

INTERSYSTEMS IRIS DATA PLATFORM: A UNIFIED PLATFORM FOR POWERING REAL-TIME, DATA-INTENSIVE APPLICATIONS

InterSystems IRIS Data Platform: A Unified Platform for Powering Real-Time, Data-Intensive Applications

Executive Summary

Organizations in every industry are looking to exploit the strategic and operational benefits of shortening and eliminating the delay between event, insight, and action. They also strive to embed data-driven intelligence into their real-time business processes.

When successful, turning these goals into reality offers myriad benefits, including:

- Delivering new and innovative business services,
- Increasing revenues,
- Improving customer experiences,
- Streamlining operations,
- Identifying and decreasing risk,
- Complying with new and ever-changing industry regulations, and
- Reducing costs.

This white paper describes the opportunities and challenges associated with shortening and eliminating these delays and presents a new technology that is simplifying the development, deployment, and maintenance of real-time, data-rich solutions in a range of industries.

Introduction

Organizations have more data at their disposal than ever. Yet many of them are challenged to gain insight from this data and act on it in real time for competitive advantage. Businesses are looking to capitalize on these opportunities by building real-time, data intensive applications using technology that can:

- Analyze real-time event and transactional data — along with large sets of historical and reference data — without delay.
- Support a range of data models and representations including relational, document, key-value, object, and unstructured text.
- Create seamless, real-time composite processes that integrate disparate applications and data sources.

- Scale to handle increasing workloads, data sizes, and user volumes.
- Embed analytic processing, including SQL queries, machine learning, predictive analytics, and natural language processing (NLP) into data driven applications.
- Leverage flexible options that support on premises, cloud, and hybrid deployments, and that support continuous delivery and DevOps methodologies.
- Provide these functional capabilities in a cost effective manner, without needing to hire a staff of experts in a broad range of disciplines.

Enabling the Real-Time Organization

Technology-industry analyst IDC recently interviewed more than 500 enterprises worldwide across a variety of industries. Over 75 percent reported that their inability to analyze current live data was actively inhibiting their ability to execute on new business opportunities. And more than half said it was limiting operational efficiencies.¹

The research found that 64 percent of companies have delays of five days or more before they can analyze operational data when using ETL (extract, transform, load) processing to move the data from their operational systems into a data warehouse.

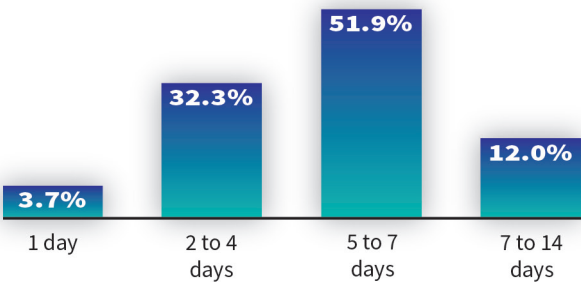


Figure 1: Average time to move operational data to the analytic database via ETL

Source: 3rd Platform Information Management Requirements Survey, IDC, October, 2016, n=502

For real-time applications that rely on change data capture (CDC) processing, organizations reported that 96 percent of their CDC processes take more than a minute before the data can be analyzed, and 65 percent take more than 10 minutes. That is too slow for critical real-time use cases, where milliseconds matter.

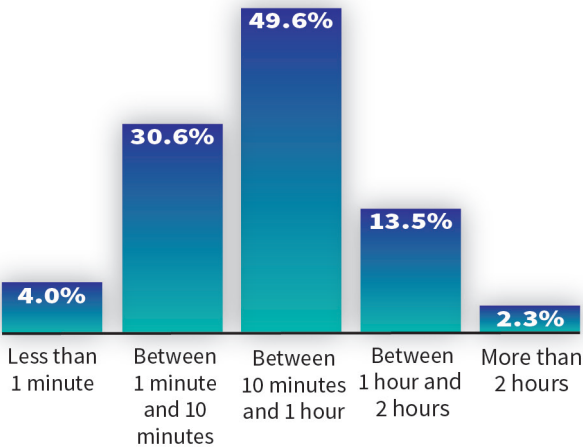


Figure 2: Average time to complete CDC processing

Source: 3rd Platform Information Management Requirements Survey, IDC, October, 2016, n=502

