ESG Technical Review

InterSystems IRIS Data Platform: A Unified, Efficient Data Platform for Analytics and Business Insight

Date: October 2019 **Author:** Kerry Dolan, Senior IT Validation Analyst

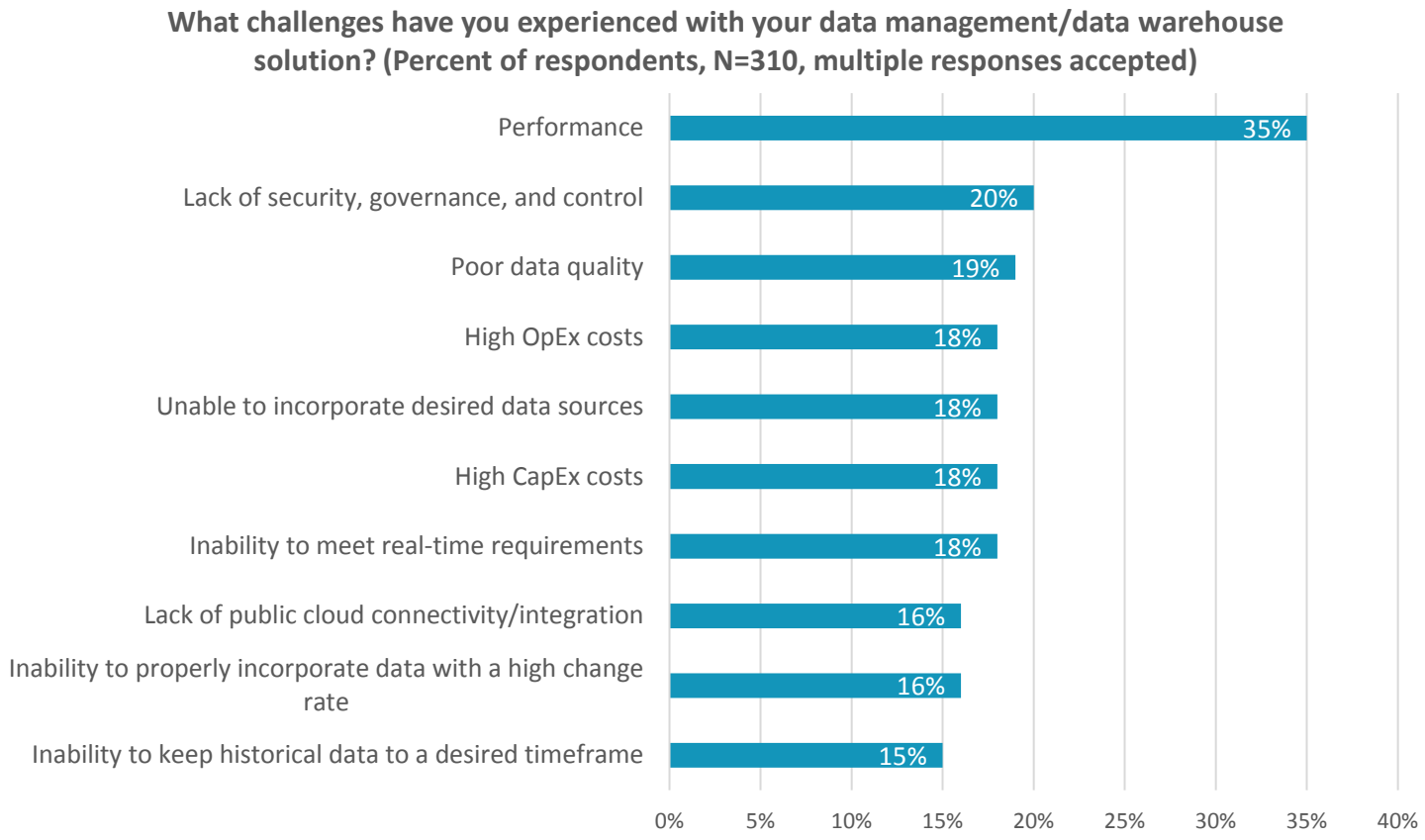
Abstract

This ESG Technical Review documents our audit of InterSystems IRIS Data Platform performance testing. Tests were executed in a proof of concept environment, with comparisons of three common databases. Testing focused on performance of common queries run in the financial services industry.

The Challenges

Data drives both transactional and analytical business processes. Transactions with customers and partners define revenue and profit, while analyzing data—whether from customers and suppliers or from sensors and machines—provides key insights to inform business decisions. The ability to transform real-time data into knowledge to make business decisions can be the difference between success and failure. Data management technologies must provide a strong foundation for analytics, and must be easily deployed, cost efficient, scalable, and most of all, high performing; delays bog down answers and business transformation. When ESG asked IT decisions makers about their data management challenges, performance was most often cited, followed by security and control issues, data quality, cost, and meeting real-time requirements.[1]

Figure 1. Top Ten Data Management Challenges



Source: Enterprise Strategy Group

[1] Source: ESG Master Survey Results, *The State of Data Analytics*, August 2019.

