

# InterSystems IRIS: High Performance Data Management Software for Concurrent Data Ingestion and Real-time Queries

**Date:** June 2020 **Author:** Kerry Dolan, Senior IT Validation Analyst

## Abstract

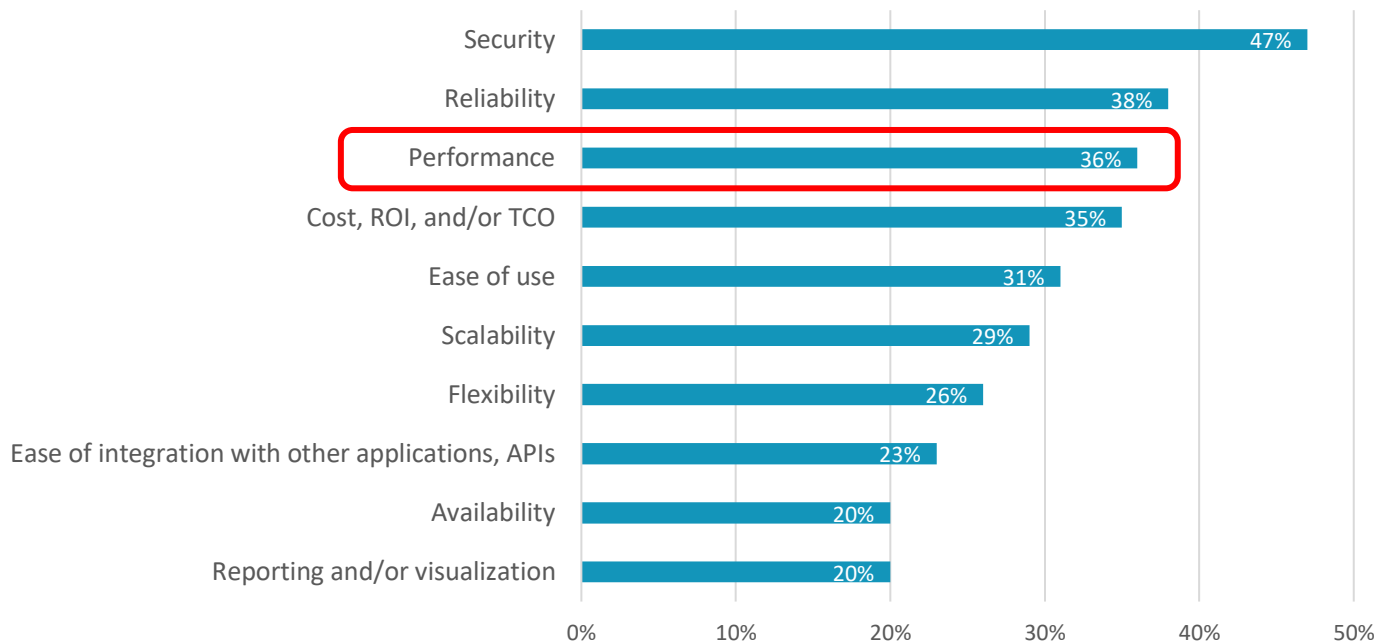
This report documents ESG’s validation of concurrent data ingestion and real-time query performance testing of various database management software products that demonstrates the ability of InterSystems IRIS data platform to ingest hundreds of millions of records and simultaneously execute millions of queries with microsecond performance, outperforming other traditional and in-memory products.

## The Challenges

For many organizations, the ability to collect data and analyze it in real time is an essential task that drives revenue, improves visibility, informs strategy, and aids decision making. For example, applications focused on financial trading, IoT, fraud detection, and real-time personalization must ingest large amounts of data and analyze it immediately. The challenge is finding a database platform with sufficient horsepower to handle large-scale ingest and querying simultaneously without impeding performance. When ESG asked database and analytics professionals about technologies supporting data analytics, performance was among the topmost important capabilities.<sup>1</sup>

**Figure 1. Top 10 Most Important Capabilities for Data Analytics Technologies**

When looking to adopt new technologies to support data analytics initiatives within your organization, what are the most important capabilities/attributes?(Percent of respondents, N=310, five responses accepted)



Source: Enterprise Strategy Group

<sup>1</sup> Source: ESG Master Survey Results, [The State of Data Analytics](#), August 2019.

