

Choose the Analytic Solution to Meet Business Requirements



Table of Contents

- 3 Choosing the Right Analytics Platform
- 3 Setting the Benchmark
- 4 High-Level Results Show Distinct Differences
- 5 Results Document Real-World Benefits
 - 5 **Tactical Queries**
 - 6 **Strategic Queries**
- 6 At the Scale of Business
- 7 Key Takeaways

Selecting the right software and hardware to support analytics and deliver answers can be confusing. There are a lot of options, both for on-premises and in the cloud. Every vendor makes a variety of claims designed to make its products stand out, which only adds to the murkiness of identifying the real business value.

Companies need to understand how they can rise above the complexity, cost, and inadequacy of today’s analytics landscape. Not all options are the same because not all software and hardware are created equally. Each option may sound like it has unique benefits, although what it ultimately delivers may not meet organizations’ current and future analytic needs, such as delivering answers to complex business problems. To meet those needs, analytic software, hardware, and solutions must enable high-value business outcomes by managing all of the data, all of the time, so businesses can analyze anything, deploy anywhere, and deliver answers that matter. They must also perform at scale.

To help organizations make an informed buying decision, Teradata conducted a benchmark that showed how software solutions compare. The benchmark evaluated the actual query performance, cost/performance, and business implications of the Teradata analytics platform software across different deployment options and against Apache Hadoop® and a leading cloud analytic database.

The benchmark helps businesses understand each analytic environment and how it performs. This allows organizations to make informed decisions about when to use each environment and what to expect from each one.

Choosing the Right Analytics Platform

Organizations require a modern analytics platform that is easy to use, delivers rapid answers, provides agility and scalability, analyzes all of their data, and allows a variety of analytic techniques to address even the most challenging use cases.

The analytics platform must offer deployment options to let companies choose where they run their analytics. After making that choice, they should be able to move the platform as their needs change so they're not locked into a single deployment option. This flexibility allows the analytics solution to move across the cloud, on-premises, and hybrid solutions.

A significant requirement for this flexibility is having the analytics solution deliver high performance and the same features and capabilities regardless of where it's used. This enables data models, applications, and queries to run unchanged across platforms. As a result, businesses benefit from seamless application portability that de-risks the investment and future-proofs buying decisions.

Setting the Benchmark

The environment for the benchmark simulated real-world customer workloads with data volume, concurrency, mixed workloads, and query types. All platforms used the same set of data, queries, and concurrency. Figure 2 shows the benchmark criteria.

Five platforms were tested. The three running Teradata software used eight nodes each. Hadoop and a leading cloud database were given four times as many—32 nodes each—to show whether hardware can compensate for software performance.

The platforms used in the benchmark were:

- **Teradata IntelliFlex**—designed specifically for Teradata analytic workloads. It has a mix of solid-state drives (SSDs) and hard disk drives (HDDs) as well as an I/O rich configuration. It is purpose-built for mission-critical operations.

Analytics at Scale Highlights Cost and Performance of Platforms

Time to process 1 million queries, based on benchmark results.

Time	Value
Less than 10 minutes for Teradata on IntelliFlex and less than 20 minutes for Teradata on AWS	 
More than 2 weeks for Apache Hadoop	
9 months for the leading cloud database	

The cost to process those 1 million queries would be:

Cost	Value
Less than \$60 for Teradata	 
\$42,700 for Apache Hadoop	
\$607,000 for the leading cloud database	

Figure 1.

- **Teradata Software for AWS**—with an SSD-only configuration, high I/O, powerful compute resource, and persistent attached storage.
- **Teradata IntelliBase**—an entry-level, HDD-only appliance that provides a fully-integrated environment for data warehousing, iterative data exploration, and low-cost data storage.
- **Hadoop with Apache® Hive™**—configured with the largest available hardware instance of Hadoop in the cloud environment.
- **A leading cloud database**—an SSD-only configuration using the largest available hardware instance in the cloud.

High-Level Results Show Distinct Differences

The benchmark showed that Teradata software performs consistently well on all hardware infrastructures. Differences in performance can be attributed to differences in the underlying hardware configuration used in the benchmark.

The benchmark also revealed that Teradata software platforms outperformed the alternatives, regardless of the underlying infrastructure. Teradata delivered a higher total number of queries, more queries per hour, the best cost per query, and the shortest time required to run a specific number of queries (see Figure 3).

During the 10-hour benchmark, Teradata Database (which is now evolved to Teradata Vantage as Teradata’s integrated, advanced analytics platform) on Teradata IntelliFlex ran more than 67,000,000 queries. During that same time, Hadoop ran fewer than 30,000 queries with some simply failing to run because they either timed out or could not be parsed. The leading cloud database ran fewer than 2,000 queries in the same period.