Most organizations leverage multiple databases, adding both infrastructure and staffing costs. If organizations can reduce the footprint required to handle all their transactional and analytics workloads (including real-time, historical, natural language processing, and business intelligence), they can save money. But they must deliver high performance, along with data durability, scalability, and management ease.

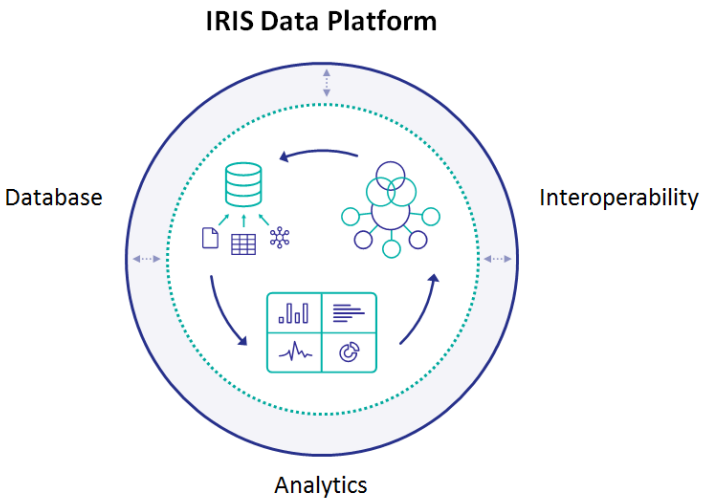
The Solution: InterSystems IRIS Data Platform

The InterSystems IRIS Data Platform was designed to help organizations build intelligent, data-intensive applications that support multiple transaction processing and analytical use cases. InterSystems IRIS is a unified, real-time platform that combines powerful data management, analytics, and interoperability into a single solution. It was designed to handle and optimize mixed workloads and data-intensive applications at scale in the cloud, on-premises, or in hybrid environments, while providing built-in security, proven reliability, and support.

InterSystems IRIS handles the varied needs of businesses today, all in a single platform, with a single copy of data. This saves organizations from having to purchase and manage different point solutions, and from storing and mapping multiple data copies. Features include:

- *Multi-workload*. InterSystems IRIS processes both transactional and analytic workloads in the same engine at scale, eliminating the complexity associated with moving and transforming transactional data for analysis, and with multi-database deployments. InterSystems IRIS can process and index transactional data, while simultaneously running real-time analytics on that transactional data, along with non-real-time data, delivering a faster response to the business and therefore better insight.
- *Multi-model*. InterSystems IRIS provides a single database that can represent multiple projections of the same data, rather than a traditional siloed database approach. It supports numerous data models, including relational, key-value, document, object, and a variety of schema-free NoSQL data.
- *Interoperability*. The comprehensive integration and open analytics platform enables streamlined data coordination and service orchestration, making it simpler to connect data sources and business processes. It also includes full life cycle API management functionality for discovering, publishing, orchestrating, and monetizing services.
- *Analytics*. InterSystems IRIS incorporates real-time and batch analytics, including distributed SQL processing, business intelligence, predictive modeling runtime engine, and natural language processing of unstructured data from sources such as emails and social media, and integrates with big data processing frameworks such as Apache Spark.
- *Enterprise-level security*. Integration with Kerberos, LDAP, and KMIP, role-based access control, and encryption of data in transit and at rest provide rock solid security.