In-memory databases offer high performance but are expensive to scale and have hard memory limits that can affect reliability and cause restart delays. Traditional databases offer persistence and reliability but lack the high performance of in-memory databases. InterSystems IRIS can process both ingestion and query workloads simultaneously, with performance equal to or better than in-memory-only databases, without their limitations. InterSystems has published an open-source test to demonstrate this claim, which ESG is validating in this report.

## The Solution: InterSystems IRIS

InterSystems IRIS is a data management software platform that was built for high performance, multi-workload processing at scale. As a multi-model DBMS, it provides native support for relational, object-oriented, document, key-value, and hierarchical data objects; in addition, it enables consistent high performance for both transactional and analytic workloads simultaneously. While a complete product description is beyond the scope of this report, some key functionality is described below.

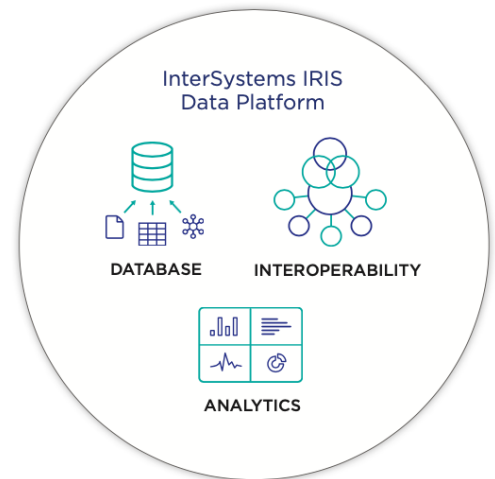
- An important feature that provides superior ingest performance is the multidimensional data engine in InterSystems IRIS that enables efficient, compact storage with a rich data structure, speeding data ingest, access, and updates while minimizing resource usage and disk consumption.
- Real-time analytics performance is achieved by using a transactional-bitmap indexing schema that enables InterSystems IRIS to process complex queries quickly, including on real-time data, without searching the entire database.
- InterSystems IRIS Enterprise Cache Protocol, an intelligent, distributed memory caching mechanism, enables it to execute sophisticated queries on very large data sets with high performance and reliability, including performing joins accessing distributed data, without making multiple data copies.

Other features include:

- In-memory performance with built-in data persistence in a format optimized for rapid data access.
- Built-in distributed caching layer with automatic and guaranteed consistency.
- Full SQL support.
- Deployment on-premises, in all major public clouds, and in hybrid environments, with a single API.

## ESG Tested

ESG validated performance benefits of InterSystems IRIS using the company's publicly available, customizable, open-source Speed Test benchmark kit.<sup>2</sup> The benchmark was designed to measure concurrent real-time ingest and query performance. This is a common use case that financial services, fraud detection, IoT, and other applications face. For example, while financial services firms are executing thousands of trades, thousands of users are querying for order status, risk management, etc. Similarly, IoT sensor data comes in fast from the field and applications must perform immediate anomaly detection and other real-time calculations. When a database is stressed in this way, having to simultaneously ingest data and execute analytic queries can slow performance.



<sup>2</sup> <https://github.com/intersystems-community/irisdemo-demo-htap>.

## Speed Test

The Speed Test benchmark was designed to simulate the combined ingestion and query pressure that applications deal with. For a typical in-memory database, ingestion fills the memory and causes the database to compress and purge data from memory, with slow disk writes. Querying with high performance requires that data be kept in memory. Executing both tasks simultaneously impedes performance.

Speed Test queries the database for a set of predefined records as quickly as possible; this tests the ability of the database to keep frequently accessed records in memory and measure query response time under ingestion pressure. The ingestion workers open multiple JDBC connections to the database and use them to insert as many records into the table as possible.

To make the test comparable with a variety of databases and database types, there are no joins or special indices used. A single, 19-column brokerage account table is used, with accountID as the primary key (the only indexed column). Data includes strings, dates/times, and big integers.