The Teradata solution performed more than 35,000 times better than the leading cloud database (see Figure 4). Teradata IntelliFlex delivered the highest performance

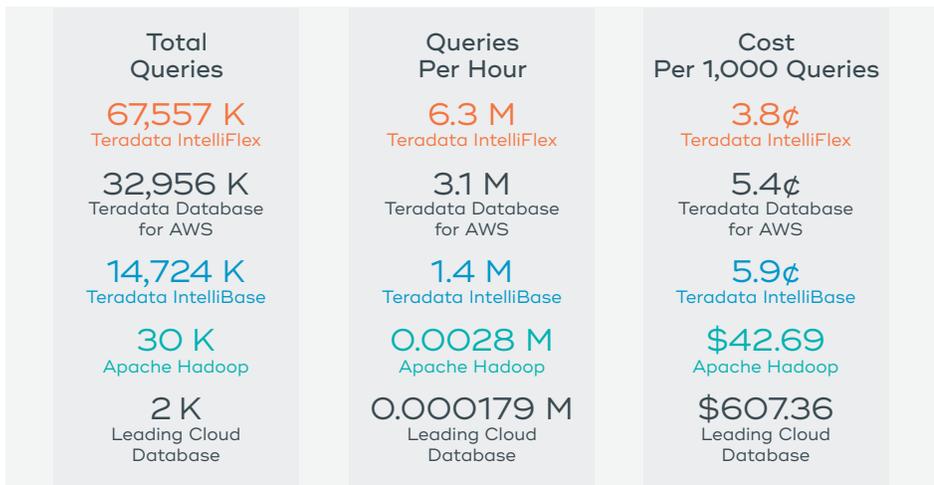


Figure 3. Benchmark Results Comparison

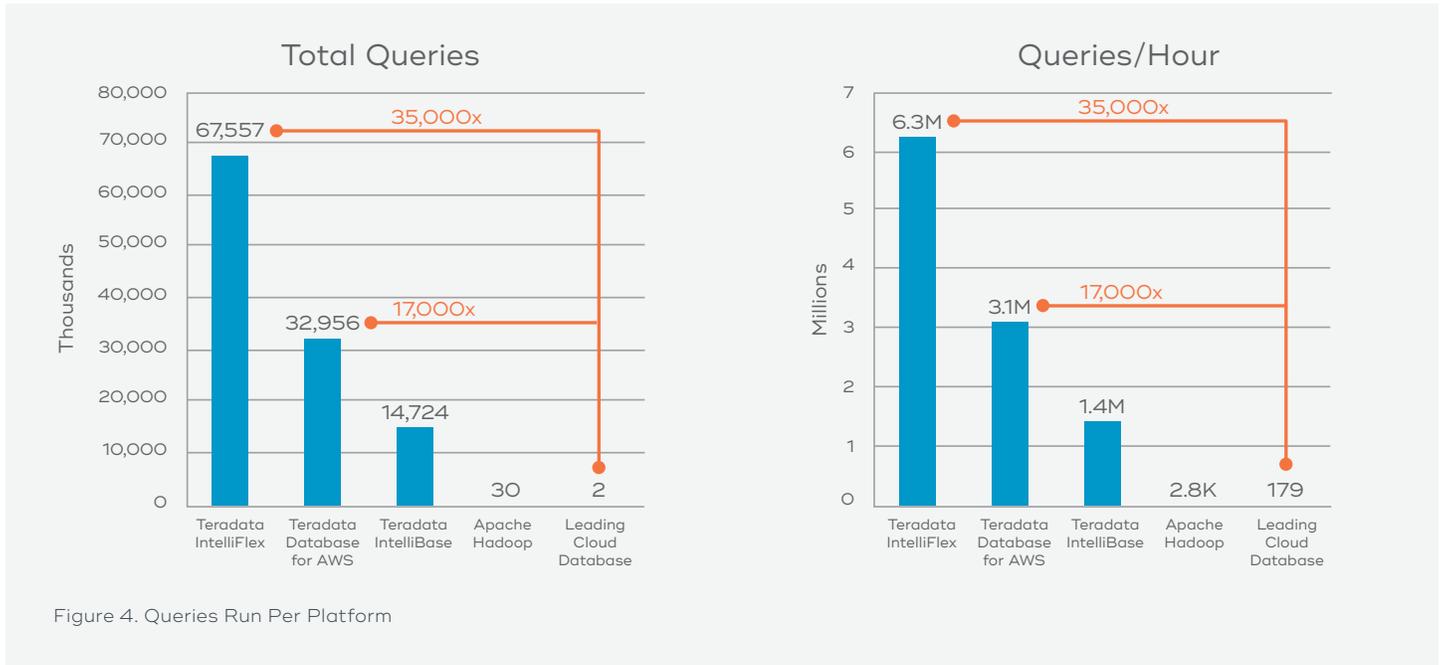
Schema	Based on industry standard schema, unmodified
Data	Retail customer and order data: 8 tables, 24 terabytes
Strategic Queries	19 queries
Tactical Queries	5 queries
Query Processing	Defined order with randomly generated parameters
Concurrency	Strategic: Up to 50 sessions Tactical: Up to 300 sessions

Figure 2. Benchmark Criteria

at 6,300,000 queries per hour. The leading cloud database was the lowest at just 179 per hour.