Applications that require real-time analytics on live data from a variety of sources are being implemented in virtually every industry:

- **Financial services**, for compliance with mandatory state and federal regulations, fraud detection, and risk management initiatives
- **Discrete manufacturing / original equipment manufacturing**, for predictive maintenance
- **Shipping and logistics**, for real-time container and shipment tracking
- **Retail**, for customer and visitor targeting and personalization
- **Public safety**, for situational awareness for first responders
- **Healthcare**, for personalized and proactive treatments at the point of care

These applications need a data platform that eliminates latency and complexity by supporting transactional and analytic workloads concurrently, in the same engine, without having to move, map, or translate the data.

InterSystems IRIS Data Platform™ delivers what is needed. It can incorporate multiple, disparate, and dissimilar data sources; support embedded real-time analytics; easily scale for growing data and user volume; interoperate seamlessly with other systems; and provide flexible, agile, DevOps-compatible deployment capabilities.

¹ "Choosing a DBMS to Address the Challenges of the Third Platform" (IDC, 2017)



InterSystems IRIS Data Platform

InterSystems IRIS Data Platform is a complete, unified platform that simplifies the development, deployment, and maintenance of real-time, data-rich solutions. It provides concurrent transactional and analytic processing capabilities; support for multiple, fully synchronized data models (relational, hierarchical, object, and document); a complete interoperability platform for integrating disparate data silos and applications; and sophisticated structured and unstructured analytics capabilities supporting batch and real-time use cases. The platform also provides an open analytics environment for incorporating best-of-breed analytics into InterSystems IRIS solutions, and it offers flexible deployment capabilities to support any combination of cloud and on-premises deployments.

InterSystems IRIS is a single product built from the ground up with a single architecture that supports a wide range of applications and scenarios.

InterSystems IRIS Data Platform provides these key features:

- Hybrid transactional/analytic processing to support real-time applications
- Multiple data models
- Embedded and open analytics
- Apache Spark integration
- Business Intelligence (BI)
- Ability to incorporate advanced analytics into real-time processes
- Natural Language Processing (NLP)
- Interoperability
- A unified development environment
- Flexible deployment options

Hybrid Transactional/Analytic Processing to Support Real-Time Applications

At the core of InterSystems IRIS Data Platform is a proven, enterprise-grade, distributed *hybrid transactional/analytic processing* (HTAP) database. It can ingest and store transactional data at very high rates while simultaneously processing high volumes of analytic workloads on real-time data (including ACID-compliant transactional data) and non-real-time data. This architecture eliminates the delays associated with moving real-time data to a different environment for analytic processing.

InterSystems IRIS's ability to deliver high performance at scale for HTAP is made possible by a number of technological innovations.

Better Sharding

InterSystems IRIS provides a powerful and efficient approach to performing queries on large data sets. An InterSystems IRIS sharded cluster can distribute workloads and data sets horizontally across a tier of application servers, partitioning the data in specific large tables across multiple nodes (called data shards²).

Sharding can benefit a wide range of applications but provides the greatest gains for use cases involving one or more of the following:

- Queries scanning very large data sets
- Complex queries on large data sets
- High data-ingestion rates and/or volumes

When an InterSystems IRIS sharded cluster receives an application query, the **shard master** pushes decomposed queries to the data shards for parallel execution, aggregates the results returned by the individual shards, and returns the final result to the application. If the data from other shards is required for a shard to complete its work, the shard can access **just the data it needs** on the other shards directly, without involving the shard master.

The result is that InterSystems IRIS achieves consistent high performance, efficiency, and reliability, even for complex queries involving multiple tables. In contrast, many other database platforms that support sharded architectures rely on broadcasting the entire table, which can result in performance penalties and timeouts.

Since sharding creates disjoint partitions of the data, each data server's cache is fully independent, and adding data servers linearly increases the cluster's overall memory. Therefore, through appropriate sizing, InterSystems IRIS can achieve the performance benefits of in-memory databases without requiring all data to fit in memory.