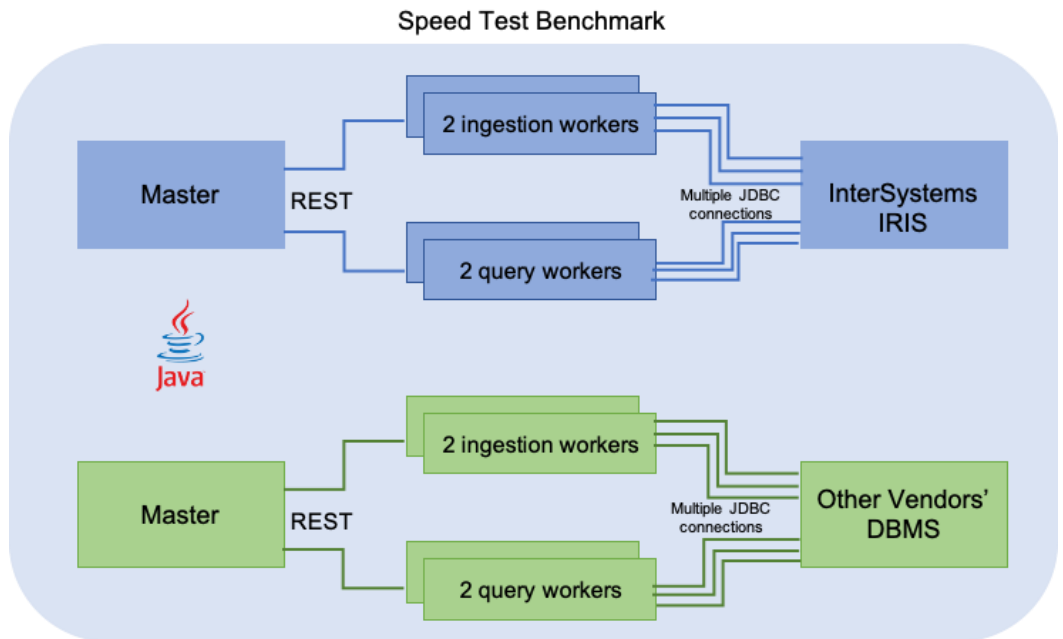
The test uses two ingestion workers and two query workers. When ingestion begins, each worker generates 1,000 random values for each column in the table. When the ingestion needs to create a new record to add to a batch, it will 1) get the next value for the primary key column, and 2) combine values from the pool of random values previously generated to create a new record. When a batch is complete, it is sent to the database. This maximizes ingestion worker efficiency and speed.

The query workers will also open multiple JDBC connections to the database and use them to fetch a set of records using eight fixed keys by accountID as fast as possible. Record retrieval was intentionally kept simple to create a level playing field for comparing different database technologies and enabled the test to focus specifically on a system's cache management as ingestion and queries run concurrently.<sup>3</sup>

The Master node collects statistics, including:

- For ingestion: number of records ingested; records ingested per second; MB ingested; and MB ingested per second.
- For query: number of records queried; records queried per second; MB queried; MB queried per second; query response time.

The query workers retrieve the records and sum the length of each column in the retrieved rows into a variable as proof of work. Retrieving the rows in this way forces the databases to send the data over the TCP/IP connection, preventing JDBC drivers from optimizing only for identification of data.



<sup>3</sup> The GitHub link (<https://github.com/intersystems-community/irisdemo-demo-htap>) includes instructions for modifying the test to add complexity.

## AWS Setup

In this testing, comparisons were made with InterSystems IRIS and three other data platforms, all running in the AWS Cloud, and all running for 1,200 seconds (20 minutes). Vendor A is a leading in-memory database, while Vendors B and C are leading disk-based traditional databases; for testing against Vendors B and C, databases were pre-expanded to optimize performance when possible.

In each case, AWS hardware configurations were matched between InterSystems IRIS and the other solution to ensure an “apples to apples” comparison (see Table 1). In the case of Vendor B, because AWS required a slightly different instance, it was configured with twice the RAM as InterSystems IRIS to ensure that InterSystems IRIS had no advantage. In addition, since that instance included replication, InterSystems IRIS was configured with mirroring so that both databases would be tasked with writing a redundant copy.