When comparing costs, the cost per query is a more relevant, and therefore a better, metric than acquisition cost alone or “cost-per-hour” because it’s what organizations pay to get answers to their business questions.

The differences in this cost metric were just as significant as the performance. The leading cloud database is 15,000 times more expensive per query than Teradata. The cost to run 1,000 queries was \$0.03 for Teradata IntelliFlex. The cost was \$40 for Hadoop, and \$600 for the leading cloud database (see Figure 5). Likewise, Teradata also proved to run queries significantly faster than competitors. The callout on Page 3 shows that what would take Teradata IntelliFlex only 10 minutes



to run would take Hadoop two weeks and the leading cloud database nine months.

Results Document Real-World Benefits

Companies deploy an analytic environment for a purpose. They expect it to serve a number of users with the analytics they need in the timeframe they’re needed. Ideally, it also allows analysts to use their preferred tools and languages across any data type.

The benchmark validates how analytics platforms meet real-world business needs and how users interact with their analytic system. The right analytics platform enables high-impact business outcomes by using all data to provide answers to any business question, no matter how complex. A low-performing platform limits outcomes while escalating costs.

Tactical Queries

Analytic solutions must support practical business operations by responding to simple but frequent queries in a “speed of thought” manner. These types of queries require sub-second response time to deliver personalized offers, provide timely answers to call center agents, and feed executive dashboards.

A query that took Teradata IntelliFlex 0.03 seconds took Hadoop 3 minutes and 50 seconds. Another query took Teradata IntelliFlex 0.24 seconds but took

Deployment Type Benefits

Each Teradata® software deployment option offers unique advantages:

- Teradata Cloud leverages an infrastructure specifically engineered for Teradata software, providing maximum scalability and availability for mission critical workloads.
- Public cloud options deliver convenience and on-demand resources. Teradata software is available via Amazon Web Services (AWS) and Microsoft® Azure®.
- Teradata IntelliFlex® is an on-premises solution optimized for performance, throughput, scalability, and mission critical availability.
- Commodity hardware is available on-premises. Organizations can deploy Teradata software on a commodity infrastructure using VMware.

Hadoop 5 minutes and 29 seconds (see Figure 6). This means the latter performance is not suitable for tactical queries, nor is it scalable in volume across the organization. Figure 2 shows the benchmark criteria that simulated real-world mixed workloads.

Strategic Queries

Analytics platforms must run queries that support strategic analysis. These queries answer complex questions and are iterative in nature—when the answer to one query spawns new questions that need to be answered in a reasonable time.

For these types of queries, the benchmark showed Teradata IntelliFlex performs much better than competitors. For example, a query that included multi-table joins with complex conditions was served by Teradata IntelliFlex in 15 minutes, while Hadoop took 14 hours, only to time out at the end without an answer (see Figure 7).

At the Scale of Business

Companies that run on analytics run at scale. Query volumes spanning thousands to millions a day are par for the course, along with hundreds to thousands of concurrent users.

Based on the benchmark results, processing 1 million queries would take:

- Less than 10 minutes for Teradata IntelliFlex
- Less than 20 minutes for Teradata Software for AWS
- More than 2 weeks for Hadoop
- 9 months for the leading cloud database

The cost to process those 1 million queries would be:

- Less than \$60 for Teradata
- \$42,700 for Hadoop
- \$607,000 for the leading cloud database

Cost to Run 1,000 Queries



Figure 5. Cost of Analytics



Consider the Hidden Costs

In addition to acquisition costs and the cost to run queries, some platforms have additional expenses that should be considered when building and maintaining an analytic environment. Non-Teradata systems may face these extra costs:

- Acquiring skilled resources to run the platform.
- Longer implementation time to set up less mature technology.
- Missed opportunities due to an inability to get timely results.
- Unpredictable performance.
- Instability across the platform.
- Poor tool integration and cohesiveness.
- Lack of portability results in costs to re-code if moving to a different environment.
- Limited sophisticated features restrict business insights.