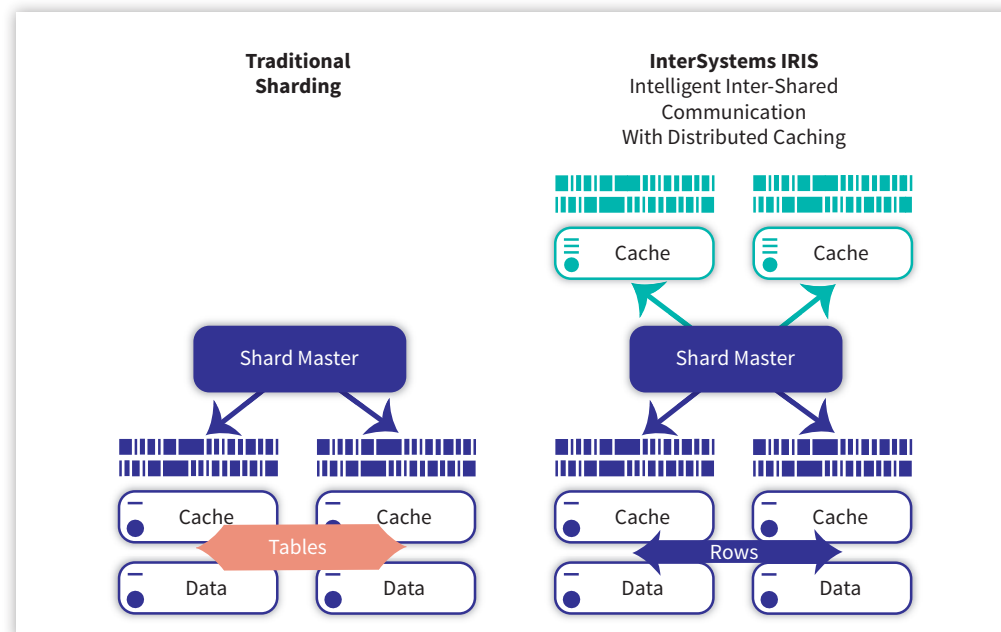


Figure 3: Intelligent Inter-Shard Communication for Analyzing Large, Distributed Data Sets

² A data shard is an InterSystems IRIS instance that stores one horizontal partition of each sharded table defined on the cluster's shard master. The node hosting this instance is called a shard data server.

An InterSystems IRIS sharded cluster provides additional performance advantages:

- The transparent parallel load capability of the InterSystems IRIS Java Database Connectivity (JDBC) driver supports the use of Java-based tools for very fast data ingestion, in parallel across the shards.
- When large, multiuser query workloads would create a bottleneck on the shard master, a tier of application servers can be added in front of the shard master to scale for user volume through distributed application logic and caching.

Because sharding is transparent to the application, it requires little or no change to application code. The distinction between sharded and non-sharded tables is entirely transparent to the application; it is strictly a design time consideration.

The InterSystems IRIS architecture enables complex multi-table joins to identify patterns and relationships in distributed, partitioned data sets without requiring co-sharding³, without replicating data, and without requiring entire tables to be broadcast across networks.

³ Cosharded data refers to distributed data that is partitioned on a common key.

