COMMISSIONED BY

ACTIAN
a division of HCLSoftware

MCKNIGHT CONSULTING GROUP

DATA WAREHOUSE PERFORMANCE

PRODUCT EVALUATION:

Actian Data Platform, Google BigQuery, Databricks and Snowflake

RELATED RESEARCH

Enterprise data integration landscape: Top capabilities and vendors exposed

MCG Stream Processing Ease of Use and TCO

William McKnight

Art Vancil

Prepared by

MCKNIGHT CONSULTING GROUP

McKnight Consulting Group
www.mcknightcg.com
September 2024

Contents

Executive Summary

Data-driven organizations rely on analytic databases to load, store, and analyze volumes of data at high speed to derive timely insights. Data volumes within modern organization's information ecosystems are rapidly expanding—placing significant performance demands on legacy architectures. Today, to fully harness their data to gain competitive advantage, businesses need modern scalable architectures and high levels of performance and reliability to provide timely analytical insights.

This report outlines the results from a McKnight Consulting Group Analytic Field Test derived from the industry standard TPC Benchmark™ H (TPC-H)[1] to compare Actian Data Platform, Google Big Query, Databricks and Snowflake. This test produced interesting results that reveal some of the performance characteristics of the four platforms.

With comparable specifications, Actian performs almost 8 times faster than Databricks, over 6 times faster than Snowflake, and over 12 times faster than BigQuery with TPC-H-like tests. When 5-user concurrency is considered, Actian performs 3 times better than Databricks, over 7 times better than Snowflake, and 9.6 times better than BigQuery.

Actian provided a price-performance that was over 8 times better than that of Snowflake and BigQuery.

These results suggest that Actian is a great choice for anyone looking to access large analytic data sets quickly and economically. Additionally, given the significant speed and cost savings that Actian provides, it is an excellent solution for organizations with large complex data sets that need to be accessed quickly and affordably.

[1] More can be learned about the TPC-H benchmark at http://www.tpc.org/tpch/.

Platform Summary

Big data analytics platforms load, store, and analyze volumes of data at high speed, providing timely insights to businesses. Data-driven organizations leverage this data, for example, for advanced analysis to market new promotions, operational analytics to drive efficiency, or for predictive analytics to evaluate credit risk and detect fraud. Customers are leveraging a mix of relational analytical databases and data warehouses to gain analytic insights.

This report focuses on relational analytical databases in the cloud, because deployments are at an all-time high and poised to expand dramatically. The cloud enables enterprises to differentiate and innovate with these database systems at a much more rapid pace than was ever possible before. The cloud is a disruptive technology, offering elastic scalability vis-à-vis on-premises deployments, enabling faster server deployment and application development, and allowing less costly storage. For these reasons and others, many companies have leveraged the cloud to maintain or gain momentum as a company.

This report compares Actian Data Platform, Google Big Query, Databricks and Snowflake—relational analytical databases based on scale-out cloud data warehouses and columnar-based database architectures. Despite these similarities, there are some distinct differences in the platforms.

Actian Data Platform

Actian Data Platform is based on its underlying technology, known as Vector. The basic architecture of Actian Data Platform is the Actian patented X100 engine, which utilizes a concept known as "vectorized query execution" where processing of data is done in chunks of cache-fitting vectors. Vector performs “single instruction, multiple data” processes by leveraging the same operation on multiple data simultaneously and exploiting the parallelism capabilities of modern hardware. It reduces overhead found in conventional "one-row-at-a-time processing" found in other platforms. Additionally, the compressed column-oriented format uses a scan-optimized buffer manager.

The measure of Actian Data Platform compute power is known as Actian Units (AU). At the time of this writing, the Actian Data Platform is priced at $2.50 per AU per hour. This price includes both compute and cluster storage.

Google BigQuery

Google BigQuery is a managed service with some interesting distinctions. Google abstracts the details of the underlying hardware, database, and all configurations. BigQuery is a hands-off database without indexes or column constraints. Defragmentation and system tuning are not required. It is serverless. Google Cloud manages the servers in a fully hands-off manner to the customer, dynamically allocating storage and compute resources. The customer does not define nodes and capacity of the BigQuery instance. The provisioning of compute is particularly fast and seamless.

You pay for the amount of data you query and store. Customers can pre-purchase computation "slots" for as short as one minute and billed by the hour. There is a separate charge for active storage of data.

Databricks

Databricks is a unified analytics platform that combines data warehousing, data lakes, and machine learning capabilities. It offers a scalable and collaborative environment for data teams to analyze, process, and extract insights from large datasets.

Databricks uses a pay-as-you-go pricing model, charging based on the resources consumed, such as compute power, storage, and network bandwidth. The exact pricing depends on various factors like the type of workload, region, and usage patterns.

Snowflake

As a cloud-only, fully managed solution, Snowflake has a clear separation between compute and storage. For Snowflake on GCP, which is what we used for the queries, data is stored in GCP GCS and is cached when queries are executed to bring the data in closer proximity to compute resources. Snowflake essentially offers two configuration "levers" — the size of the warehouse cluster and how many clusters are permitted to spin up to handle concurrency. Snowflake scales by cluster server count in powers of 2 (i.e., 1, 2, 4, 8, 16, and so on.) If enabled, Snowflake will spin up additional clusters to handle multi-user concurrent query workloads. Snowflake would automatically spin the additional clusters down once demand has passed. If not enabled, it will place paused queries in a queue until resources free up.

For Snowflake, you pay a flat hourly fee for when hourly compute resources are being used. We paid $3.00 per hour for the Enterprise tier. Once the compute warehouse goes inactive, you no longer pay. However, there is a separate charge for data storage.

Test Setup

The setup for this Field Test was informed by the TPC Benchmark™ H (TPC-H)[2] spec validation queries. This is not an official TPC benchmark. The queries were executed using the following setup, environment, standards, and configurations.

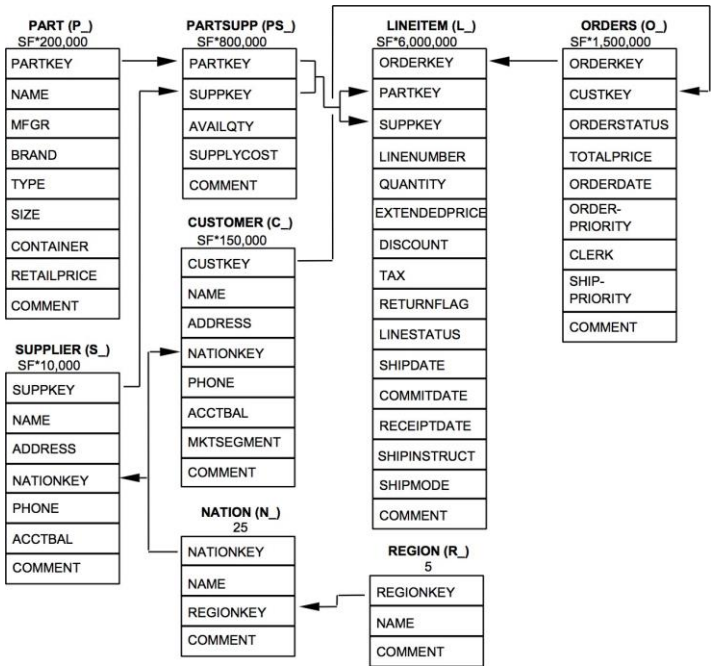
Benchmark Data

The data sets used in the benchmark were a workload derived from the well-recognized industry standard TPC Benchmark™ H (TPC-H).

From tpc.org: "The TPC-H is a decision support benchmark. It consists of a suite of business-oriented ad-hoc queries and concurrent data modifications. The queries and the data populating the database have been chosen to have broad industry-wide relevance. This benchmark illustrates decision support systems that examine large volumes of data, execute queries with a high degree of complexity, and give answers to critical business questions."

To show the data model, the following diagram was taken from page 13 of the TPC-H Revision 2.17.3 specification document.

Figure 1. TPC-H Data Model



[2] More can be learned about the TPC-H benchmark at http://www.tpc.org/tpch/.

To give an idea of the data volumes used in our benchmark, the following table gives row counts of the database when loaded with 30TB of TPC-H data:

Table 1. TPC-H Database Row Count given 30TB

TPC-H Table	30TB Row Count
Customer	4,500,000,000
Line Item	180,000,000,000
Orders	45,000,000,000
Part	6,000,000,000
Supplier	300,000,000
Part Supp	24,000,000,000

Cluster Environments

Our benchmark included the following cluster environments:

Table 2. Cluster Environments

TPCH 30TB	Actian	Snowflake	BigQuery	Databricks
Tier	Enterprise	Enterprise	Enterprise	Serverless
Size	128AU	4X-Large	4XL	2X-Large
Units	128 AUs	128 nodes	4,800 Slots	144 DBUs

Queries

We sought to replicate the TPC-H Benchmark queries modified only by syntax differences required by the platforms. The benchmark is a fair representation of enterprise query needs. The TPC-H testing suite has 22 queries, which are described by the table below.

