Cloud-Based Analytics Platforms

A workload comparison study

John Myers
Senior Director, Market Intelligence, Teradata

teradata.

Executive Summary

Selecting the right cloud-based analytics platform is crucial for organizational success. This white paper presents a comprehensive workload comparison analysis of three leading cloud analytics and data platforms from Teradata, Snowflake, and Databricks.[1]
  In this study focusing on enterprise price performance—a critical metric balancing cost efficiency with analytical capabilities at scale—Teradata VantageCloud Lake emerged as the leader in cost efficiency and query throughput, highlighting the importance of sophisticated workload management and query optimization in cost-conscious cloud environments.

Key insights for decision-makers

Cost management. The significant differences in price performance suggest substantial potential for cost savings through careful platform selection. This could have a major impact on overall cloud analytics budgets for business stakeholders and information technology (IT).

Scalability. Teradata's ability to process substantially more queries without proportional cost increases indicates better scalability for growing data volumes and user bases. This is crucial for organizations anticipating rapid growth or facing unpredictable analytical demands as their analytics requirements mature.

Operational efficiency. Higher query throughput can translate to faster insights and decision-making, potentially providing a competitive edge for data-driven organizations where rapid response to market changes is critical.

Total cost of ownership (TCO) and return on investment (ROI). The workload comparison study results suggest Teradata could offer a lower TCO and higher ROI for organizations with complex analytical needs. Decision-makers should consider these long-term implications when evaluating platforms.

Workload management. Teradata's performance underscores the importance of sophisticated workload management and query optimization in cloud environments. Organizations should evaluate not just individual query performance but how well a platform can manage diverse, concurrent workloads at scale.

---

1. For more information on the Teradata competitive workload comparison methodology and process, go to **www.teradata.com/how-we-help/work-load-comparisons** or contact **John.Myers@Teradata.com**

Introduction

Elastic cloud-based compute resources have revolutionized data analytics by providing unprecedented scalability, flexibility, and access to computational power. As organizations increasingly migrate their analytical workloads from on premises to the cloud, the promise of improved agility and reduced infrastructure costs is a key modernization driver. However, as organizations manage these efforts, the need for cost efficiency and financial governance in their cloud analytics practices becomes an additional imperative.

In today's competitive business environment, the ability to derive insights quickly and accurately from vast amounts of data is no longer just an advantage—it's a necessity. Cloud analytics platforms provide the tools and resources to meet this demand, enabling organizations to process, analyze, and visualize data at scale that was once unimaginable. Yet, as data volumes grow exponentially and analytical workloads become more complex, the costs associated with cloud analytics can escalate rapidly if not managed effectively.

This workload comparison study aims to provide insights for decision-makers navigating the growing landscape of cloud-based analytical platforms, balancing the need for powerful analytical capabilities with the imperative of cost effectiveness.

Why dare to compare?

Rigorous **workload comparisons** in cloud analytics aren't just about evaluating numbers. They're about understanding the practical implications of choosing one platform over another in the context of an organization's unique needs and goals—especially financial governance and budgeting objectives. Workload comparison testing is essential for:

- Objectively analyzing performance, cost efficiency, and scalability across different platforms

- Identifying strengths and weaknesses in handling specific analytical workloads

- Providing actionable insights beyond marketing claims

- Aligning technology choices and business objectives with budgetary constraints

These evaluations help to reveal the strengths and weaknesses of each platform in handling specific workloads. They can demonstrate how individual platforms impact real-world requirements, going beyond qualitative claims to provide actionable insights for platform selection.

With a clear picture of what to expect in terms of performance and cost, customers can align their technology choices with business objectives.

teradata.

Workload comparison methodology

Our methodology was meticulously designed to ensure fair and meaningful comparisons across often disparate platforms. It consists of three key components that work in tandem to create a balanced testing environment.