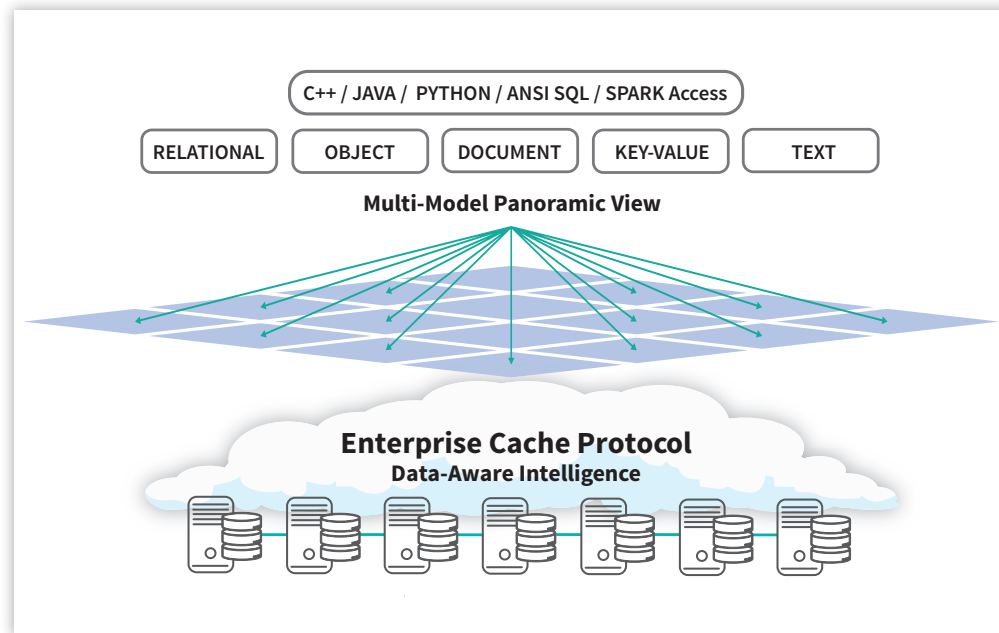


Figure 4: Unified Access to Multi-Model, Distributed Data With InterSystems IRIS

Higher Performance, Lower Cost

In addition to performing efficient analytical processing, InterSystems IRIS processes concurrent transactional and analytic workloads with high performance and at scale. There is no need to move transactional data to a different environment for analysis. InterSystems IRIS can process transactions, make the data durable on persistent storage, and make the transactional data available for analytic queries all within tens of nanoseconds on commercially available hardware.

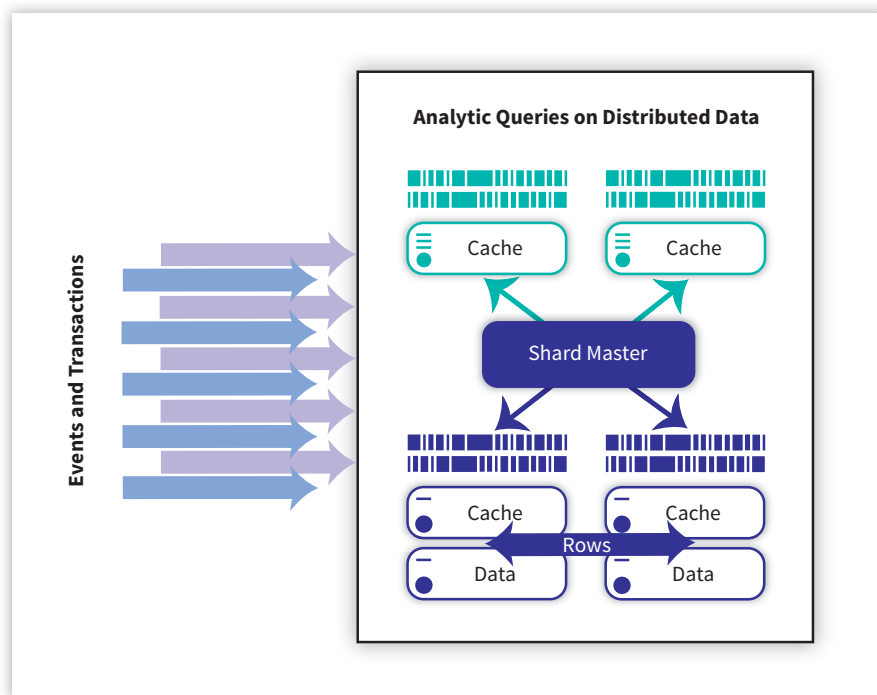


Figure 5: Horizontally Distributed HTAP

InterSystems IRIS supports direct shared memory writers and client/server distributed SQL processing simultaneously to support high-performance concurrent transactional/analytic use cases. As a result, InterSystems IRIS can reliably process and analyze real-time data in combination with data stored in distributed and partitioned data sets, in less time and at lower operational cost.

For high availability of both non-sharded and sharded tables, all nodes storing data can be mirrored. Compute nodes can be easily added and removed to support user workload fluctuations. InterSystems IRIS provides strong enterprise-level security; integration with Kerberos, LDAP, and KMIP (Key Management Interoperability Protocol); role-based access control; and encryption for data in transit and at rest.