Table 3. TPC-H Query Parameters

Q#	Description	Sum	Sub-query	Join	Min/ Max	Avg	Count	Top/ Limit
1	Pricing Summary Report	✓				✓	✓	
2	Minimum Cost Supplier		✓	✓	✓			✓
3	Shipping Priority	✓		✓				✓
4	Order Priority Checking		✓	✓			✓	
5	Local Supplier Volume	✓		✓				
6	Forecasting Revenue Change	✓						
7	Volume Shipping	✓	✓	✓				
8	National Market Share	✓	✓	✓				
9	Product Type Profit Measure	✓	✓	✓				
10	Returned Item Reporting	✓		✓				✓
11	Important Stock Identification	✓	✓	✓				
12	Shipping Modes and Order Priority	✓		✓				
13	Customer Distribution		✓	✓			✓	
14	Promotion Effect	✓		✓				

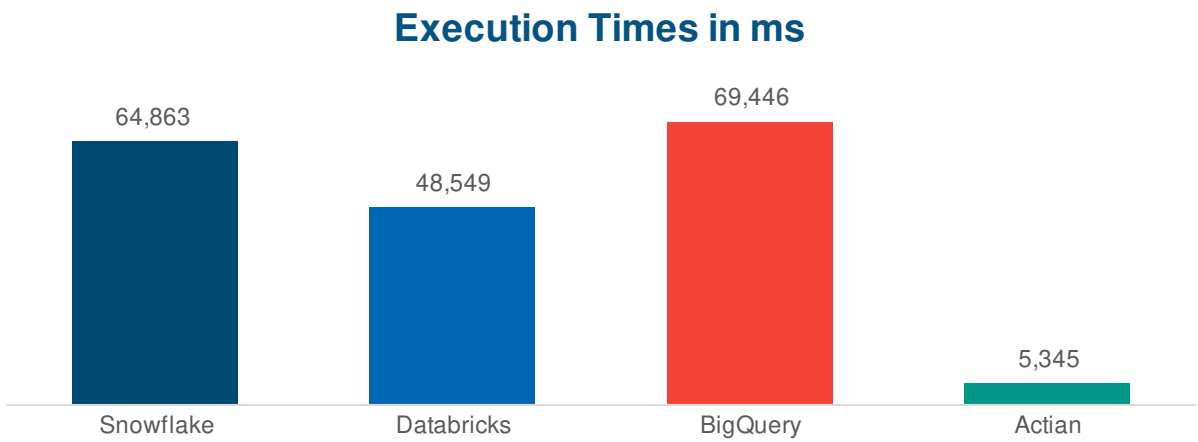
15	Top Supplier		✓	✓	✓			
16	Parts/Supplier Relationship		✓	✓			✓	
17	Small Quantity Order Revenue	✓	✓	✓		✓		
18	Large Volume Customer	✓	✓	✓				✓
19	Discounted Revenue	✓		✓				
20	Potential Part Promotion	✓	✓	✓				
21	Suppliers Who Kept Orders Waiting		✓	✓			✓	✓
22	Global Sales Opportunity	✓	✓	✓		✓	✓	

Test Results

This section analyzes the query results from the fastest runs of the three sets of 22 TPC-H queries[3] described in Table 4.

Tests with 1 user (no concurrency)

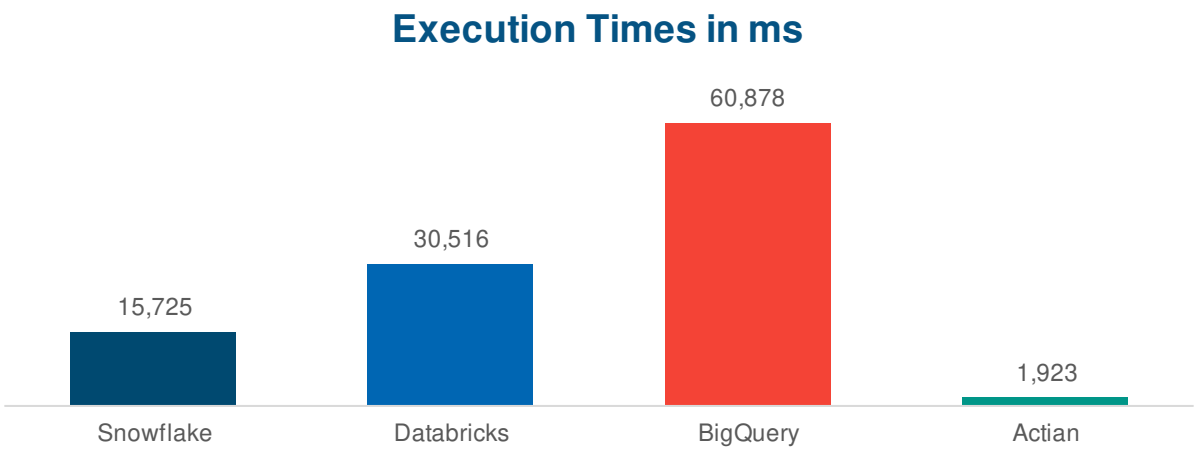
Figure 2. TPC-H Query 1: "Pricing Summary Report" Execution Times



Query 1 is the only query that uses only Sum, Average, and Count. Actian well outperformed the field and BigQuery was the slowest.

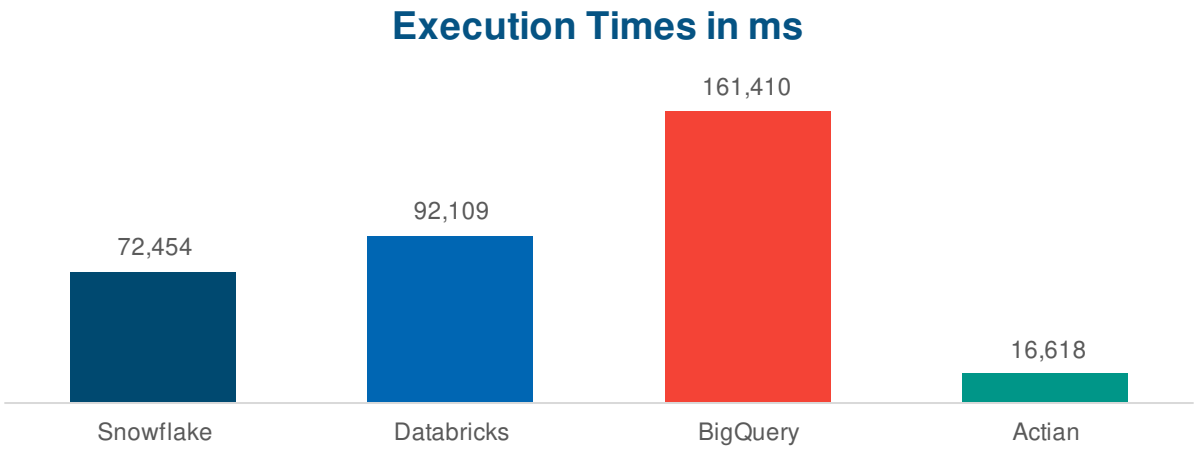
[3] The order of the queries for the power run and 5 stream runs was in the randomized order suggested by TPC in TPC Benchmark H Appendix A.

Figure 3. TPC-H Query 2: "Minimum Cost Supplier" Execution Times



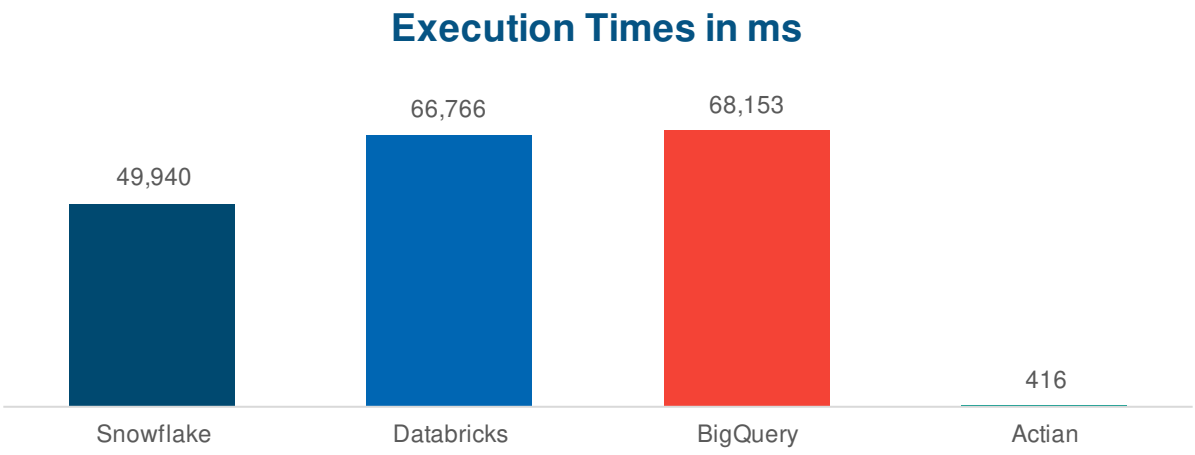
Query 2 was one of two queries that contained a Min/Max. Actian was the fastest by over 8 times the second fastest, Snowflake.

