



InterSystems IRIS Overview

Version 2019.4
2019-12-18

InterSystems IRIS Overview

InterSystems IRIS Data Platform Version 2019.4 2019-12-18

Copyright © 2019 InterSystems Corporation

All rights reserved.



InterSystems, InterSystems Caché, InterSystems Ensemble, InterSystems HealthShare, HealthShare, InterSystems TrakCare, TrakCare, InterSystems DeepSee, and DeepSee are registered trademarks of InterSystems Corporation.



InterSystems IRIS Data Platform, InterSystems IRIS, InterSystems iKnow, Zen, and Caché Server Pages are trademarks of InterSystems Corporation.

All other brand or product names used herein are trademarks or registered trademarks of their respective companies or organizations.

This document contains trade secret and confidential information which is the property of InterSystems Corporation, One Memorial Drive, Cambridge, MA 02142, or its affiliates, and is furnished for the sole purpose of the operation and maintenance of the products of InterSystems Corporation. No part of this publication is to be used for any other purpose, and this publication is not to be reproduced, copied, disclosed, transmitted, stored in a retrieval system or translated into any human or computer language, in any form, by any means, in whole or in part, without the express prior written consent of InterSystems Corporation.

The copying, use and disposition of this document and the software programs described herein is prohibited except to the limited extent set forth in the standard software license agreement(s) of InterSystems Corporation covering such programs and related documentation. InterSystems Corporation makes no representations and warranties concerning such software programs other than those set forth in such standard software license agreement(s). In addition, the liability of InterSystems Corporation for any losses or damages relating to or arising out of the use of such software programs is limited in the manner set forth in such standard software license agreement(s).

THE FOREGOING IS A GENERAL SUMMARY OF THE RESTRICTIONS AND LIMITATIONS IMPOSED BY INTERSYSTEMS CORPORATION ON THE USE OF, AND LIABILITY ARISING FROM, ITS COMPUTER SOFTWARE. FOR COMPLETE INFORMATION REFERENCE SHOULD BE MADE TO THE STANDARD SOFTWARE LICENSE AGREEMENT(S) OF INTERSYSTEMS CORPORATION, COPIES OF WHICH WILL BE MADE AVAILABLE UPON REQUEST.

InterSystems Corporation disclaims responsibility for errors which may appear in this document, and it reserves the right, in its sole discretion and without notice, to make substitutions and modifications in the products and practices described in this document.

For Support questions about any InterSystems products, contact:

InterSystems Worldwide Response Center (WRC)

Tel: +1-617-621-0700

Tel: +44 (0) 844 854 2917

Email: support@InterSystems.com

Table of Contents

InterSystems IRIS Overview.....	1
1 Sharding to Distribute Data	1
2 Multiple Data Models	3
3 Embedded and Open Analytics	3
3.1 Apache Spark Integration	5
3.2 Business Intelligence	5
3.3 Predictive Model Markup Language Support	5
3.4 Natural Language Processing	5
4 Interoperability	5
5 Unified Development Environment	5
6 Flexible Deployment Options	6
7 Summary	6

InterSystems IRIS Overview

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)
- Complete interoperability platform for integrating disparate data silos and applications
- Sophisticated structured and unstructured analytics capabilities supporting batch and real-time use cases

The platform also provides an open analytics environment for incorporating other analytics packages into InterSystems IRIS solutions, and 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.

Key capabilities include:

- Hybrid transaction-analytic processing (HTAP) to support real-time applications
- Multiple data models
- Embedded and open analytics
- Apache Spark integration
- Business intelligence
- Ability to incorporate advanced analytics into real-time processes
- Natural language processing
- Interoperability
- Unified development environment
- Flexible deployment options

At the core of InterSystems IRIS is a proven, enterprise-grade, distributed, hybrid transactional-analytic processing (HTAP) database. It can ingest, process, and store transactional data at very high rates while simultaneously processing high volume analytic workloads involving both real-time data (including ACID-compliant transactional data) and historical data. This architecture enables InterSystems IRIS to process transactional data and make it durable on persistent storage and available for analytic queries within tens of nanoseconds on commercially available hardware, eliminating the delays associated with moving real-time data to an analytic environment. InterSystems IRIS is built on a distributed architecture to support large data volumes, enabling organizations to analyze very large data sets while simultaneously processing large amounts of real-time transactional data.

1 Sharding to Distribute Data

InterSystems IRIS provides a powerful and efficient approach to performing queries on large data sets. An InterSystems IRIS sharded cluster distributes workloads and data sets horizontally, partitioning large tables and their associated indexes across multiple InterSystems IRIS instances called data shards. Each data shard stores one horizontal partition of each sharded table defined on the cluster's shard master; the node hosting the data shard instance is called a shard data server.

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

- Relatively large data sets, queries that return large amounts of data, or both.
- Complex queries that do large amounts of data processing, such as those that scan a lot of data on disk or involve significant compute work.
- High-volume or high-speed data ingestion, or a combination.