Multiple Data Models

InterSystems IRIS is built on a true multi-model database. This means the data is stored once and can be accessed via multiple data models, including relational and object models, which are always synchronized. This eliminates the need to duplicate data or provide mappings between different representations (e.g., object-to-relational mapping). The ability to natively support multiple data types enables organizations to model, store, and use data in the most appropriate format and representation, for flexible solution development, higher performance, and reduced complexity.

How Important are the New Data Types?

(Rating scale: 1 = Not very important, 5 = very important)

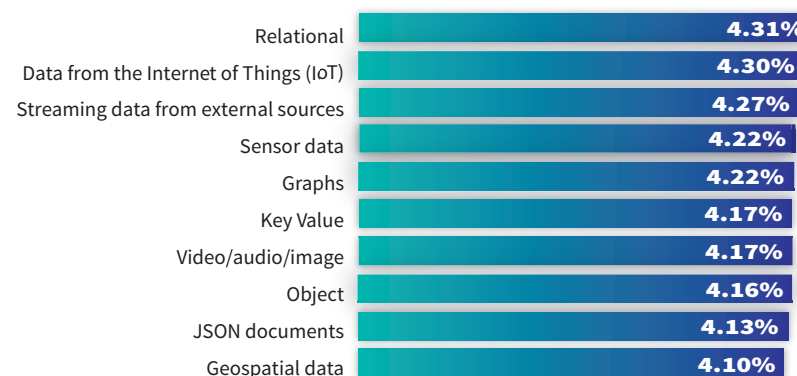


Figure 6: Importance of Supporting Various Data Types in a Data Platform

Source: 3rd Platform Information Management Requirements Survey, IDC, October, 2016, n=502

Embedded and Open Analytics

InterSystems IRIS supports a wide range of analytics to meet the varied requirements of today's data-intensive, real-time applications. InterSystems IRIS provides embedded state-of-the-art analytics capabilities for distributed SQL, BI, and NLP and can incorporate a wide range of third-party and open-source analytics packages as needed.

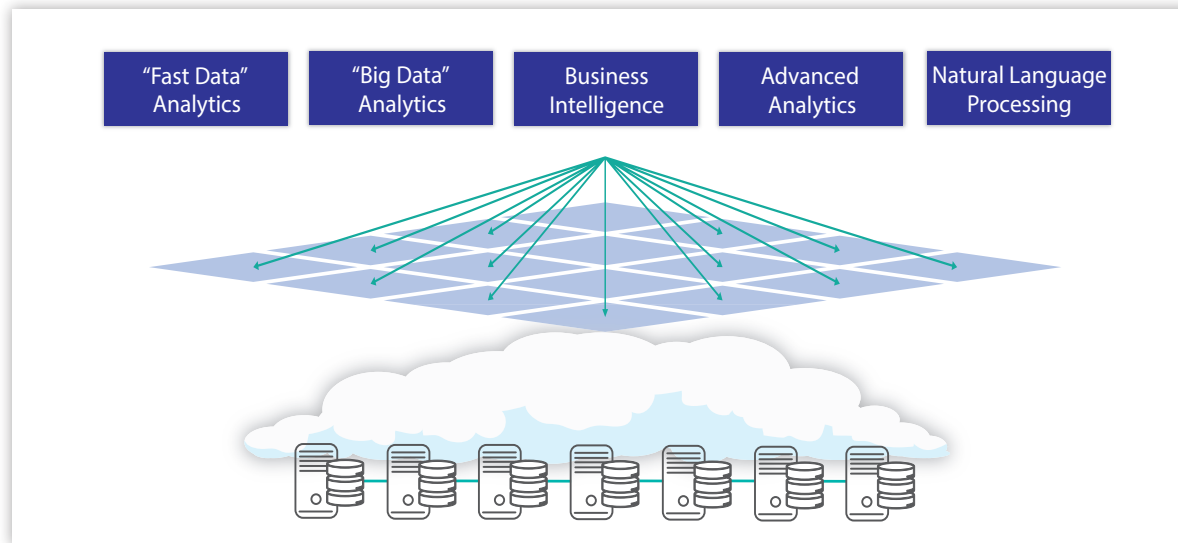


Figure 7: InterSystems IRIS Embedded and Open Analytics Capabilities

According to a 2017 survey of large businesses by research firm Gartner, 45% of the 1,931 respondents said they planned to use data mining and predictive analytics, 39% planned to use Apache Hadoop or Spark, and 25% planned to use the advanced analytics capabilities provided by Apache Hadoop or Spark.⁵