Figure 4. Query 3: "Shipping Priority" Execution Times



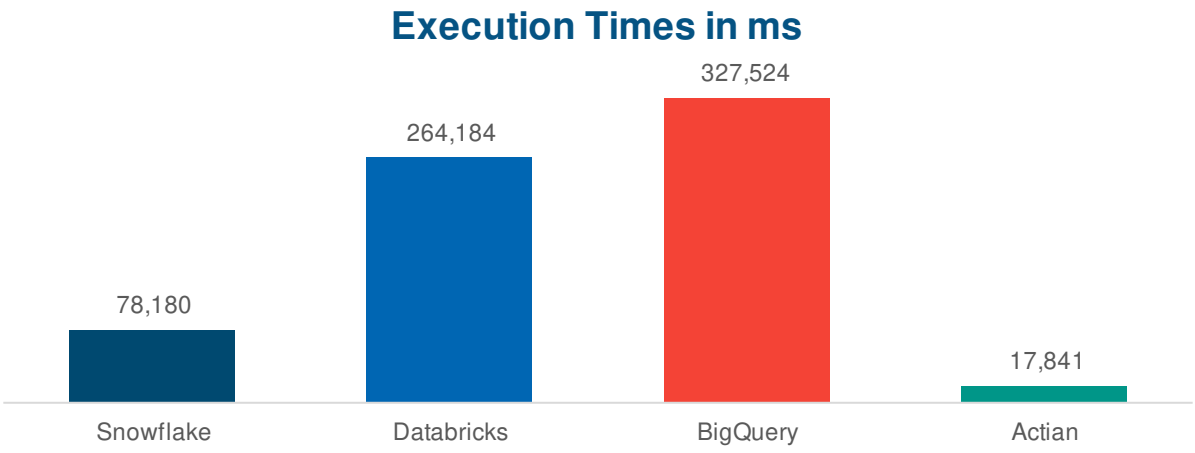
Actian was over 4 times as fast as Snowflake, with BigQuery the slowest. Most queries performed in this order.

Figure 5. Query 4: "Order Priority Checking" Execution Times



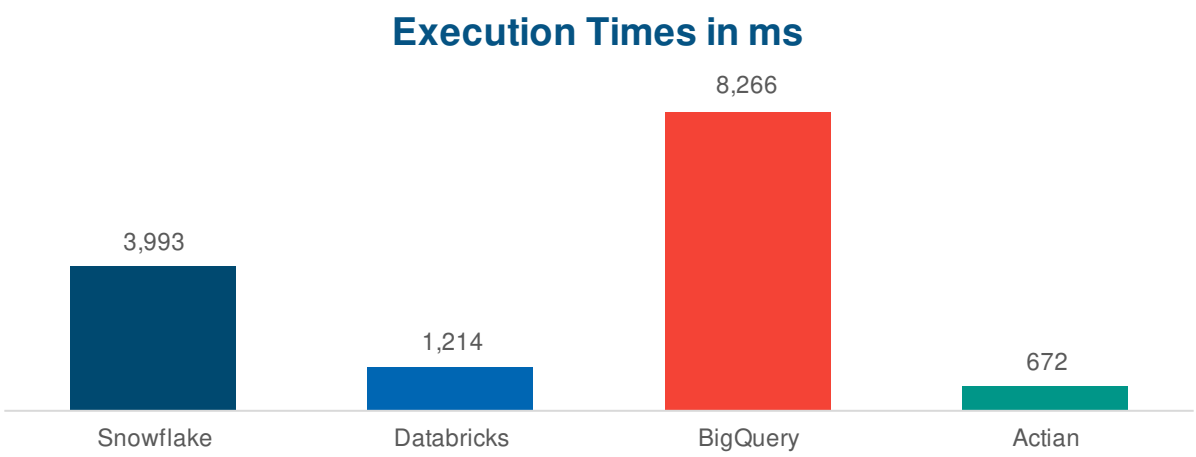
In Query 4, with a subquery and a count, we see the familiar pattern of Actian being the fastest, this time by over 100 times to second place Snowflake.

Figure 6. Query 5: "Local Supplier Volume" Execution Times



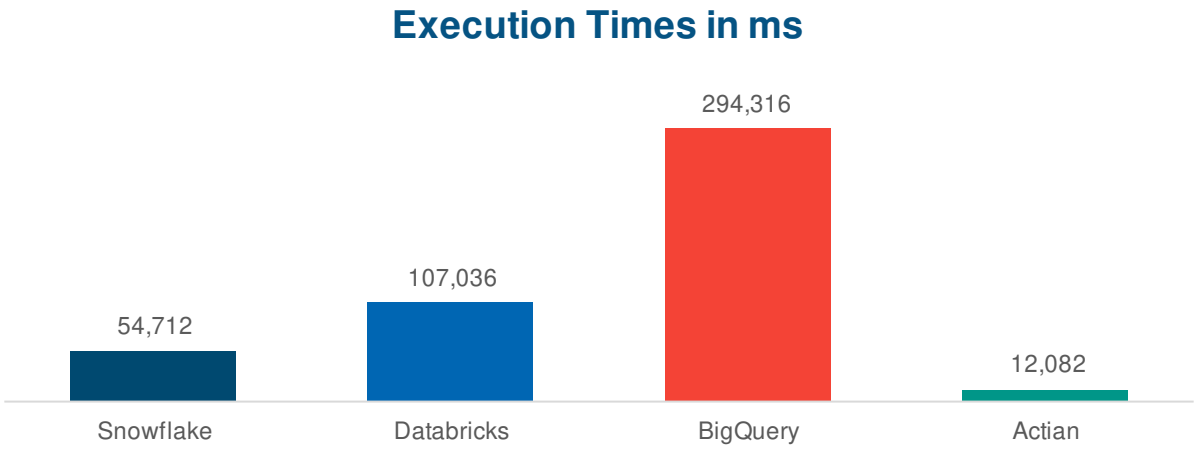
Query 5, which only employs a sum aggregation, favored Actian as well.

Figure 7. Query 6: "Forecasting Revenue Change" Execution Times



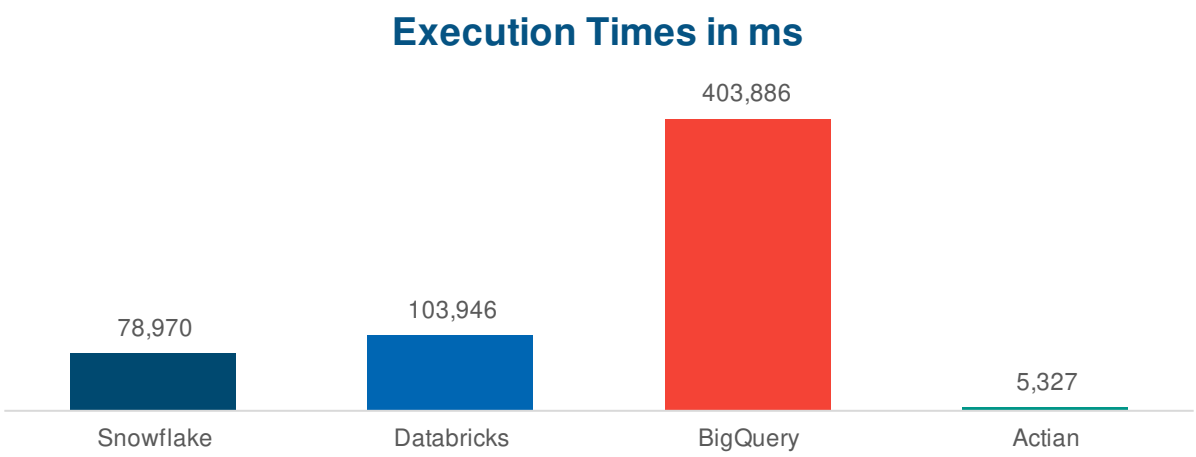
The simple SUM of Query 6 was a very high relative performer for Actian.

