized management

In conjunction with enhanced cost visibility, VantageCloud Lake enables centralized management across the organization. From a financial governance standpoint, this means that administrators have the power to centrally access, manage, and bill each department for the resources they use. Management capabilities also extend to queries with Teradata’s Active Query Monitor, which provides a summary view of active and completed queries. Having these capabilities on one centralized platform ensures access to valuable, detailed insights on each limb of the business, helping administrators to make informed, accurate decisions and govern their investments better to meet the needs of the business.





Lower total cost of ownership

VantageCloud Lake offers several advantages that allow customers to lower their total cost of ownership, including:

- **Separation of compute and storage:** Computing and storage resources are now separated, so users can grow their compute without having to incur the costs of growing their storage (and vice versa).
- **Flexible compute clusters:** Compute clusters can be provisioned and deprovisioned based on fluctuating needs (ad hoc workloads, weekends, holidays, seasonal lulls, etc.). This allows for increased financial control over when to tactically grow or shrink compute resources in accordance with project timelines and query complexity/priority.
- **Flexible object storage:** Massively scalable, low-cost storage options include the OFS, open file formats, and open table formats. The OFS is a Teradata-managed, fully ACID-compliant data lake, optimized for performance including caching, spool, and index metadata. Open file formats (Parquet, JSON, CSV) and open table formats (Iceberg, Delta Lake) enable shared access to customer-managed data lakes. This architecture provides customers with flexibility to choose the most appropriate storage solution based on workload requirements and budget.
- **Self-service accessibility:** Self-service accessibility reduces IT overhead by reducing labor-intensive administrative tasks, such as provisioning the system, managing environments, and installing software.
- **Reduced data duplication and movement:** The LFS enables organizational teams on the same system to share one copy of the same data. QueryGrid enables push-down processing that function-ships query instructions to the data at its source. As a result, data duplication and movement are significantly reduced, decreasing any associated egress costs.
- **Auto-scaling guardrails:** Users can designate policy-driven guardrails for each compute cluster to ensure that costs associated with auto-scaling meet expectations. These guardrails include parameters such as maximum number of auto-scalable nodes and on/off scheduling.
- **Auto-scaling trigger:** Instead of triggering a compute cluster to auto-scale after an ambiguous limit has been surpassed, Teradata offers an optimized algorithm that ensures compute resources are fully utilized before initiating the next batch of resources. This ensures that customers only pay for the resources they use.

Summary

In summary, VantageCloud Lake offers a transformative solution that enables customers to innovate faster, scale smarter, and govern better. Its robust architectural design and cloud-enabled capabilities empower users of any skill level to execute all types of data analytics use cases across the entire organization. The integration of the generative-AI powered ask.ai tool further reinforces a new standard for knowledge intelligence, propelling organizations into the future of data-driven success. VantageCloud Lake is available for purchase at www.teradata.com.