Advanced analytics technologies are rapidly gaining adoption. These approaches and technologies include machine learning, predictive analytics, artificial intelligence, and real-time big-data processing frameworks like Apache Spark.

In addition to its real-time (HTAP) and big (distributed) data processing capabilities, InterSystems IRIS provides the following analytic capabilities and integrations.

Apache Spark Integration

Apache Spark is a high-performance, open-source cluster-computing framework and is often used when performance on large distributed data sets is critical. Apache Spark can be 100 times faster than Apache Hadoop (MapReduce), and many common machine learning and statistical algorithms are available.

InterSystems IRIS integrates directly with Apache Spark via a shard-aware native Spark connector, so that InterSystems IRIS applications can incorporate Spark processing, and Spark applications can incorporate distributed data from InterSystems IRIS. The Apache Spark connector presents the data shards of an InterSystems IRIS sharded cluster as a native partition for the highest performance. The connector is aware of the partitioned nature of the InterSystems IRIS database, allowing the Apache Spark worker nodes to automatically connect directly to the shards, and work in parallel on disjoint pieces of data. These parallel, direct connections also allow much higher throughput (since less data needs to be passed through each connection) and support high-speed data ingestion to the sharded cluster.

⁵ Rita L. Sallam, et al., "Survey Analysis: BI and Analytics Spending Intentions, 2017" (Gartner, 2017)

Business Intelligence

InterSystems IRIS provides fully integrated capabilities for BI modeling, analysis, and end-user dashboards. A BI model represents dimensions that are meaningful to the business, including aggregate concepts (such as product line, sales area, market segment, and so on) and numeric measures (such as revenue, expenses, year-to-year growth, defect rate, and so on). An InterSystems IRIS BI model can be based directly on transactional data and other data that might be needed. A fully automated synchronization option avoids the need for ETL processing. Drag-and-drop analysis capabilities enable nontechnical users to examine the data at any level and perform complex queries with ease. InterSystems IRIS dashboards can display live business metrics and give restricted analysis options to users.

Ability to Incorporate Advanced Analytics Into Real-Time Processes

Organizations can incorporate predictive models created by data mining and machine learning algorithms using external tools and applications through InterSystems IRIS embedded support for the Predictive Model Markup Language (PMML). PMML is an XML standard that fully defines all the parameters of a predictive model developed using an external analytics application or framework. When a PMML model is loaded into InterSystems IRIS, native code is generated to allow execution of the model in real time, without requiring any

external tool or the performance-inhibiting passing of data across systems. This integration enables predictive models created by data scientists and other specialists to be seamlessly incorporated into data-processing pipelines and business processes within InterSystems IRIS.

Natural Language Processing

InterSystems IRIS provides NLP capabilities that infer meaning and sentiment from natural language text. InterSystems IRIS can automatically identify concepts and relationships in text without requiring upfront work or domain knowledge. These advanced NLP capabilities are embedded in InterSystems IRIS and can be included in business processes, enabling organizations to include information from notes fields, social media, and other sources in data-rich applications.

Since there are many different kinds of specialized NLP tools, each with a specific type of functional or domain applicability, some applications may require these tools to be used in sequence. InterSystems IRIS supports the Apache Unstructured Information Management Architecture (UIMA) standard, which enables a standards-based pluggable NLP pipeline to be defined and executed. Apache UIMA support brings open interoperability to the NLP capabilities in InterSystems IRIS.