Figure 8. Query 7: "Volume Shipping" Execution Times



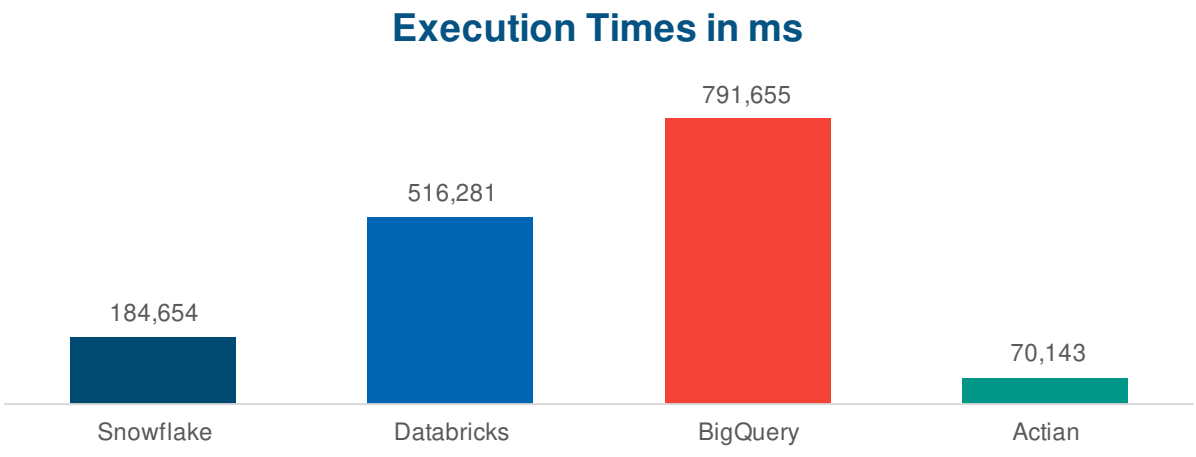
Actian was a super strong performer on this query as well.

Figure 9. Query 8: "National Market Share" Execution Times



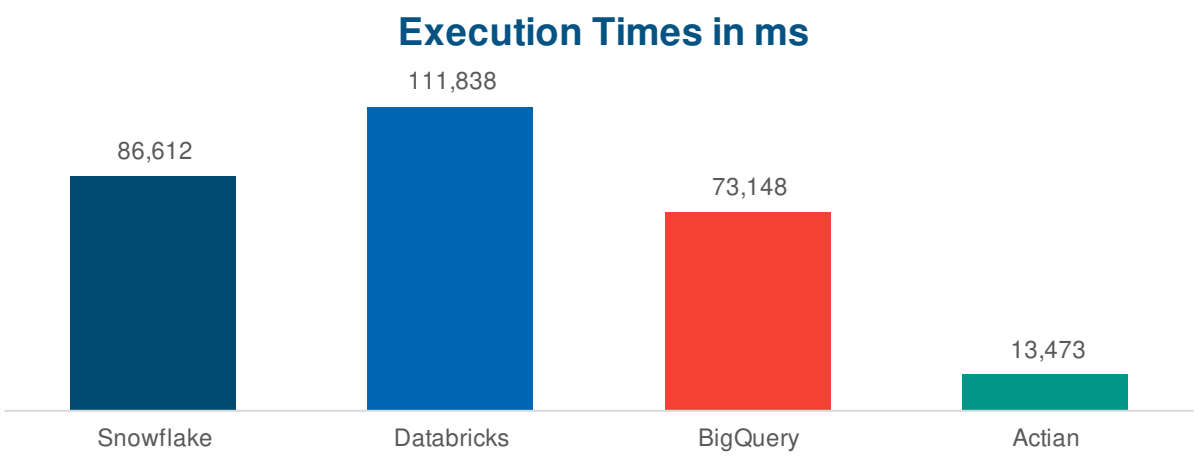
Query 8 performance favored Actian over Snowflake by 14 times. The gaps are actually wider than our 2022 report.

Figure 10. Query 9: "Product Type Profit Measure" Execution Times



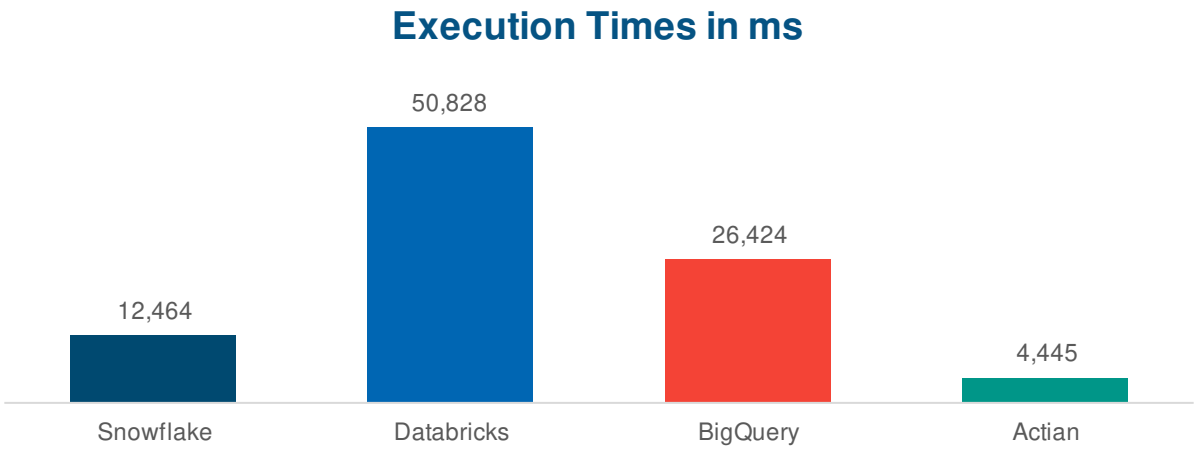
Snowflake was a relatively close only 2.6 times slower than Actian on this query.

Figure 11. Query 10: "Returned Item Reporting" Execution Times



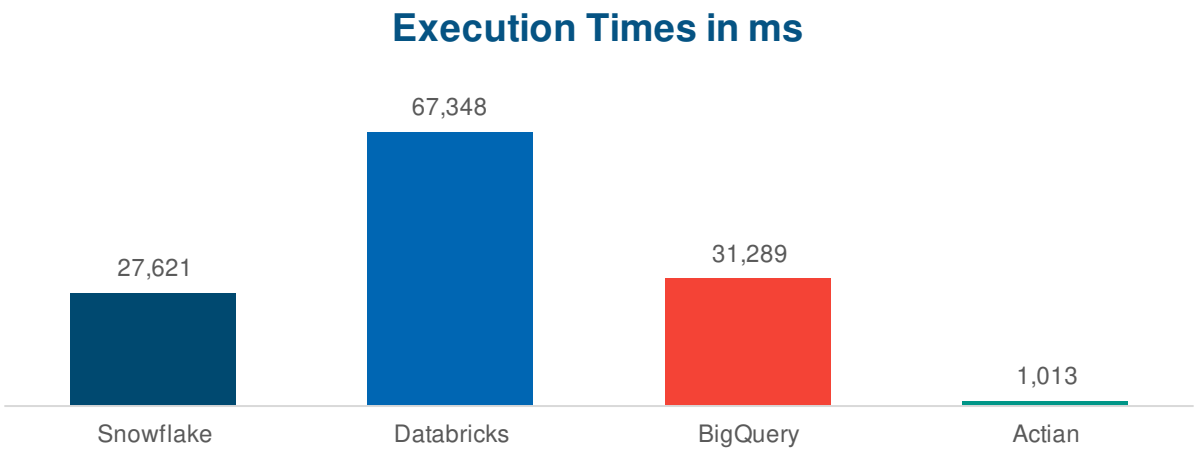
Query 10 uniquely has a Sum and a Top/Limit. Actian was the top performer.