1. **Developed a workload mix to test the platforms utilizing real-world workload situations.** This comprehensive mix combines queries from an industry standard decision support benchmark with those that simulate diverse, concurrent, operational analytics queries. This query mix reflects data-driven organizations' need to capitalize on real-time decision-making capabilities to drive their businesses and establish competitive advantages in fast-moving industries. This diverse query set ensures that our workload analysis accurately represents complex and varied demands placed on today's cloud-based analytics platforms, rather than a less sophisticated historically focused query set.

2. **Established cost-equivalent environments for each platform.** This approach allows each platform to leverage its strengths within a standardized budget, mirroring the constraints that businesses often face when deploying cloud analytics solutions.

3. **Minimized system tuning.** Overall, no specific platform tuning was performed for the defined workloads. Performance features that would be commonly enabled by database administrators (DBAs) and were generally available (GA) on the platforms were deployed.
   A specific feature—often referred to as "materialized views"—was deployed on all platforms—if available.[2] This feature was deployed on select larger tables in accordance with DBA best practices.

This three-pronged approach aims to level the playing field, allowing for apples-to-apples comparisons despite the inherent architectural differences between systems. By carefully controlling these conditions, we've created a workload comparison evaluation that provides valuable, actionable insights for decision-makers evaluating cloud-based analytics platforms.

---

2. Teradata VantageCloud Lake refers to this feature as a "single table join index." Snowflake refers to this as "materialized view."
   Databricks did not have this concept as a generally available (GA) feature until June 2024 and was not included in the configuration.

teradata.

Our real-world workload

Today's analytical environments must enable organizations to understand what's happened historically, what's happening today, and what's likely to happen in the future. The set of commonly used benchmarks doesn't effectively test all the requirements of a data-driven organization. The methodology detailed below more effectively mimics the workloads executed at the world's most sophisticated, and often largest, data-driven companies.

Building on an industry-accepted foundation

Our methodology starts with one of the currently available benchmarks generally accepted by the data and analytics industry: the Transaction Processing Performance Council's (TPC) TPC-H specification. This specification provides the following base components for our workload comparison:

- Data schema including table structures and relationships

- Data generation tools for standardized data creation

- Base historical analytics queries

TPC-H is a decision support benchmark specification developed by the TPC in the late 1990s/early 2000s. It was designed to evaluate the performance of database systems in handling complex, ad-hoc queries on large datasets typical of business intelligence and data warehousing scenarios.

  We utilize the TPC-H specification's suite of 22 business-oriented queries. These queries operate on a structured dataset that models a realistic business scenario, including customers, orders, and products. The benchmark measures query execution time, throughput, and price-performance metrics.

Today's analytical environments must enable organizations to understand what's happened historically, what's happening today, and what's likely to happen.

Accounting for data-driven scenarios

To this specification, we add 28 queries designed to simulate real-world, data-driven scenarios. Unlike the design for the TPC-H specification, this enhanced workload definition focuses on simulating the diverse query types and workloads typical in modern, data-driven environments. This enhanced workload incorporates a mix of complex analytical queries, reporting, and operational analytics queries that reflect the demands of data-driven organizations. It aims to test a system's ability to handle concurrent, mixed workloads efficiently. By doing so, it provides a more comprehensive assessment of a database system's capabilities in supporting the dynamic needs of contemporary businesses.

Changing the question

As part of our methodology, queries are parameterized to ensure that the same data isn't requested for each round of execution. This limits the amount of data that can be cached—a typical benchmark performance tuning technique. This approach also allows for different parts of the data storage layer to be accessed.

Weighing serial vs. throughput tests

In evaluating cloud-based analytical platforms, serial and throughput tests serve distinct purposes.

  A serial test, often described as a "straight line" test—like the American motorsports concept of drag racing—executes queries sequentially, measuring individual query performance. And much like drag racing, serial tests focus on a narrow band of raw performance metrics. This approach shows how individual queries perform in a vacuum. It evaluates a system's ability to handle one query after another in a sequential or serial fashion.

  In contrast, a throughput test simulates real-world scenarios by continuously running over a set amount of time and continually changing queries. This would be like an endurance motorsports event, such as the 24 Hours of Le Mans. This approach aims to replicate a system's ability to handle concurrent operations and manage resources efficiently, making it a more appropriate evaluation to simulate production-level, real-world workloads.

teradata.