Figure 2. InterSystems IRIS Data Platform



Source: InterSystems Corporation

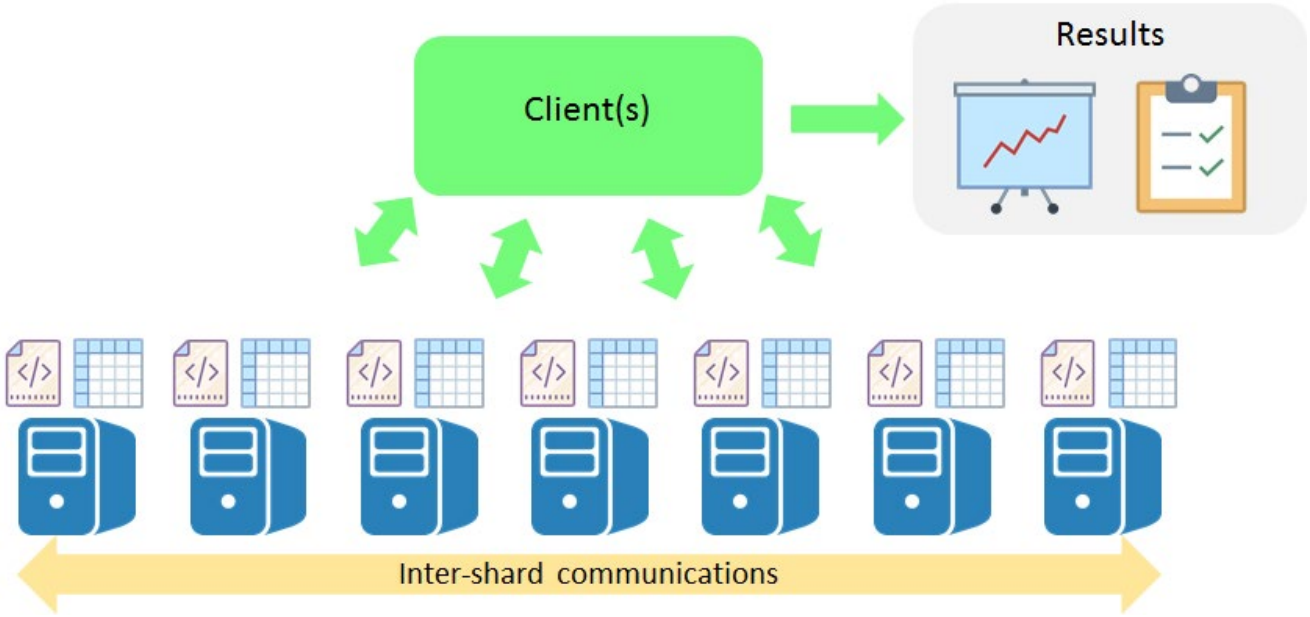
InterSystems IRIS Architecture

InterSystems IRIS can ingest and store transactional data fast, while at the same time processing analytic workloads on batch and real-time, ACID-compliant data; this eliminates the delays from moving real-time data to an analytics platform for processing. It scales vertically with parallel SQL processing, and employs a differentiated approach to horizontal scaling.

For large data sets and complex queries, InterSystems IRIS employs a specialized approach to horizontal scaling. Using commodity servers to distribute query processing, it leverages intelligent sharding and distributed caching. Workloads and tables can be distributed to multiple nodes by sharding the database; sharding is transparent to the application, and requires few if any code changes. InterSystems IRIS uses an intelligent Shard Master that pushes out application code for parallel execution, aggregates the results of each shard, and returns a final result (see Figure 3). Each shard is aware of the location of all of the other data; if data needed for a query resides in another shard, one shard can reach out directly to another to obtain *only* the data it needs. With other solutions, the entire table would have to be broadcast to the other nodes, increasing latency and timeouts. This enables InterSystems IRIS to achieve high performance even for queries that involve complex joins of sharded data.

Each server's cache is independent, as the sharding creates disjoint partitions. Scaling out servers will therefore increase the cluster's memory in a linear fashion. As a result, when sized appropriately, InterSystems IRIS deployments can achieve levels of performance similar to in-memory databases, but without the requirement to fit all data into memory because it uses the cache for frequently used data, with "colder" data durably persisted on disk in a format optimized for rapid random access.

Figure 3. Horizontal Scalability



Source: Enterprise Strategy Group

ESG Tested

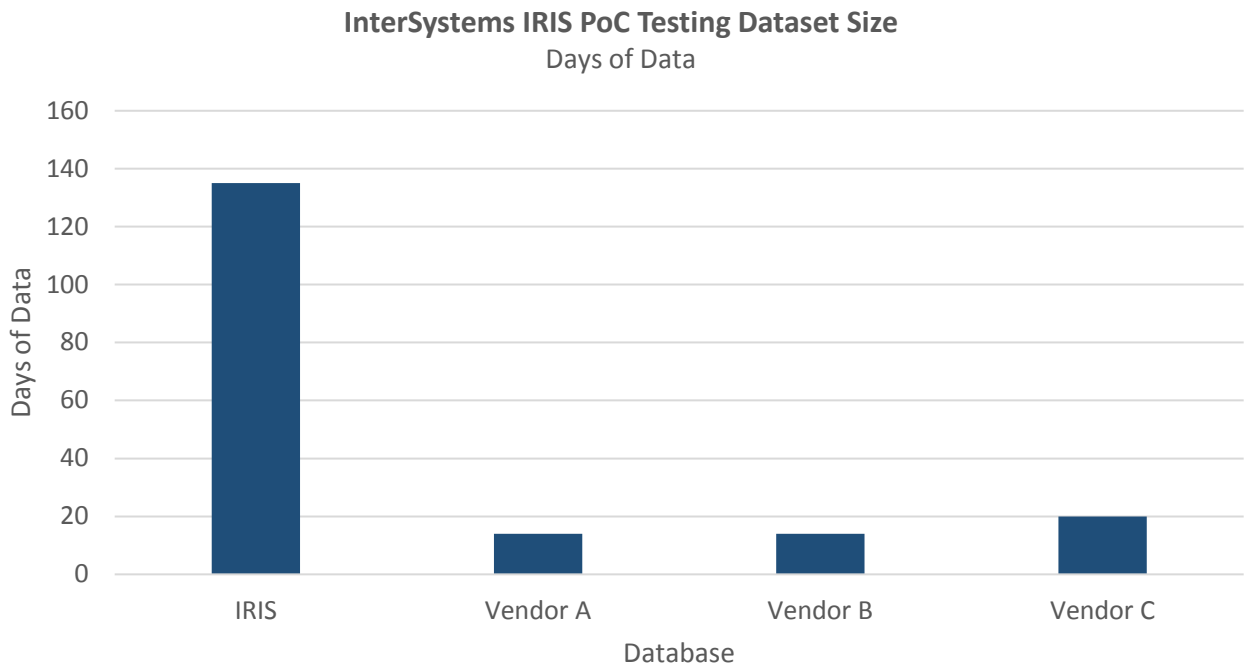
ESG audited results from tests performed on a database application proof of concept by a financial industry firm. Testing focused on analytical query performance of InterSystems IRIS and several other databases commonly used in the financial industry, leveraging thousands of financial instruments. The application supports multiple, firm-wide mission-critical lines of businesses; algorithmic trading apps require the SQL queries to successfully execute within strict performance constraints, while hundreds of users across the firm need ad hoc access to the query results. Millisecond performance is critical; if the queries cannot be completed within the required performance SLAs, the algorithmic trades will not execute and the firm cannot successfully carry out its trading strategies.