Figure 12. Query 11: "Important Stock Identification" Execution Times



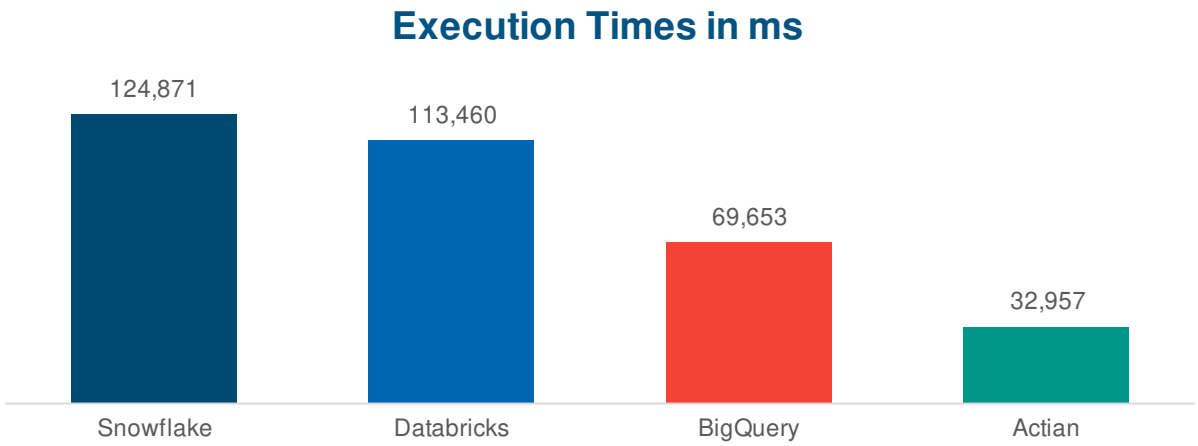
Query 11 (another sub-select and a sum operation) saw performance slightly favoring Actian by 2.8x over Snowflake.

Figure 13. Query 12: "Shipping Modes and Order Priority"



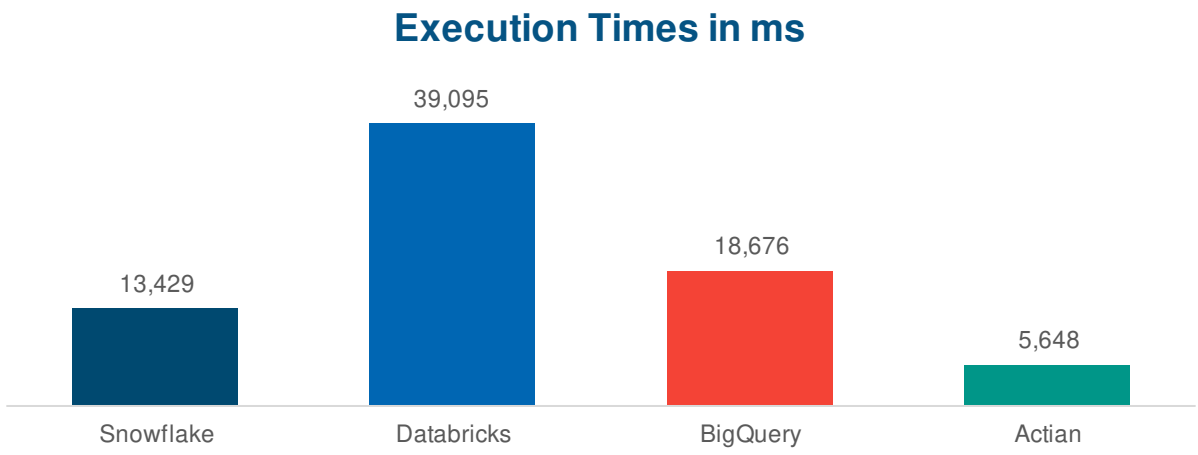
Query 12, another simple SUM, was another plus performance query for Actian, with Actian outperforming the analogous Snowflake configuration by 27 times.

Figure 14. Query 13: "Customer Distribution" Execution Times



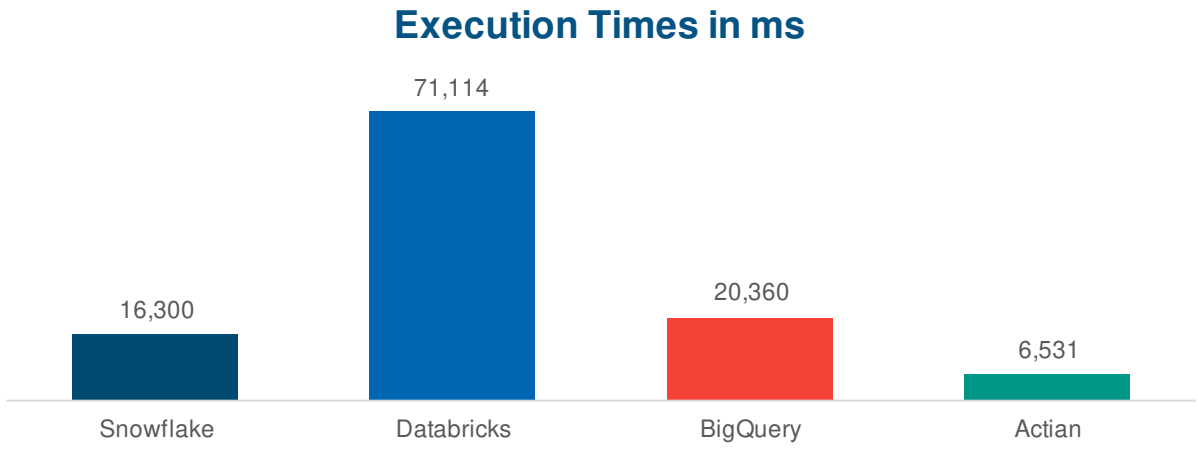
Query 13 is the only TPC-H query with an explicit JOIN. Actian outperformed the field, and this time Snowflake was the slowest.

Figure 15. Query 14: "Promotion Effect" Execution Times



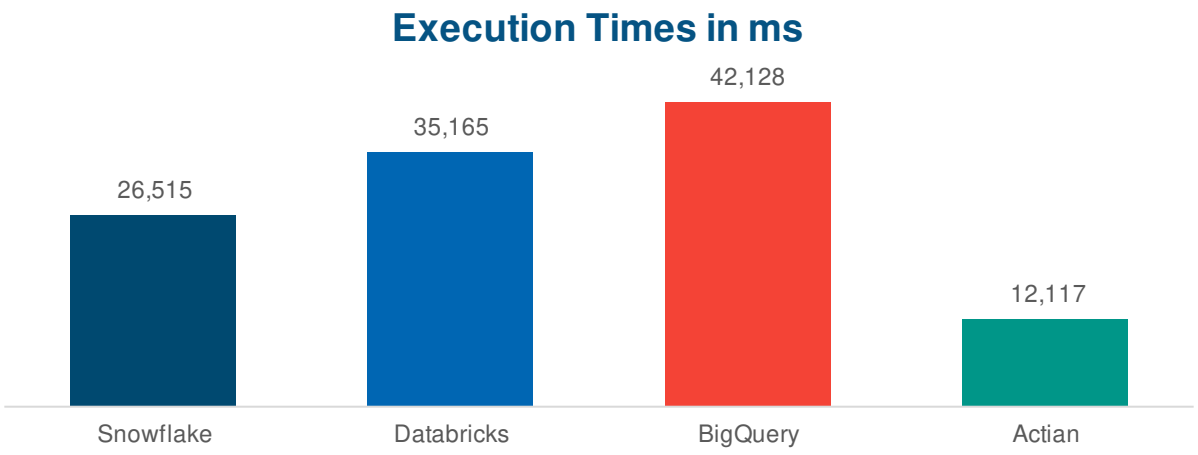
For Query 14, Actian took only an impressive 5,648 ms.

Figure 16. Query 15: "Top Supplier" Execution Times



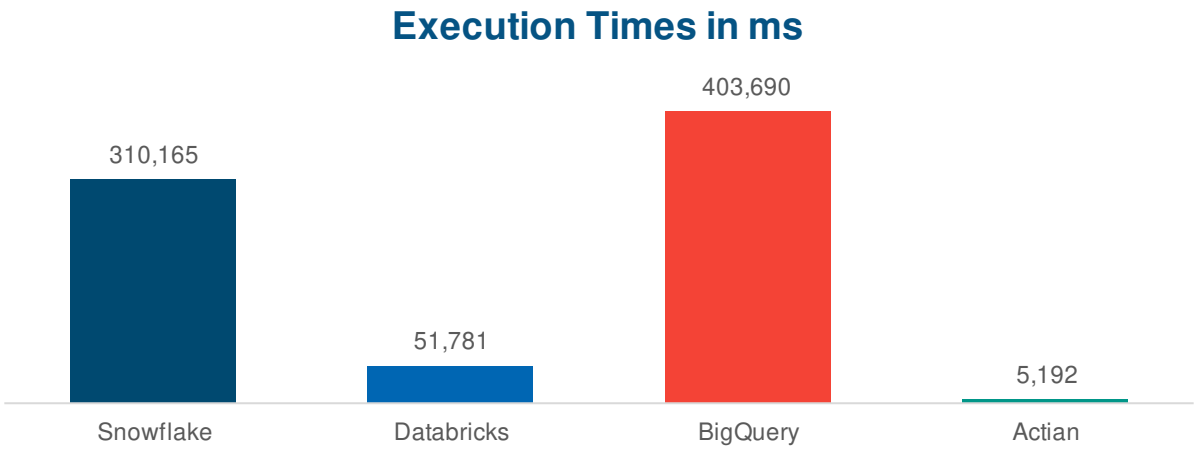
Snowflake was relatively close on this query, slower than Actian by only 2.5 times.