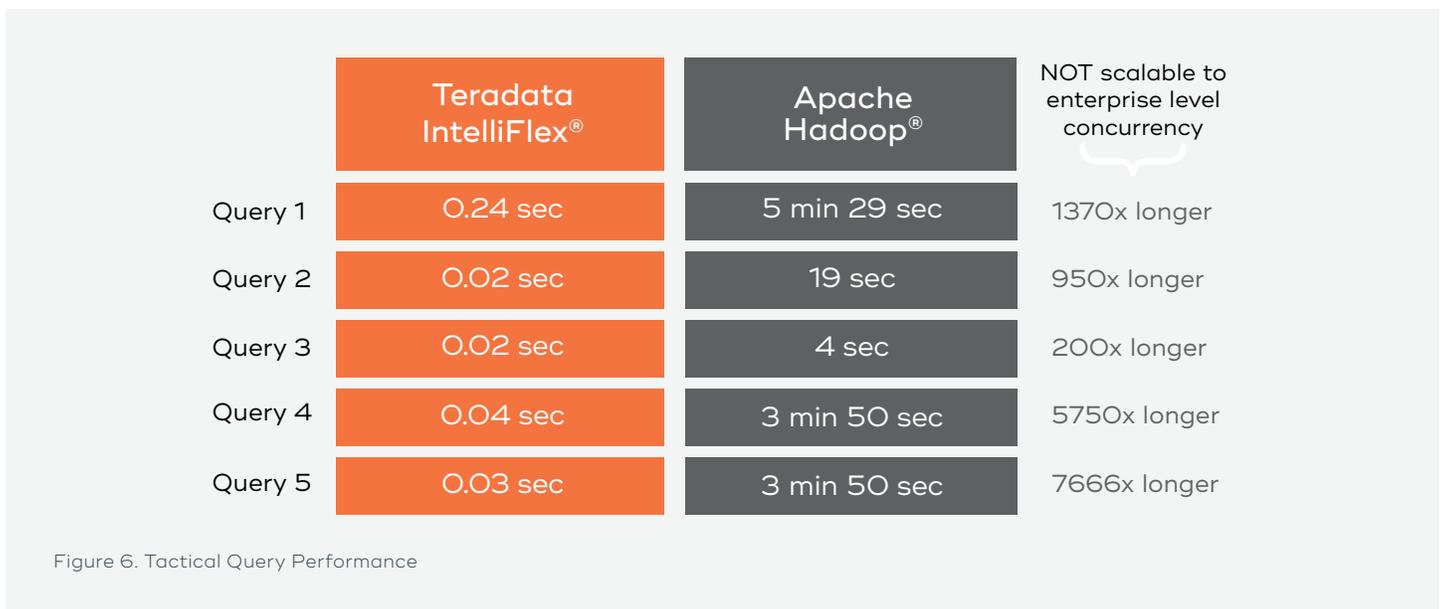
Key Takeaways

The key takeaways from the benchmark include:

- Organizations cannot successfully run their business using analytics that do not align with their performance or budget requirements.
- Teradata provides the highest-performance analytic solution that scales regardless of where it runs.
- Teradata provides the most cost-effective solution regardless of where it runs.
- Companies have a choice of where to deploy their analytics based on:
 - Business criticality
 - Query volume and throughput
 - Cloud mandate
 - Data gravity

Organizations need to choose an analytics platform based on their current and future needs. Teradata solutions give organizations the highest performance, the best cost per query, and the flexibility to move the solution across deployment options as their business needs change.

For more information, visit Teradata.com/vantage.



A benchmark comparison shows the overall cost and performance of analytic platforms to help organizations choose the one that meets their needs

Teradata Vantage Future-Proofs the Business

Teradata Vantage helps you meet your data and analytic needs today and in the future by delivering flexibility, agility, scalability, and choice across four high-level capabilities:

Analyze Anything—Enables analytic users throughout the organization to use their preferred analytic tools and engines across data sources, at scale.

Deploy Anywhere—Provides analytic processing across flexible deployment options, including AWS, Azure, and Teradata Cloud as well as on-premises on Teradata IntelliFlex or commodity hardware.

Buy Any Way—Empowers companies to purchase software in more accommodating ways based on specific use cases through simplified pricing bundles, subscription-based licenses, and as-a-service options.

Move Anytime—Future-proofs buying decisions by taking advantage of our software license portability that provides flexibility to run analytics across deployment options.

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

Hive is a trademark, and Apache and Hadoop are registered trademarks of the Apache Software Foundation in the United States and other countries. Amazon Web Services and AWS are trademarks of Amazon.com, Inc. or its affiliates in the United States and/or other countries. Microsoft and Azure are registered trademarks of Microsoft Corporation. Teradata and the Teradata logo are registered trademarks of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or Teradata.com for more information.

© 2018 Teradata Corporation All Rights Reserved. Produced in U.S.A. 10.18 EB9908