**Table 1. AWS Configurations**

	AWS Instance	Cores	RAM	EBS IOPS	Replication/Mirroring
InterSystems IRIS	i3.xlarge	4	32GB	3,500	No
Vendor A	i3.xlarge	4	32GB	3,500	No
InterSystems IRIS	m5.xlarge	4	16GB	2,500	Yes
Vendor B	db.r5.xlarge	4	32GB	2,500	Yes
InterSystems IRIS	m5.xlarge	4	16GB	3,500	No
Vendor C	m5.xlarge	4	16GB	3,500	No

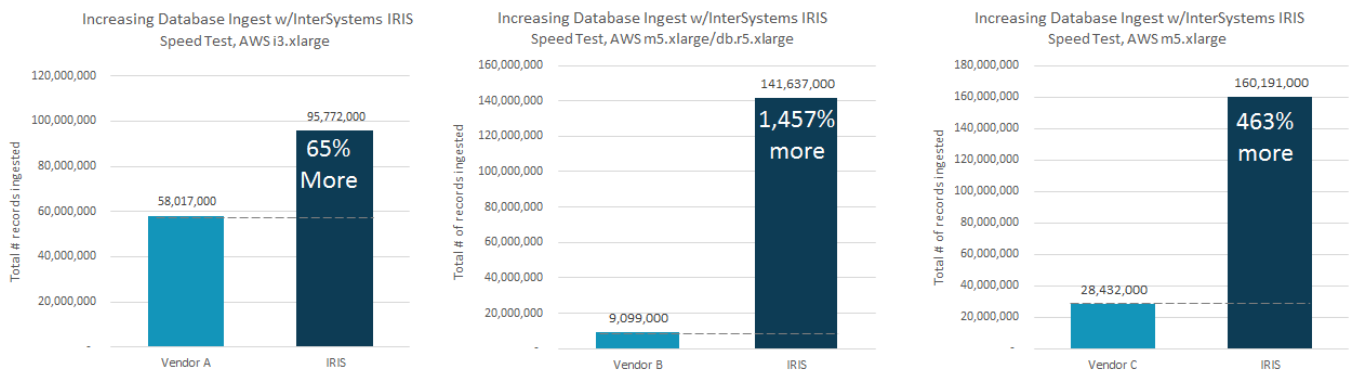
Source: Enterprise Strategy Group

## Results

ESG reviewed all the data collected by the Speed Test kit, which measured performance while both ingestion and query workloads were running concurrently. We then compared InterSystems IRIS with Vendors A, B, and C in four key categories: total number of records ingested; average number of records ingested per second; total number of records queried; and query response time.

Figure 2 shows the total number of records ingested during the tests. This demonstrates the ability to ingest large amounts of data while performing concurrent real-time queries within a fixed time window. In each case, InterSystems IRIS ingested hundreds of millions of records, resulting in 65% more than Vendor A; 1,457% more than Vendor B; and 463% more than Vendor C.

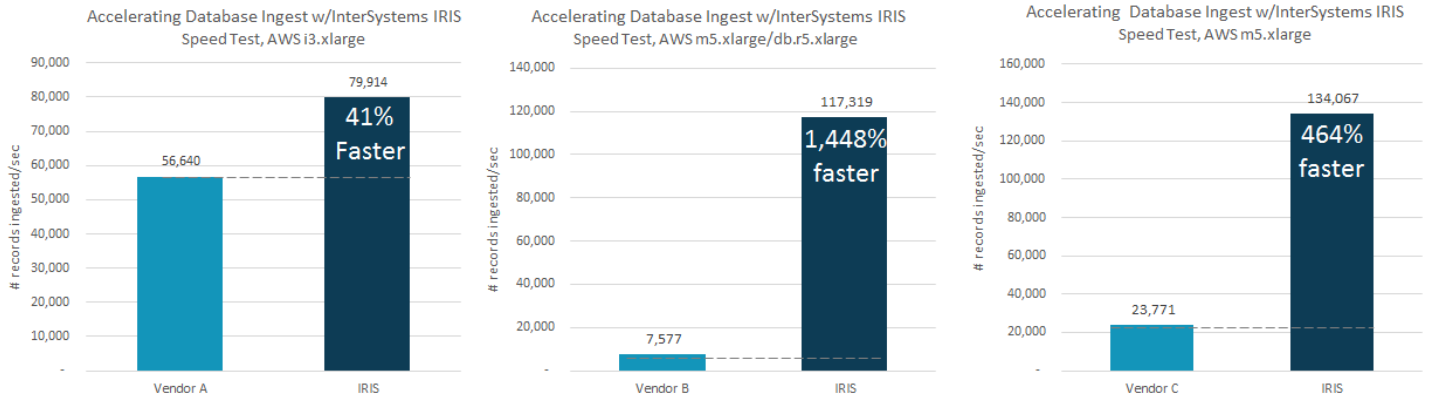
**Figure 2. Increasing Database Ingest with InterSystems IRIS**



Source: Enterprise Strategy Group

Next, we looked at the average number of records ingested per second. This shows the rate at which each data platform can ingest data while simultaneously querying. As Figure 3 shows, InterSystems IRIS was ingesting data an average of 41% faster than Vendor A throughout the test, 1,448% faster than Vendor B, and 464% faster than Vendor C.