Figure 17. Query 16: "Parts/ Supplier Relationship" Execution Times



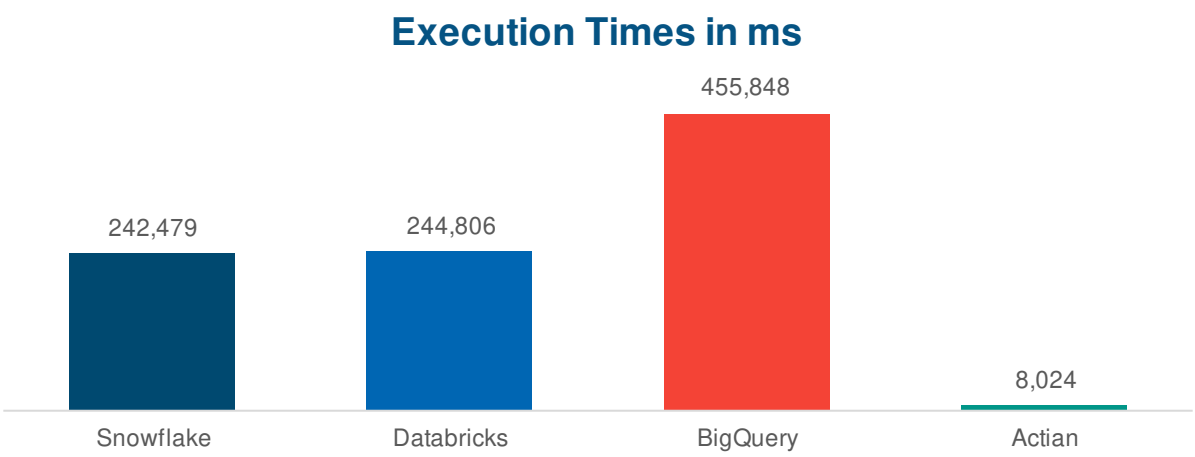
Actian was twice as fast as second-place Snowflake for Query 16.

Figure 18. Query 17: "Small Quantity Order Revenue" Execution Times



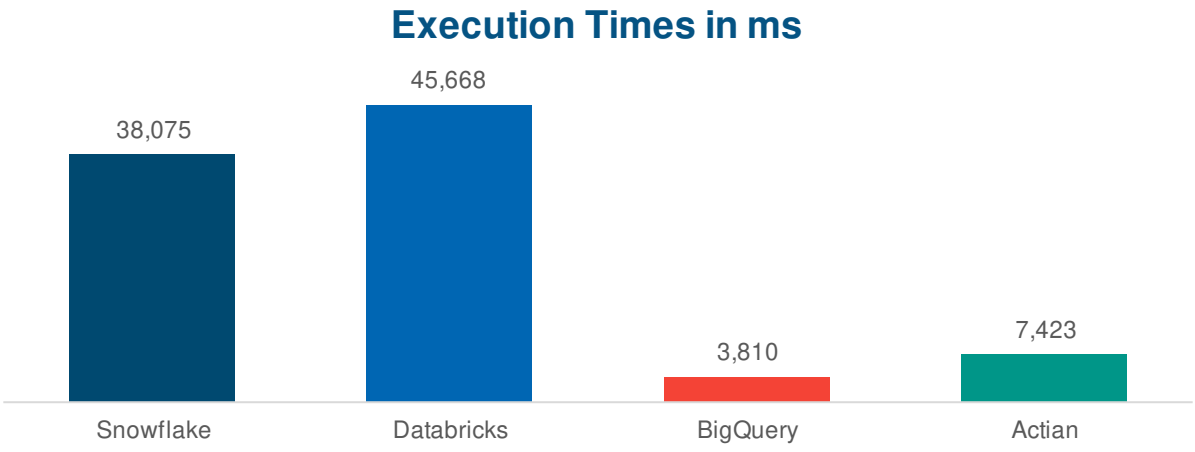
Query 17 yielded major performance advantages for Actian.

Figure 19. Query 18: "Large Volume Customer" Execution Times



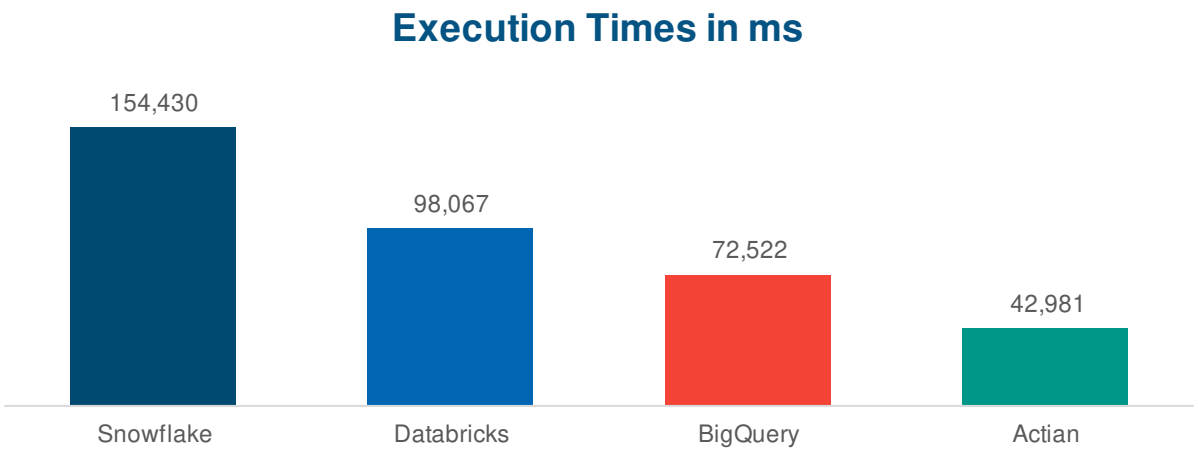
Actian shines here for Query 18, with Snowflake second and Databricks a close third.

Figure 20. Query 19: "Discounted Revenue" Execution Times



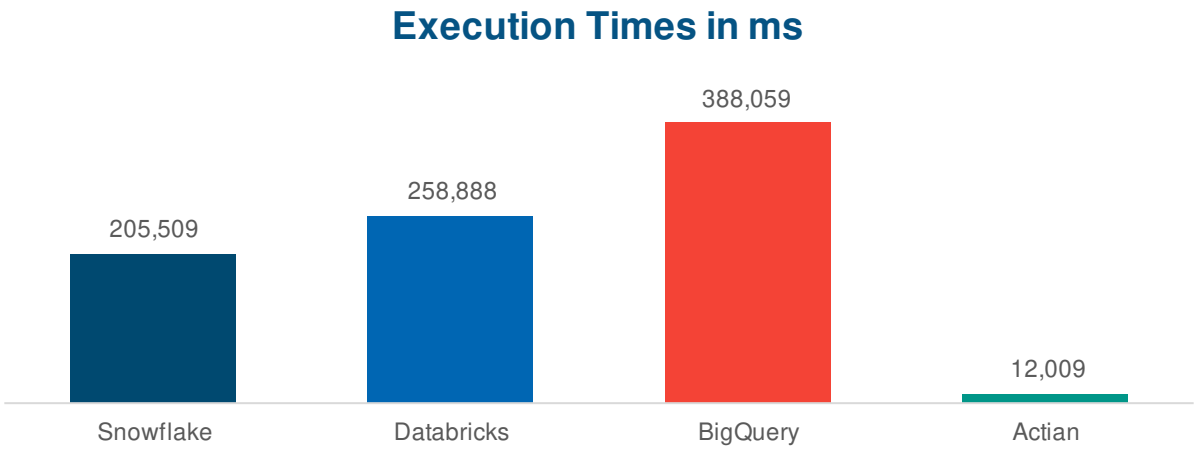
Query 19, a SUM, was a strong performer for BigQuery.