46% of large businesses planned to incorporate sentiment analysis of unstructured content into their applications in 2017.⁶

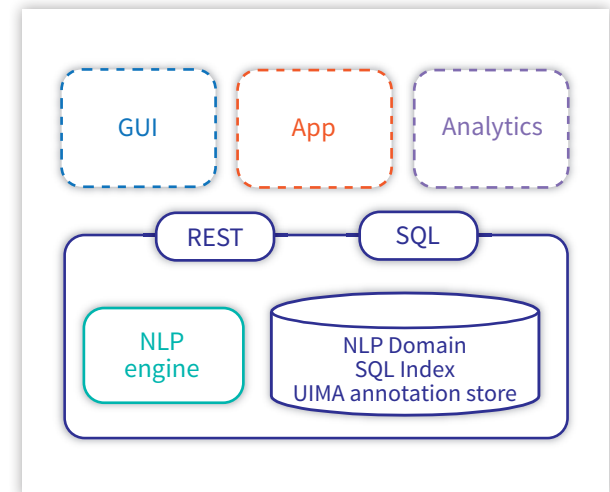


Figure 8: InterSystems IRIS Natural Language Processing Capabilities

⁶ Rita L. Sallam, et al., "Survey Analysis: BI and Analytics Spending Intentions, 2017" (Gartner, 2017)

Interoperability

InterSystems IRIS provides a complete set of native integration and interoperability features. It provides out-of-the-box connectivity and data transformations for a wide range of packaged applications, databases, industry standards, protocols, and technologies. Flexible data-transformation capabilities enable InterSystems IRIS to resolve differences in semantics and data

schemas that exist between applications or services.

Application developers can create seamless business processes that connect with internal and external data sources, applications, and services. InterSystems IRIS provides graphical tooling to visually diagram processes, rules, and workflows, allowing developers to focus on the logical interactions between systems, minimizing concerns about application interfaces, adapters, or middleware mechanisms. The graphical models

enable collaboration between the lines of business and IT, resulting in faster development of solutions that meet business requirements, and easier modification and extension of existing processes. The embedded role-based workflow engine supports manual interactions in business processes, automating the distribution of tasks among users and incorporating their decisions and actions.