The test bed included InterSystems IRIS and three other common database deployments. *Table 1 shows that the InterSystems IRIS configuration used less compute infrastructure (including fewer VMs, cores, and RAM), while running queries against more data than the competitors (measured in days, GBs, and rows in Table 1 and Figure 4).*

Table 1. Test Bed Configurations

Database	Configuration	Days of Data	Dataset Size (GB)	Dataset Size (Rows)
InterSystems IRIS	**4 VMs: 8 cores, 96GB RAM, 400GB local storage**	**135**	**320**	**200,000,000**
Vendor A	8 VMs: 16 cores, 256GB RAM, 400GB local storage	14	33	20,740,000
Vendor B	8 VMs: 16 cores, 256GB RAM, 400GB local storage	14	33	20,740,000
Vendor C	3 dedicated physical servers, each with 24 cores, 256GB RAM, 2TB shared SAN storage	20	47	29,620,000

Figure 4. Dataset Size Comparison

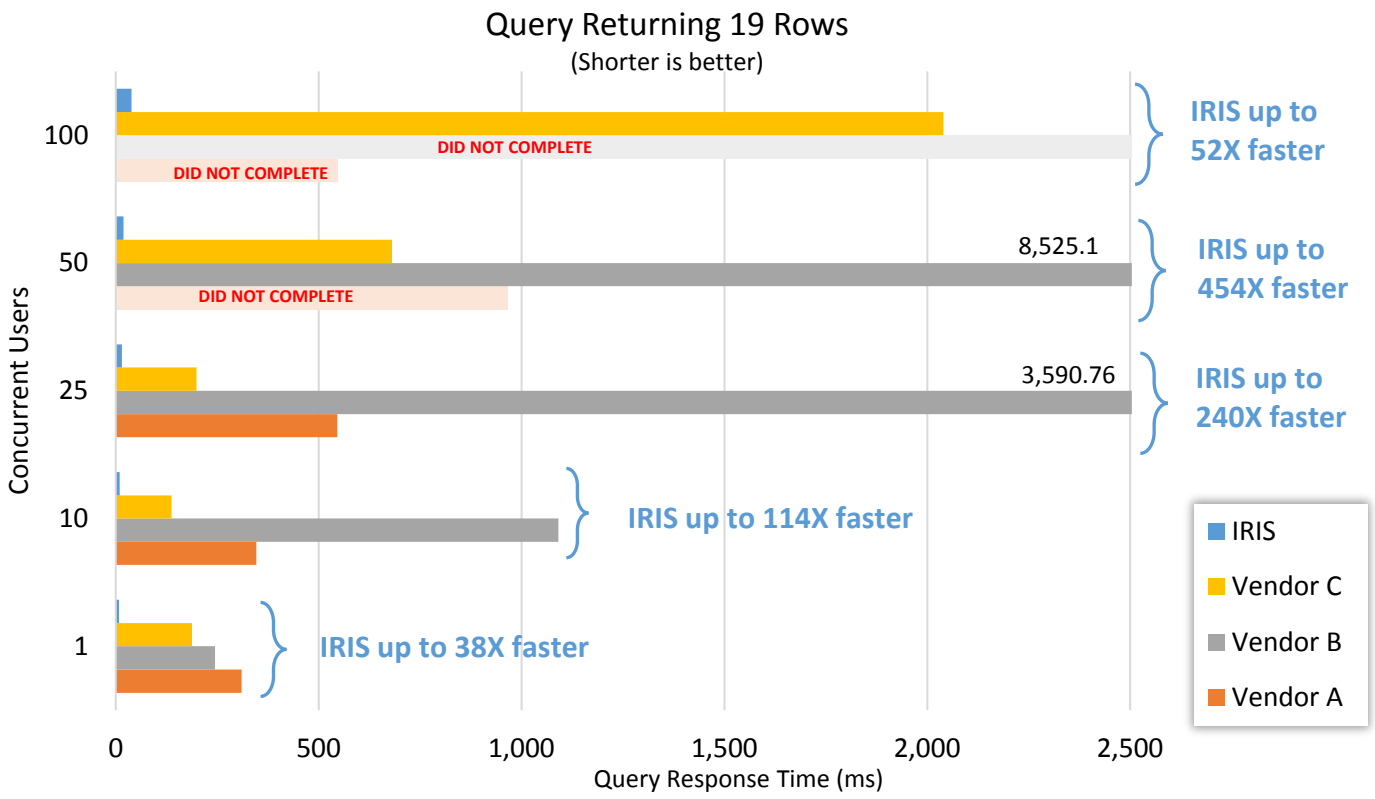


Source: Enterprise Strategy Group

Five queries were tested. All of them were complex joins among five to six tables with eight to ten predicates in the WHERE clause, with results sets containing between 19 and 6,504 rows, and each row containing hundreds of fields. Response times were measured as each query was executed against an increasing number of concurrent users: 1, 10, 25, 50, and 100.

Figure 5 presents the results of Query 1, returning 19 rows. Performance was measured by how many milliseconds each query required to complete as the number of concurrent users scaled.

Figure 5. Query 1 Performance as Concurrent Users Scaled



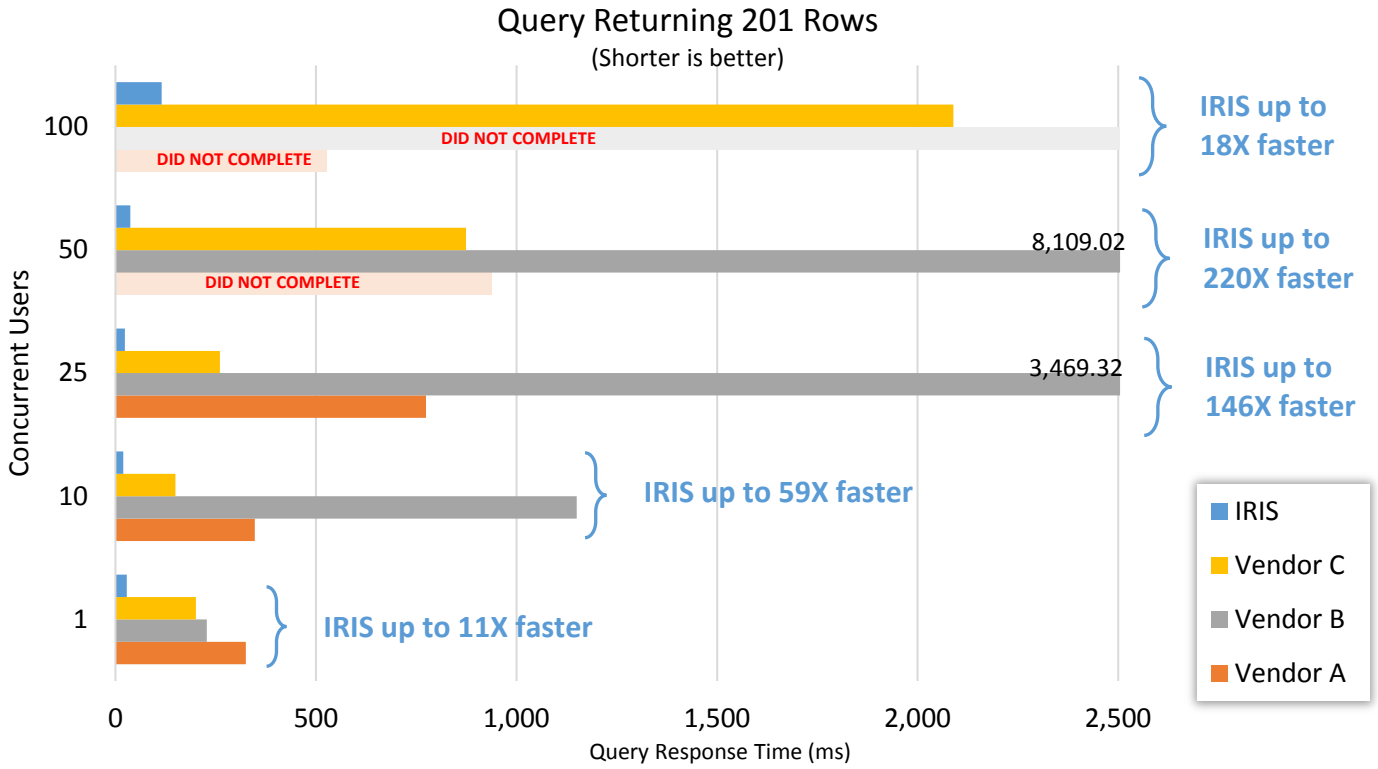
Source: Enterprise Strategy Group

Notes:

- InterSystems IRIS was operating on less infrastructure, and accessing 5-8X more data than the other vendors.
- Some solutions were unable to complete queries at scale, limiting their usefulness. These results are indicated with a notation and a pale version of the bar color.

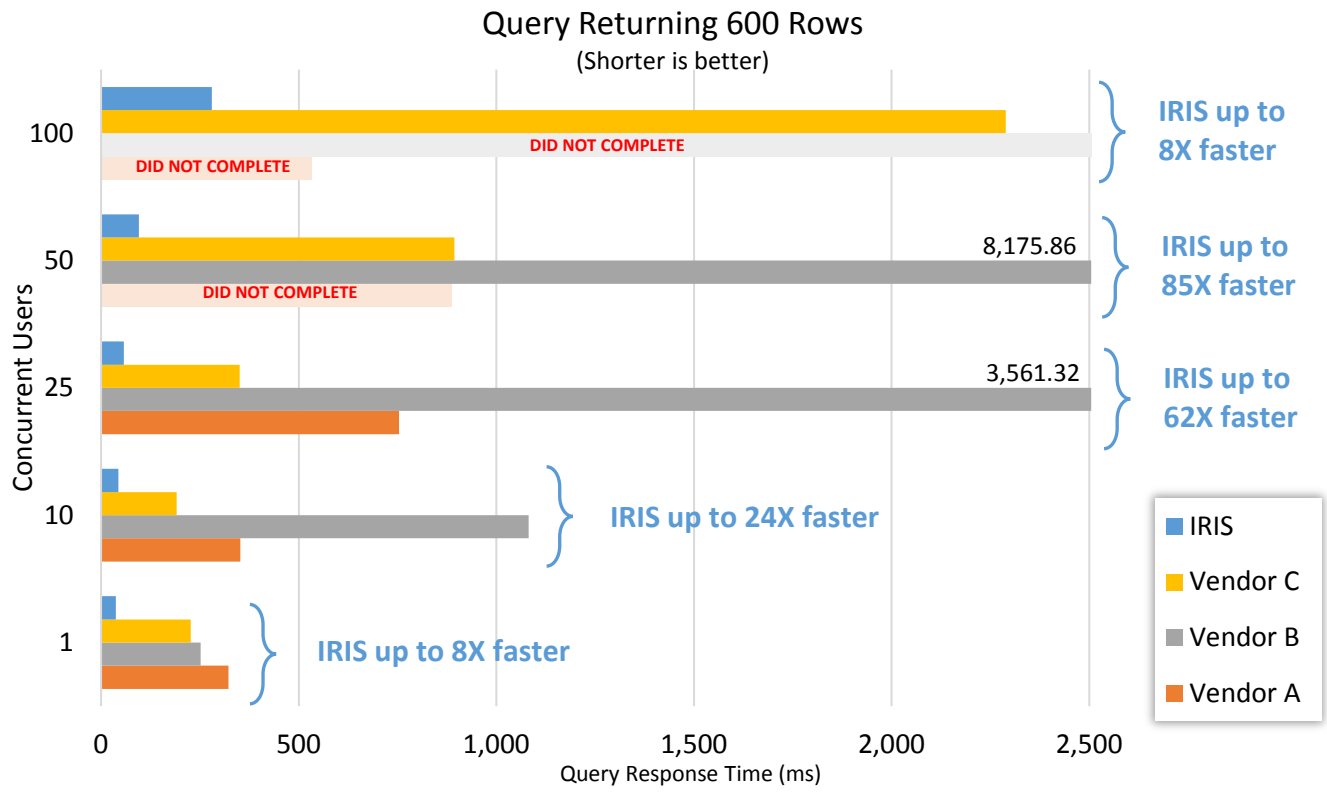
Queries 2-4 delivered similar results, with InterSystems IRIS being consistently faster. Query 2 returned 201 rows, Query 3 returned 600 rows, and Query 4 returned 984 rows; Figures 6, 7, and 8 show these results.