Figure 21. Query 20: "Potential Part Promotion" Execution Times



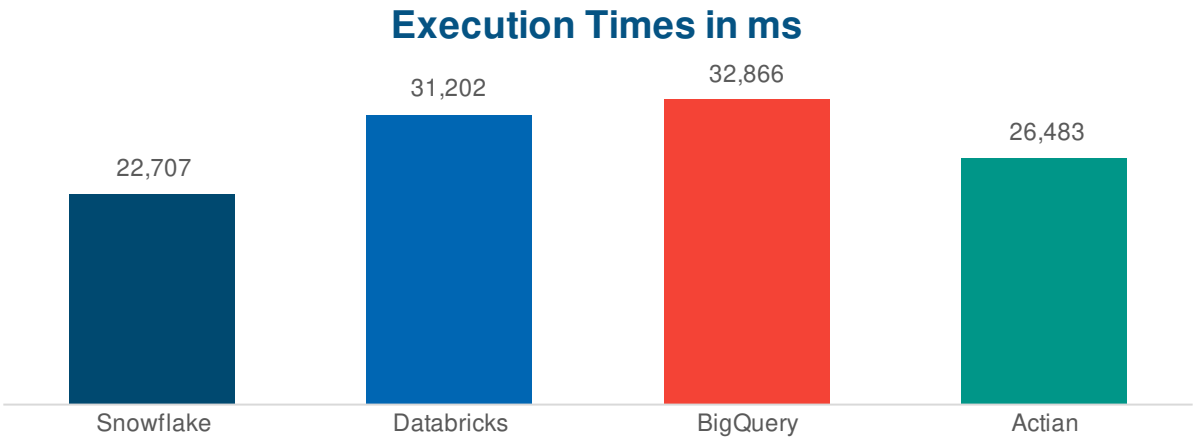
Query 20 showed Actian to be faster in analogous configurations to Snowflake and BigQuery.

Figure 22. Query 21: "Suppliers Who Kept Orders Waiting" Execution Times



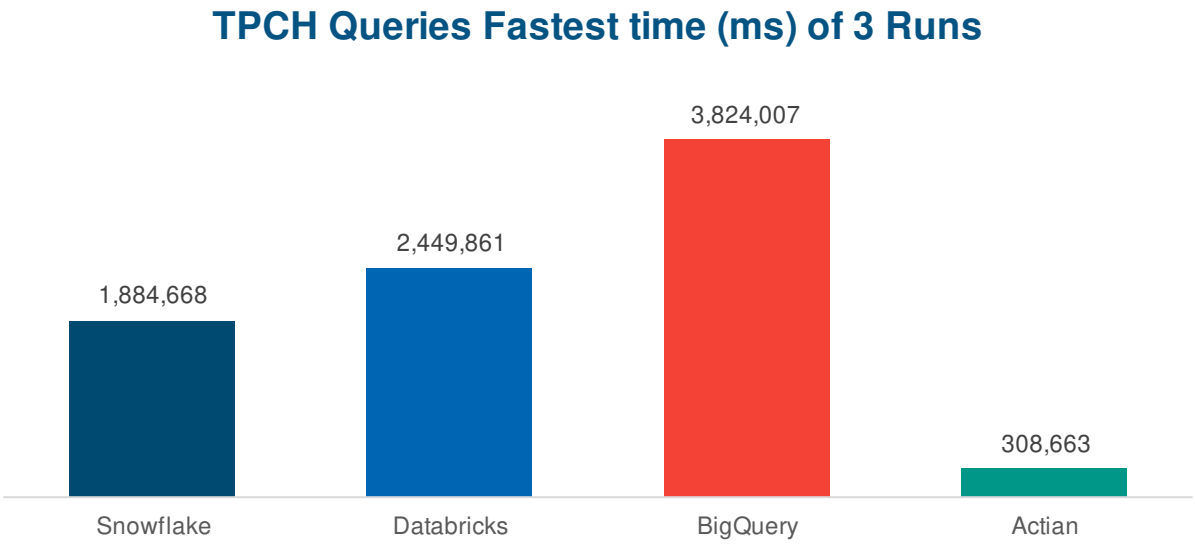
Query 21 shows the familiar pattern of high-performance differences between Actian and the analogous configurations.

Figure 23. Query 22: "Global Sales Opportunity" Execution Times



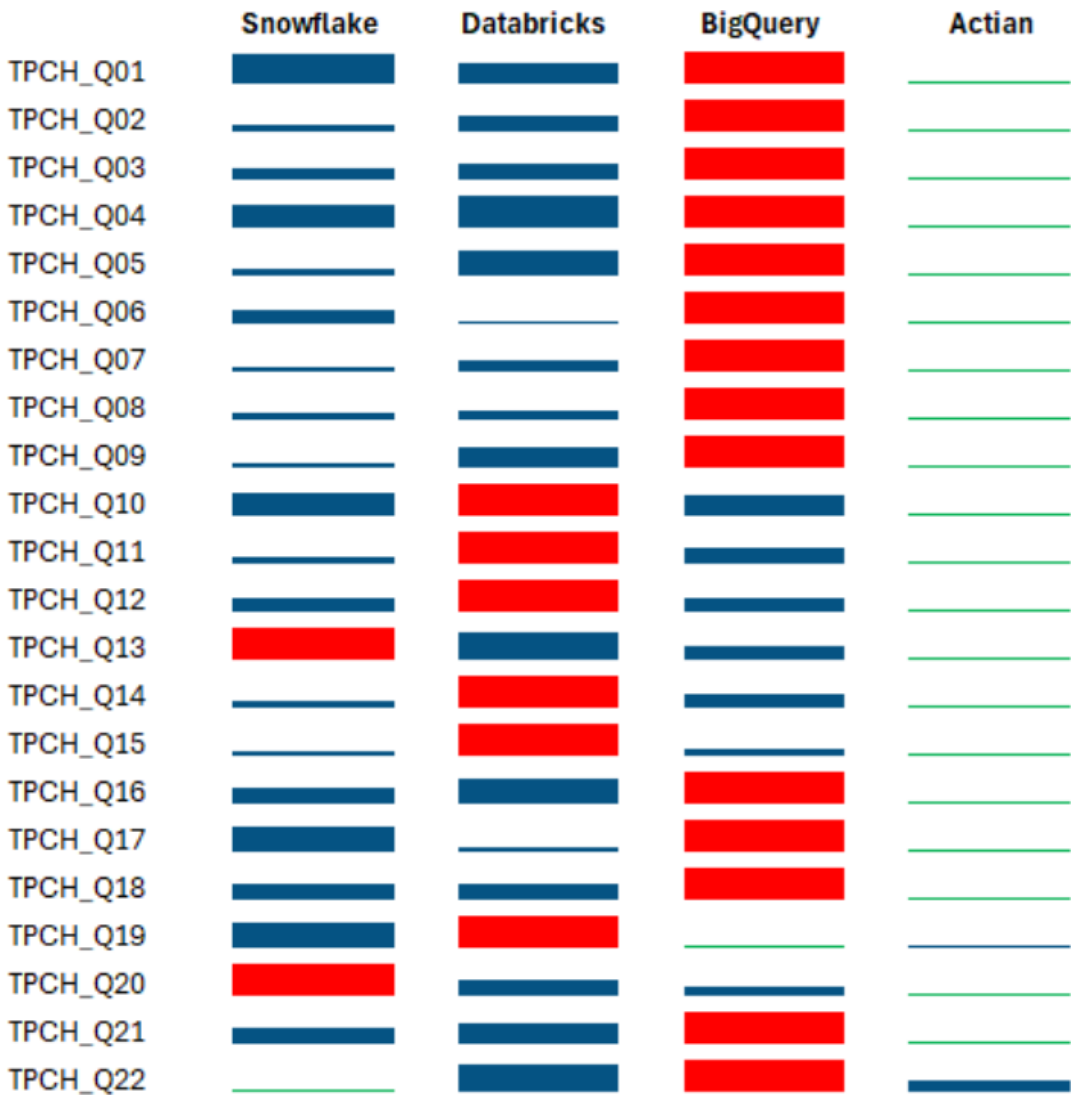
Snowflake outperformed the field for Query 22, although it was close with second place Actian.

Figure 24. TPC-H total time using fastest runs of each query



Actian's performance of 308,663 ms for a single user TPC-H test is significantly faster than Snowflake's 1.9M ms, Databricks 2.4M ms and BigQuery's 3.8M ms. Actian is a great choice for applications that require fast query processing performance.