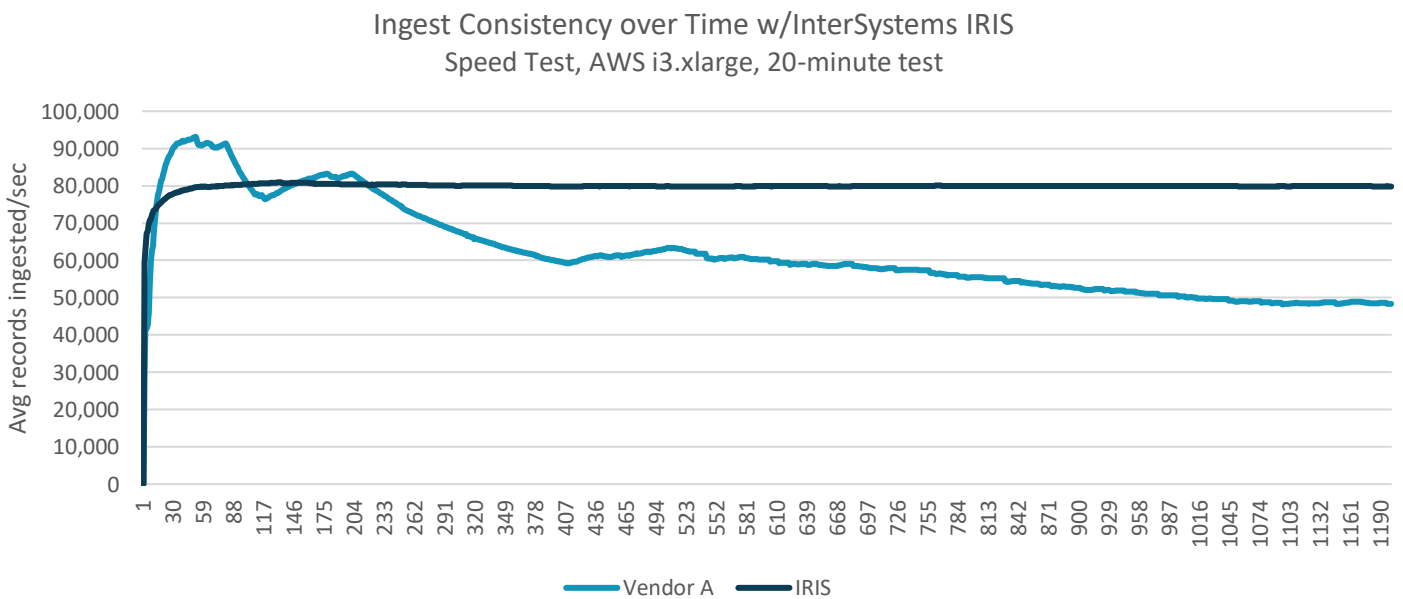
**Figure 3. Accelerating Database Ingest with InterSystems IRIS**



Source: Enterprise Strategy Group

In addition, in all three cases, the minimum and maximum rates of records ingested per second by InterSystems IRIS were higher than the others. This means organizations can count on consistently fast data ingestion rates. ESG reviewed the variability of ingestion rate by reviewing the standard deviation and graphs of ingestion over time; IRIS showed less variability in all cases. For example, the graph in Figure 4 compares Vendor A with InterSystems IRIS in terms of ingest rate. While both started with high ingest rates, the in-memory database (Vendor A) declined 48% from its maximum. While the in-memory database ingested data quickly at first, performance degraded as the memory filled. Compression becomes more difficult with more data, and more data in memory forces an in-memory database to write to disk. In contrast, InterSystems IRIS quickly reached a high ingestion rate and remained steady throughout the test.