Creating our real-world workload

Building upon multiple levels of data analytics use cases and anchored in the challenges of data-driven organizations of all sizes around the globe, our real-world workload test consists of:

- 50 queries spanning multiple query types to simulate a modern, data-driven workload

- A throughput workload comparison evaluation that runs continuously for two hours, stressing all the components of a cloud-based analytics platform—from query optimizer to workload management to storage—instead of simply running through a set of queries

This diverse query mix was designed to test each platform's ability to handle complex analytical queries, operational reporting, and tactical queries in a way that reflects the demands of data-driven organizations.

Understanding costs

Cost efficiency is paramount in cloud-based analytics platforms. Organizations typically operate under strict budget constraints, making cost-effective solutions crucial for expanding analytical capabilities without increasing expenditure. These platforms must efficiently handle diverse workloads, from batch processing to real-time analytics, with costs varying significantly based on their ability to manage these varied requirements.

Furthermore, the exponential growth of data presents a challenge, as costs can escalate rapidly. Platforms that can efficiently process growing data volumes without proportional cost increases are increasingly valuable. In this context, cost-efficient analytics platforms enable companies to do more with less, optimizing their resources while meeting evolving analytical needs in a data-driven landscape.

Cloud-based analytics platforms often have fundamentally different architectures and design philosophies, making direct comparisons difficult. Each platform may have unique features, optimization techniques, and resource allocation strategies that don't directly correspond to those of competitors.

Determining base unit cost

For this workload comparison, we establish a cost equivalency for the base "unit" of processing for each system. In this evaluation, approximately $70 per compute hour is used as the target for system equivalency:

Teradata: $69.31
Snowflake: $72.00
Databricks: $72.80
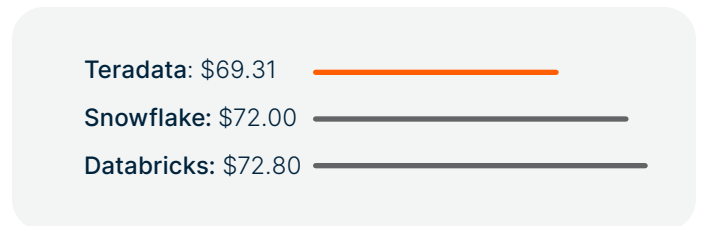
This target range reflects differences in vendor pricing and node specifications. Exact pricing to that $70 target is often not possible.

A compute hour in this context represents the hourly cost of running each cloud-based platform under the following conditions:

- Generally available software versions

- Mid-tier platform offering

- One-year service pricing commitment

Each system is configured to maximize value within this budget constraint, utilizing best practices and publicly available configuration guidance for setup.

This approach promotes a fair comparison by standardizing the cost basis while allowing each platform to leverage its strengths within a given budget. It reflects real-world choices made by decision-makers when deploying cloud-based analytics solutions.

teradata.

How we arrive at enterprise price performance

There's a simple equation for calculating enterprise price performance of a production analytics platform:

**Enterprise price performance =
Operational cost / Queries completed**

This is measured in **cost per query** and provides organizations with a view into how their platforms perform when executing production workloads.

Our study lays out the costs and query performance to show the enterprise price performance of VantageCloud against the competition. Our workload comparison evaluation cost equation will differ slightly from the one above. We'll look at the hourly cost to run our test—a substitution for **operational cost.** We'll use the average number of workload queries completed in our test—an alternative for **queries completed.**

This will give us the following cost per query equation:

**Enterprise price performance =
Hourly testing cost / Queries completed per hour**

Actual costs

While we establish a cost equivalency for the initial base "unit" of processing, the actual testing costs vary among the platforms due to their operational performance. Platforms with more aggressive scaling paradigms, like Snowflake and Databricks, quickly provision additional compute resources as processing demands increase. In contrast, Teradata—with its more sophisticated software—efficiently and fully utilizes the available compute resources at hand before asking for more. Teradata VantageCloud can do this by leveraging its patented capabilities in advanced query optimization and industry-leading workload management.