Figure 6. Query 2 Performance as Concurrent Users Scaled



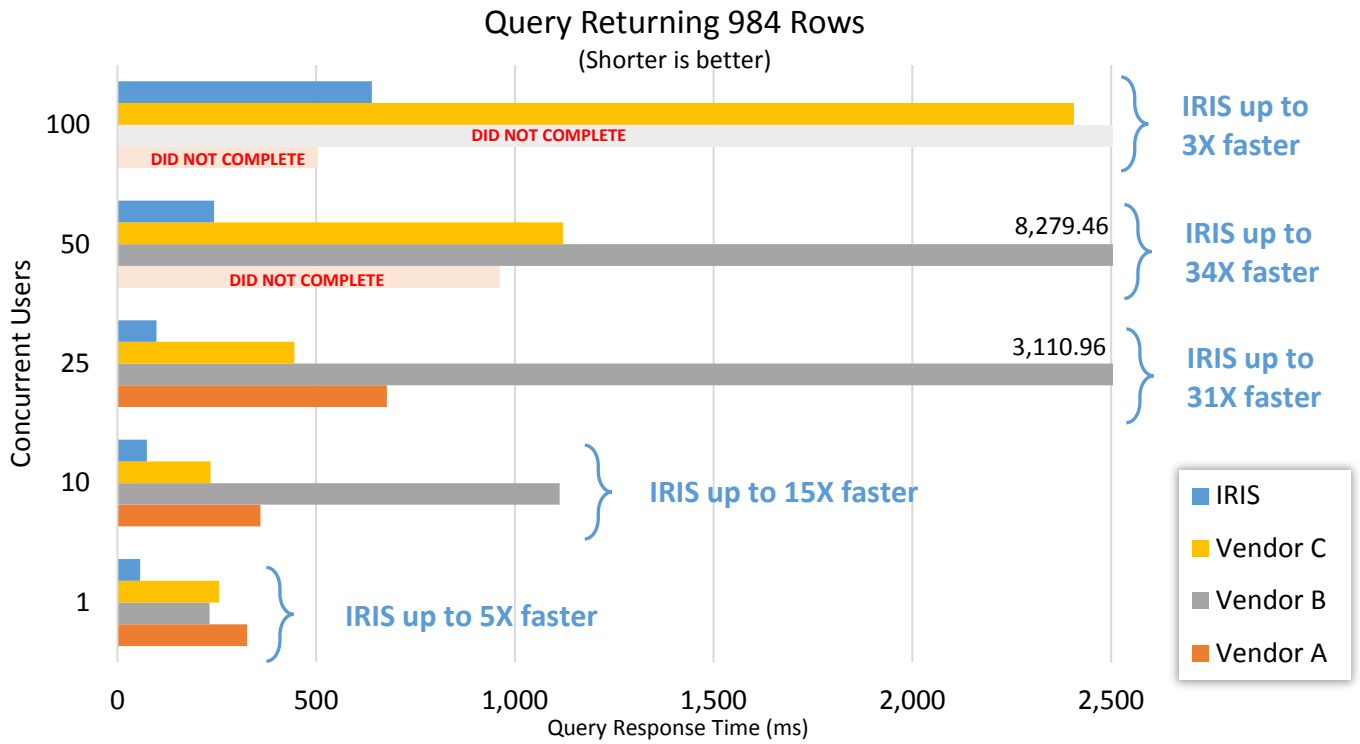
Source: Enterprise Strategy Group

Figure 7. Query 3 Performance as Concurrent Users Scaled



Source: Enterprise Strategy Group

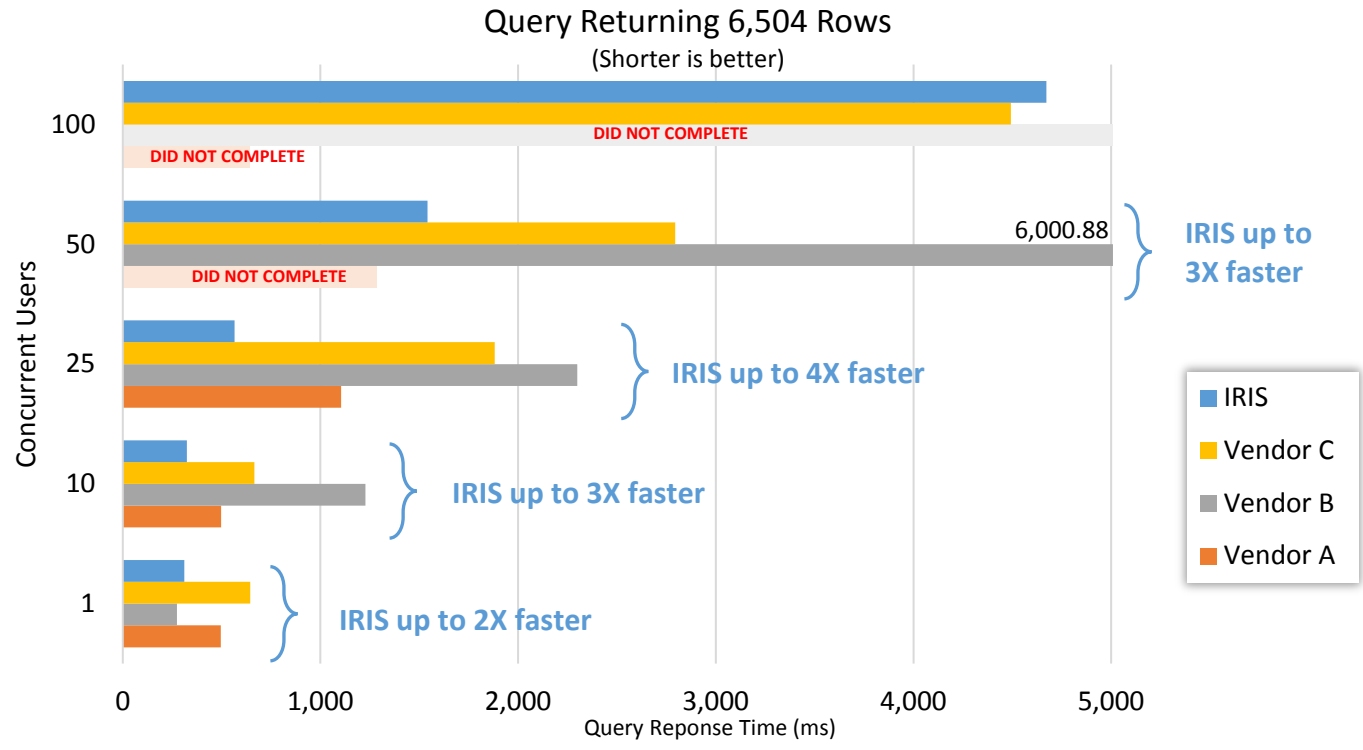
Figure 8. Query 4 Performance as Concurrent Users Scaled



Source: Enterprise Strategy Group

Figure 9 presents the results of Query 5, which returned 6,504 rows. For this more complex query, performance was more impacted as the number of concurrent users scaled.

Figure 9. Query 5 Performance as Concurrent Users Scaled



Source: Enterprise Strategy Group

Notes:

- InterSystems IRIS was operating on less infrastructure, and accessing more days of data than the other vendors.
- InterSystems IRIS performance ranged from 312 ms for a single user to 4,673 ms for 100 users, an increase of 14X.
- Vendor C performed slightly faster than InterSystems IRIS at 100 users, achieving 4,492 ms to InterSystems IRIS's 4,673.

Why This Matters

According to ESG research, organizations are focused on data analytics to achieve critical business objectives: better operational efficiency; reduced cost of business operations; better business decisions, product quality, and forecasting accuracy; and faster product time to market.[2] In short, they are counting on analytics to improve every part of their businesses, and they need fast answers to make accurate, timely decisions. The more data they can access, and the faster their queries return results, the better their decisions will be.

ESG validated that InterSystems IRIS provides the robust data management capability and performance needed for transactional and analytics workloads. IRIS performed consistently *faster* than other tested databases, while using *less infrastructure* and accessing *more data* than the others. InterSystems IRIS handled all the query types with fast performance, scaled easily, required less infrastructure, and experienced no failures. By accessing more data and delivering faster results, InterSystems IRIS can deliver better business insight.

The Bigger Truth