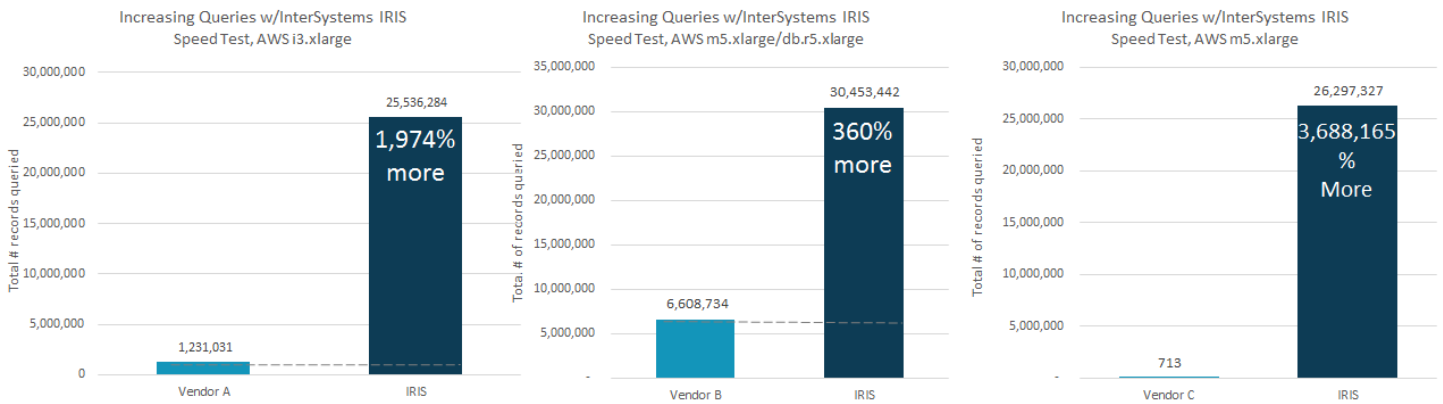
**Figure 4. Ingest Consistency over Time with InterSystems IRIS**



Source: Enterprise Strategy Group

Next, we looked at the total number of records queried while concurrently maximizing data ingestion. The more queries a database can execute, the greater the potential for insights, better decision-making, and real-time actions. InterSystems IRIS queried 1,974% more records than Vendor A, 360% more than Vendor B, and 3,688,165% more than Vendor C.<sup>4</sup>

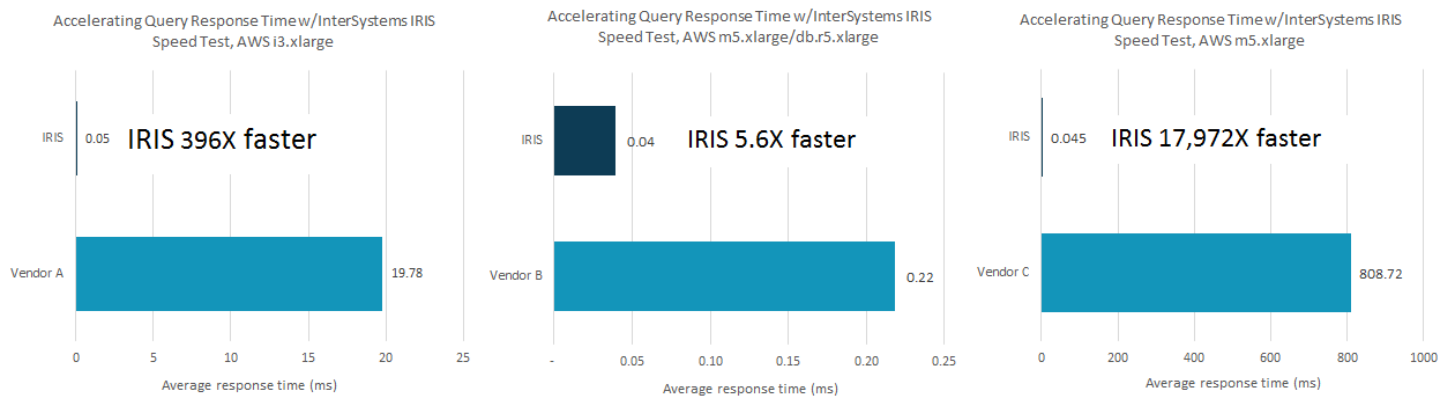
**Figure 5. Increasing Queries with InterSystems IRIS**



Source: Enterprise Strategy Group

Finally, ESG reviewed the average query response times for each data platform during concurrent ingest. The average query response times for InterSystems IRIS were many times faster than the other data platforms. While simultaneously ingesting data, InterSystems IRIS consistently delivered microsecond query response times (see Figure 6 and Table 2). InterSystems IRIS queried the data 396X faster than Vendor A, 5.6X faster than Vendor B, and 17,972X faster than Vendor C.