Consequently, the actual costs for the two-hour workload comparison vary significantly:

Teradata: $358.25

Snowflake: $431.25

Databricks: $513.05
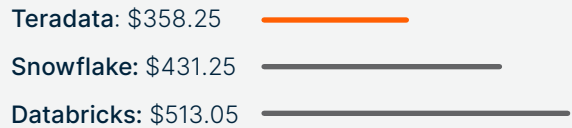
These figures reflect how each platform's scaling strategy and resource utilization efficiency impacts overall costs in a real-world scenario, despite starting from an equivalent baseline.

Queries completed

Now, we subject each platform to the previously described real-world workload over a fixed two-hour period. To simulate real-world cloud elasticity, systems are permitted to scale up to four "units" of processing to manage compute resource demands. The results, measured in average queries completed per hour, are:

Teradata: 197,366 queries

Snowflake: 3,144 queries

Databricks: 23,825 queries
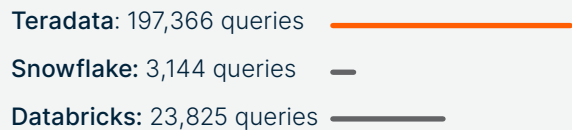
These figures demonstrate the relative performance of each platform in handling the workload, scaling effectively, and efficiently executing the query mix.

The differences in query throughput highlight varying platform capabilities in managing and optimizing resource utilization, query processing, and scalability. Teradata's higher number of completed queries suggests superior efficiency in handling this workload.

teradata.

The results: Cost per query

Next, we combine the costs and query throughput data above into our enterprise price performance equation:

🖩 **Enterprise price performance =
Hourly testing cost / Queries completed per hour**

We can now determine the enterprise price performance for each cloud-based analytics platform. This metric quantifies each platform's cost efficiency in delivering analytical capabilities and processing workloads at scale. The initial results, measured in USD per query (representing the cost efficiency of processing workloads), are:

Teradata: $0.0009 ▬

Snowflake: $0.0686 ————————

Databricks: $0.0108 ———————

To make these minute fractions of a dollar more comprehensible and to better illustrate the enterprise price performance, we normalize the costs—using Teradata's cost per query as the baseline—to $1.00. This normalization provides a clearer picture of the relative cost efficiency of each platform in handling complex, diverse queries for decision-makers across the spectrum—from business stakeholders to IT to finance.

This normalization reveals that for every dollar spent on Teradata to process the established real-world workload, an organization would need the following to equal the workload completed by VantageCloud Lake:

Teradata: $1 ▬

Snowflake: $20+ ————————

Databricks: $12 ——————
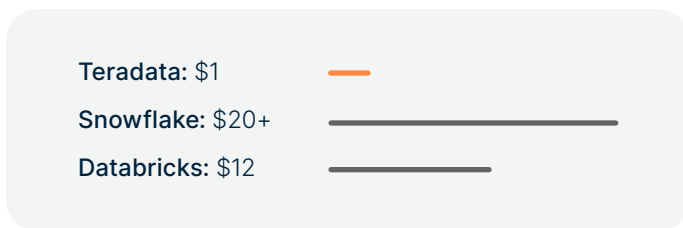
These figures starkly illustrate the significant differences in cost efficiency among the platforms, with Teradata demonstrating a substantial price-performance advantage in this real-world workload scenario.