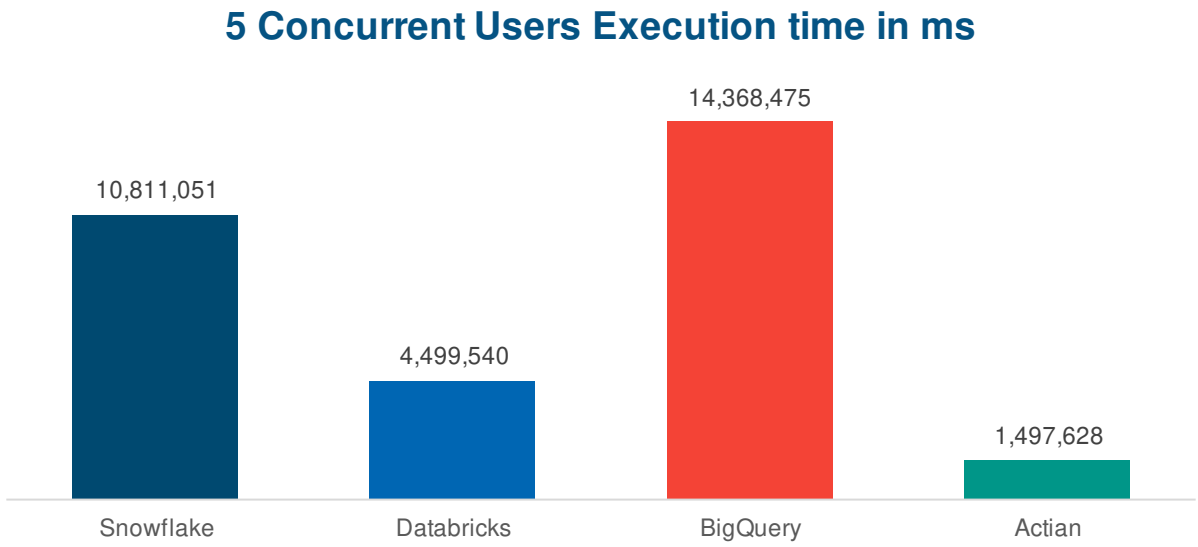
Figure 25. TPC-H query times using fastest runs of each query



Actian was the top performer on 20 of the 22 TPC-H queries. BigQuery outperformed on Query 19 and Snowflake on Query 22. BigQuery was the slowest on 14 of the queries followed by Databricks, which was the slowest on 6 of the queries and Snowflake, which was only slowest on 2 of the queries.

Tests with 5 concurrent users

Figure 26. TPC-H All Queries (1 Stream), 5 Users Execution Times



The elapsed time of the test is actually the duration of the slowest running thread of the concurrency test. To complete the 5-user test, we have to wait until all 5 users complete all their queries. Thus, the slowest thread represents the elapsed time of the test from beginning to end.

Actian's performance of 1.5m ms for a 5-concurrent TPC-H test is significantly faster than Databrick's 4.5m ms, Snowflake's 10.8m ms, and BigQuery's 14.4m ms, making it 3.2 times faster than Databricks, 7 times faster than Snowflake and 9.6 times faster than BigQuery once you add in concurrency. We noted Snowflake swapped second place with Databricks once concurrency was added.

Actian is a great choice for applications that require fast query processing performance. Let's add price to the equation now.

Price-Performance

System cost can be a difficult aspect to compare systems, because vendor platforms vary on their pricing and licensing models. However, all four platforms have clear and consistent on-demand hourly cloud pricing that we can use to determine price per performance.

Actian has a clear pricing model. For the Actian software usage and the underlying platform, there is a $2.50 per Actian Unit (AU) per hour cost. Thus, with 128 AUs, we paid $320 per hour.

Google BigQuery charges either monthly or hourly for Slot commitments. To ensure a like-for-like comparison, we used the Enterprise hourly rate of $0.06 per slot per hour. With a 4XL reservation (4,800 slots), we paid $288 per hour in the US region. We rounded BigQuery performance to next whole hour for slot-hour billing.

Snowflake has several pricing tiers with varying levels of features and security capabilities. We used the Enterprise tier on GCP for a cost of $3.00 per cluster node. We used a 4X-Large cluster with 128 nodes. Thus, we paid $384.00 per hour.

For Databricks, we used 144 units of the 2X-Large, which came to $100.80 per hour.

With the hourly cost of the configuration, to calculate the price-per-performance, we used the following formula:

$$\frac{\text{Elapsed time of test (seconds)} \times \text{Cost of platform (\$/hour)}}{\text{(seconds/hour)}} \; 3{,}600$$

The following tables detail the price-performance for the different tests.

Table 4. Price Performance

30TB	Actian	Snowflake	BigQuery	Databricks
Tier	Enterprise	Enterprise	Enterprise	Serverless
Size	128AU	4X-Large	4XL 4,800 Slots	2X-Large
Units	128	128	4800	144
Cost $/unit/hour	$2.50	$3.00	$0.06	$0.70
$/hour	$320.00	$384.00	$288.00	$100.80
Single User Total Time (seconds)	309	1,885	7,200	2,450
Price Perf Single User	**$27.47**	**$201.07**	**$576.00**	**$68.60**
Times Cheaper		***7.32x***	***20.97x***	***2.50x***
5 Concurrent Users Total Slowest Time (seconds)	1,498	10,811	14,400	4,500
Price Perf 5 Concurrent Users	**$133.16**	**$1,153.17**	**$1,152.00**	**$126.00**
Times Cheaper		***8.66x***	***8.65x***	***0.95x***

BigQuery total times are rounded to next whole hour for slot-hour billing.

Actian's price-performance for a single user was 2.5 times better than the next competitor Databricks and over 7 times better than Snowflake. BigQuery's price-performance was 20 times that of Actian.

For the concurrent user test, Databricks and Actian showed comparable results, while Snowflake and BigQuery were over 8 times the price-performance of Databricks and Actian.

Actian is a great choice for applications that require fast query processing performance, with concurrency.

Conclusion

Cloud databases are a way for enterprises to avoid large capital expenditures, provision quickly, and provide performance at scale for advanced analytic queries. Relational databases with analytic capabilities continue to support the advanced analytic workloads of the organization with performance, scale, and concurrency. In a representative set of corporate-complex queries from the well-known TPC-H standard, Actian consistently outperformed the competition.

Overall, the benchmark results were insightful in revealing query execution performance and some of the differentiators for Actian, Snowflake, Google BigQuery and Databricks. Using TPC-H-like tests, Actian outperforms Databricks by nearly eight times, Snowflake by more than six times, and BigQuery by more than twelve times, all with similar specs. Actian outperforms Snowflake by over 7 times, BigQuery by 9.6 times, and Databricks by 3 times when 5-user concurrency is taken into account.

Price and performance are critical points of interest when it comes to selecting an analytics platform, because they ultimately impact total cost of ownership, value, and user satisfaction. Our analysis reveals Actian to be very powerful and comparative in value.

About Actian

Actian, the hybrid data management, analytics and integration company, delivers data as a competitive advantage to thousands of customers worldwide. Through the deployment of innovative hybrid data technologies and solutions Actian ensures that business critical systems can transact and integrate at their very best – on premise, in the cloud or both. Thousands of forward-thinking organizations around the globe trust Actian to help them solve the toughest data challenges to transform how they run their businesses, today and in the future. For more, visit **http://www.actian.com**.

About McKnight Consulting Group

Information Management is all about enabling an organization to have data in the best place to succeed to meet company goals. Mature data practices can integrate an entire organization across all core functions. Proper integration of that data facilitates the flow of information throughout the organization which allows for better decisions – made faster and with fewer errors. In short, well- done data can yield a better run company flush with real-time information... and with less costs.

However, before those benefits can be realized, a company must go through the business transformation of an implementation and systems integration. For many that have been involved in those types of projects in the past – data warehousing, master data, big data, analytics - the path toward a successful implementation and integration can seem never-ending at times and almost unachievable. Not so with McKnight Consulting Group (MCG) as your integration partner, because MCG has successfully implemented data solutions for our clients for over a decade. We understand the critical importance of setting clear, realistic expectations up front and ensuring that time-to-value is achieved quickly.

MCG has helped over 100 clients with analytics, big data, master data management and “all data” strategies and implementations across a variety of industries and worldwide locations. MCG offers flexible implementation methodologies that will fit the deployment model of your choice. The best methodologies, the best talent in the industry and a leadership team committed to client success makes MCG the right choice to help lead your project.

MCG, led by industry leader William McKnight, has deep data experience in a variety of industries that will enable your business to incorporate best practices while implementing leading technology. See www.mcknightcg.com.

MCKNIGHT CONSULTING GROUP

Our services span strategy through implementation for turning information into the asset it needs to be for your organization. We offer a wide range of services related to the important asset of enterprise information.

For Enterprises

- ✓ 2-Week Data Initiative Jumpstart
- ✓ Proof-of-Concepts and Minimum Viable Products
- ✓ Information Management Action Plan
- ✓ Data Platform Selection
- ✓ Data Maturity Assessment and Improvement
- ✓ Data Warehousing and Analytics
- ✓ Big Data
- ✓ Master Data Management, Data Governance and Data Quality
- ✓ Enterprise Business Intelligence Architecture
- ✓ Predictive Analytics Accelerator
- ✓ Compliance Accelerator
- ✓ Artificial Intelligence Accelerator
- ✓ Education and Workshops

For Vendors

- ✓ Benchmark Services
- ✓ Competitive Market Research
- ✓ Engineering Projects
- ✓ Research Development

For Your Audience

- ✓ Keynotes, Webinars, Seminars and Podcasts