**Figure 6. Accelerating Query Response time with InterSystems IRIS**



Source: Enterprise Strategy Group

In addition, in all three cases, InterSystems IRIS demonstrated significantly less variability in response times than the others; this means that InterSystems IRIS consistently delivered high performance querying while ingesting data at a high rate. Table 2 shows query response time minimums, maximums, and standard deviations for each test.

<sup>4</sup> Query results (both number of records queried and response time) for Vendor C were significantly lower than others. Troubleshooting the environment to improve results was unsuccessful with the configured RAM.

**Table 2. Query Response Time Statistics**

	Minimum Query Response Time (milliseconds)	Maximum Query Response Time (milliseconds)	Standard Deviation (milliseconds)
InterSystems IRIS	0.04	1.78	0.05
Vendor A	0.2	3,000	122.64
InterSystems IRIS	0.033	0.088	0.005
Vendor B	0.173	45.454	1.306
InterSystems IRIS	0.04	0.35	0.009
Vendor C	20.41	1,000	273.52

Source: Enterprise Strategy Group

## Why This Matters

Financial trading, IoT, fraud detection, gaming, personalization, and other applications stress databases with simultaneous data ingestion and querying. When these stresses reduce performance, they can negatively impact revenue streams, customer loyalty, and infrastructure costs. Organizations need data platforms that can provide high performance for processing both ingestion and analytics workloads simultaneously, even when workloads spike, to eliminate these problems and enable better and faster insights, decisions, and actions using more data.

ESG validated that InterSystems IRIS significantly outperformed both traditional and in-memory data platforms, ingesting more data, faster, while simultaneously querying more data with faster query rates and microsecond query response times. InterSystems IRIS performance was also much more consistent than the other solutions.

## The Bigger Truth

Many of today's applications require high throughput and concurrent high-performance data access, such as for financial services, trading, fraud detection, real time personalization, IoT, and other scenarios.

Unfortunately, traditional disk-based databases are often too slow for the high throughput and data access rates required. In-memory databases can offer high performance but are expensive to scale and have hard memory limits that can affect reliability and cause restart delays.

InterSystems IRIS is multi-model data management software that was designed for high-performance, resource-efficient, multi-workload processing at scale. Its data engine and real-time bitmap indexing help IRIS deliver in-memory performance for ingestion and querying with built-in data persistence.

InterSystems created the Speed Test benchmark to be a simple test of data platform horsepower and to evaluate a data platform's performance capabilities for concurrent high-volume data ingest and querying. ESG validated the results of performance testing that compared IRIS with three other data platforms, all running in the AWS Cloud. Results of 20-minute testing included:

- InterSystems IRIS ingested more records faster and queried more records faster than both traditional and in-memory databases.
- InterSystems IRIS ingested hundreds of millions of records in each test, 65% to 1,457% more than other platforms.

- InterSystems IRIS queried significantly more records than other platforms and maintained microsecond query response times that were significantly faster than other platforms.
- InterSystems IRIS ingest and query results were much more consistent over time compared with other platforms.

These results demonstrate that InterSystems IRIS can provide an excellent foundation for these types of applications. Having a platform with this caliber of performance can assure organizations of maximum uptime and performance during events that drive data ingestion and query spikes. In addition, InterSystems IRIS can be more cost-effective than other data platforms that depend on large memory volumes or that must create separate data copies, increasing disk costs.

The results documented in this test were based on testing in a controlled environment; however, InterSystems has made the benchmark kit available [as open-source software](#) for anyone to run either in AWS or using Docker on local machines. Still, in production, your mileage may vary, as the variables in any production data center will impact performance. It is important to plan and test in your own environment to validate the efficacy of any solution.

ESG believes that InterSystems IRIS is a scalable, high performance data platform that can easily handle the high-volume simultaneous data ingestion and querying that modern applications demand. If you are in search of a robust data platform for your organization, InterSystems IRIS deserves a good look.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.

The goal of ESG Validation reports is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Validation reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.