Organizations gain competitive advantage by leveraging information. Companies in finance, healthcare, manufacturing, telecommunications, and just about every other industry know that data analytics leads to better decision making. But it can be frustrating and time consuming to set up infrastructure silos and move data between transactional and analytics platforms, delaying insight and driving up costs. The more—and faster—organizations can understand customers, suppliers, and market trends, the better they can make business-driving decisions. The challenge is getting insight from as much data as possible, as fast as possible, and turning that into action quickly. Organizations can make real-time changes if they can execute real-time analytics on a flexible, scalable, cost-efficient, high-performance platform.

InterSystems IRIS is a unified, cost-efficient platform that was designed to improve decision making by letting organizations combine data-driven intelligence with real-time business processes. InterSystems IRIS is an innovative knowledge platform that speeds analytical queries on real-time data by eliminating the need for different transactional and analytics databases. Organizations can run transactional and analytic workloads using the same engine, without having to move, map, or translate data; the data can be represented in multiple ways at the same time, eliminating the need to deploy multiple point solutions, and resulting in faster performance.

ESG validated InterSystems IRIS performance against several common database solutions. Results demonstrated that InterSystems IRIS provided faster performance, leveraging more data, and using less infrastructure—enabling organizations to gain better insight, faster, for less money. With more in-depth insight, organizations can increase revenue, improve customer experiences, create tailored products and services, improve compliance, reduce risk, make operations more efficient, and reduce costs. If your organization is looking for the strong data management platform for data-centric insights, we recommend evaluating InterSystems IRIS.

[2] Source: ibid.

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.