This indicates superior cost efficiency in delivering analytical capabilities at scale. This analysis should help organizations understand the potential impact on their TCO and ROI when choosing a cloud-based analytics platform.

teradata.
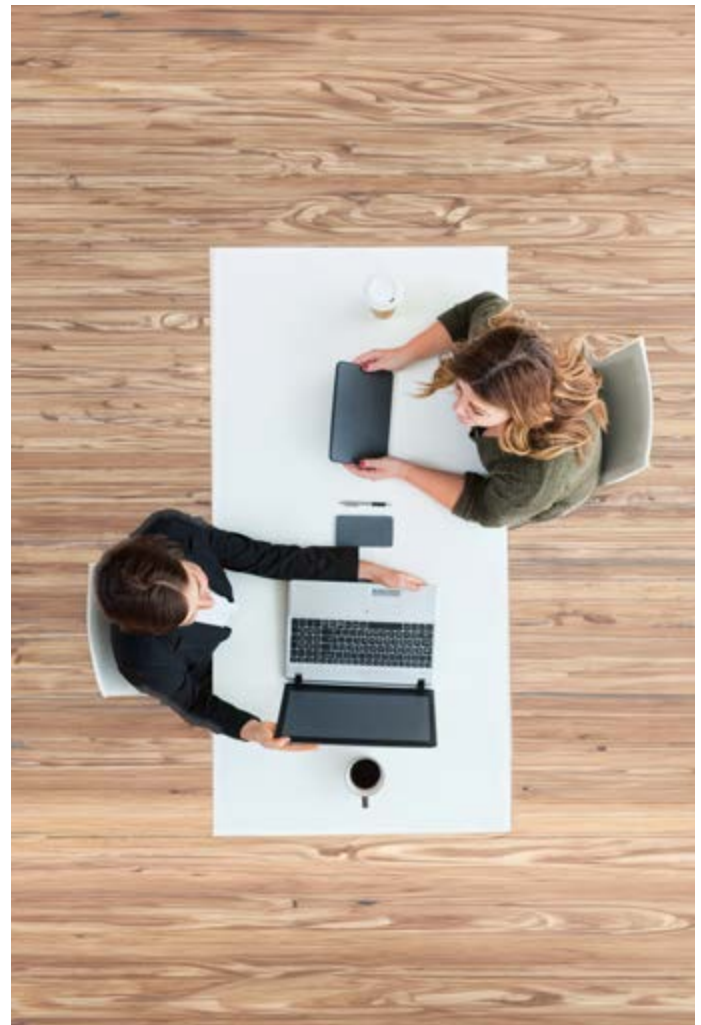
Conclusion

This study offers valuable insight into the enterprise price performance of leading cloud-based analytical platforms. Through rigorous testing with a real-world workload mix and cost-equivalent environments, we've demonstrated significant differences in efficiency, scalability, and cost effectiveness among these platforms.

Teradata VantageCloud Lake has emerged as the clear leader in this comparison, showcasing superior cost efficiency and query throughput. Its ability to process a substantially higher number of queries at a fraction of the cost of its competitors underscores the importance of sophisticated workload management and query optimization capabilities in cloud environments.

These findings highlight the importance of looking beyond marketing hype and considering holistic value propositions, including cost efficiency, scalability, and adaptability to diverse workloads. The dramatic differences in enterprise price performance underscore the potential to significantly reduce costs and improve performance through careful platform selection.

Organizations that can execute significantly more queries can analyze more data, resulting in more accurate predictions and better customer experiences. This also equips companies to proactively avoid risks like customer churn and fraudulent activity.

In the pursuit of data-driven decision-making, the cost efficiency with which insights are derived is just as crucial as the insights themselves.

As the demands on cloud-based analytics evolve, evaluations like this one can guide organizations in their digital transformations and cloud modernization journeys.

We've seen that superior software, efficient resource management, and patented workload management can deliver substantial advantages in both performance and cost effectiveness.

About Teradata

At Teradata, we believe that people thrive when empowered with trusted information. That's why we built the most complete cloud analytics and data platform for AI.

By delivering harmonized data, Trusted AI, and faster innovation, we uplift and empower our customers—and our customers' customers—to make better, more confident decisions. The world's top companies across every major industry trust Teradata to improve business performance, enrich customer experiences, and fully integrate data across the enterprise.

We drive positive impact for hundreds of millions of people every day around the world with faster, flexible data integration and trusted, cost-effective AI innovation. **Speak to an expert** to learn how Teradata can improve your data analytics today and learn more about our **workload comparisons**.

---

teradata.