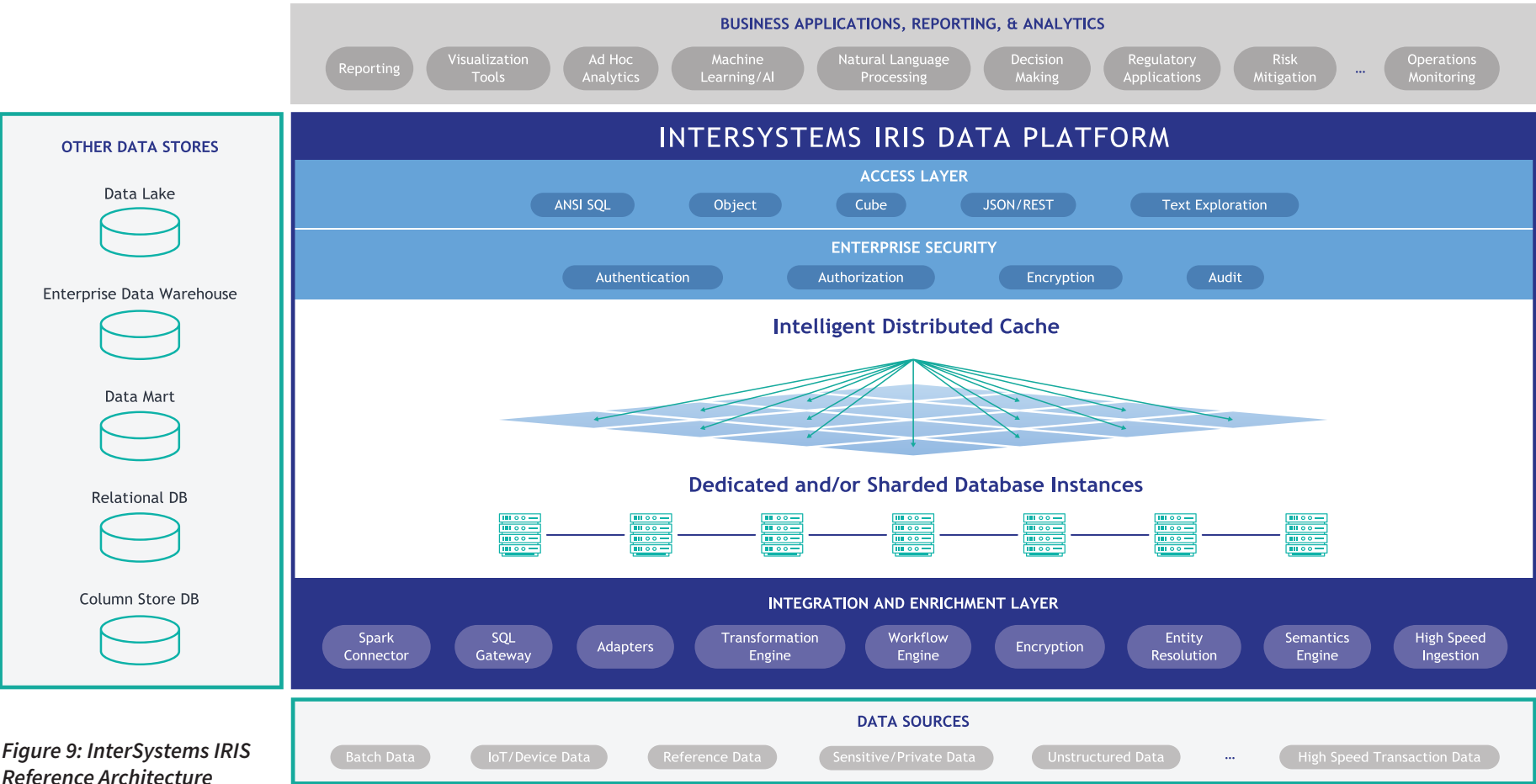


Figure 9: InterSystems IRIS Reference Architecture

Since InterSystems IRIS includes embedded database and analytics capabilities, sophisticated analytics can be seamlessly incorporated into business processes, leveraging data stored in the database as well as real-time data. All data, including in-flight data or data associated with long-running asynchronous transactions, can be automatically persisted in the database and available for reporting and analysis.

The platform supports a wide range of standards used in various industries, such as healthcare, financial services, retail, and telecommunications, including REST architectures and web services (e.g., JSON, XML, XPATH, XSLT, SOAP, and DTDs).

Unified Development Environment

The unified graphical and code-based environment of InterSystems IRIS delivers a consistent representation of diverse programming models, programming interfaces, and data formats, providing a single development environment across all functionality.

Flexible Deployment Options

InterSystems IRIS provides a simple, intuitive way to provision and deploy services on cloud-based and on-premises infrastructures. InterSystems IRIS delivers the benefits of infrastructure as code, immutable infrastructure, and containerized deployment of InterSystems IRIS-based applications. It eliminates the need for major investments in new technology and associated training, as well as trial-and-error system configuration and management efforts.

InterSystems IRIS allows organizations to take advantage of the efficiency, agility, and repeatability that cloud computing and containerized software offer, without requiring major development or retooling. It can also provision and deploy InterSystems IRIS configurations on existing virtual and physical clusters, and it supports deployment of containers on enterprise-level operating system platforms, including preexisting infrastructure and commercial cloud platforms.

Conclusion

InterSystems IRIS is a complete, unified data platform that simplifies the development, deployment, and maintenance of real-time, data-rich solutions. InterSystems IRIS provides concurrent transactional and analytic processing capabilities; support for multiple, fully synchronized data models (including relational, hierarchical, object, and document); a complete interoperability platform for integrating disparate data silos and applications; and sophisticated structured and unstructured analytics capabilities supporting both batch and real-time use cases. The platform also provides an open analytics environment for incorporating best-of-breed analytics into InterSystems IRIS solutions and offers flexible deployment capabilities to support any combination of cloud and on-premises deployments.

InterSystems IRIS is being used in multiple industries to help deliver a range of important strategic and operational benefits, by leveraging more data while eliminating delays between event, insight, and action.

We are also proud to offer the InterSystems IRIS Experience, a self-directed, hands-on opportunity to discover for yourself the power of InterSystems IRIS. Learn more at **InterSystems.com/Experience**



[InterSystems.com](https://www.intersystems.com)