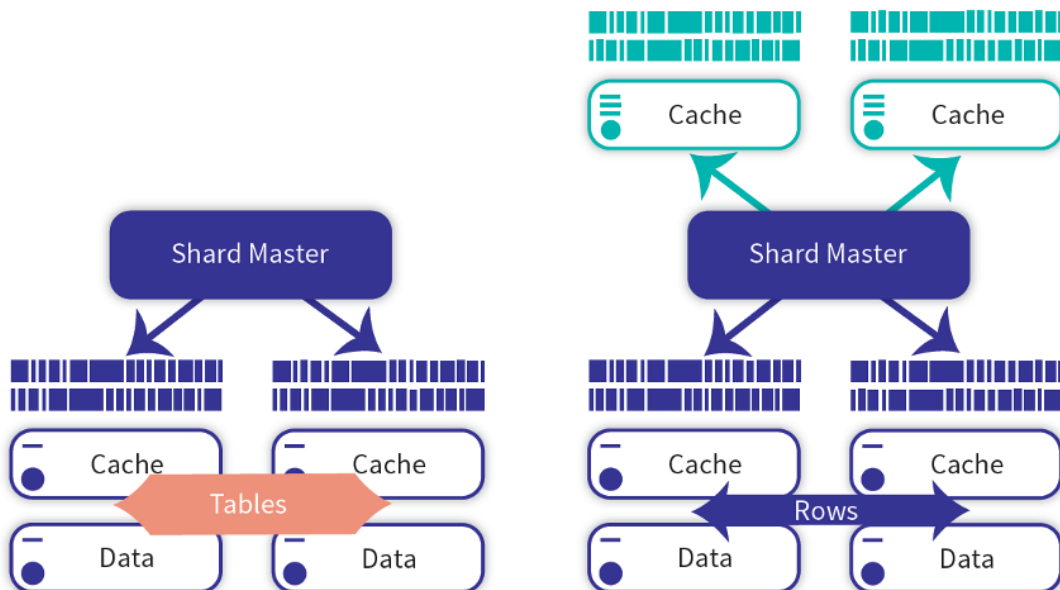
When an InterSystems IRIS sharded cluster receives an application query, the shard master pushes decomposed shard-local queries to the data shards for parallel execution, then combines the results from the individual shards and returns the final result to the application. Unlike other sharding solutions, a data shard does not need to store all the data it needs to complete its work; because the shards communicate directly with each other through the sharding manager, each shard can access the data it needs on the others.

Since sharding creates disjoint partitions of the data, each shard data server's cache is fully independent; adding shard data servers linearly increases the cluster's overall caching capacity. This allows InterSystems IRIS to achieve the performance benefits of in-memory databases without requiring the entire working set to fit in a single system's memory.

An InterSystems IRIS sharded cluster provides these additional performance advantages:

- The transparent parallel load capability of the InterSystems IRIS 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.

The following illustration compares traditional sharding solutions to an InterSystems IRIS sharded cluster with horizontal partition of tables, intershard communication through the sharding manager, and distributed caching:



InterSystems IRIS sharding requires little or no change to application code. The distinction between sharded and nonsharded tables is entirely transparent to the application; it is strictly a design time consideration.

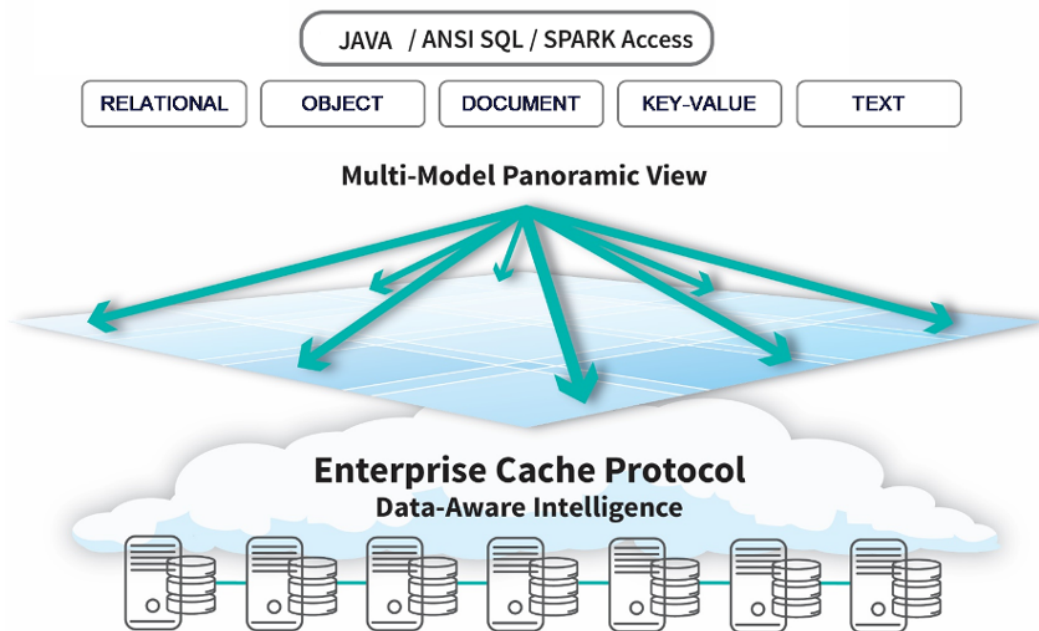
InterSystems IRIS architecture enables complex multiple table JOINS to identify patterns and relationships in distributed, partitioned data sets without requiring cosharding (requiring a common key), without replicating data, and without requiring entire tables to be broadcast across networks.

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 identify information and patterns using real-time data in combination with data stored in distributed data sets, in less time and with less operational cost.

For high availability of both nonsharded and sharded tables, all nodes storing data can be mirrored. Host 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; role-based access control; and encryption for data in transit and at rest.

2 Multiple Data Models

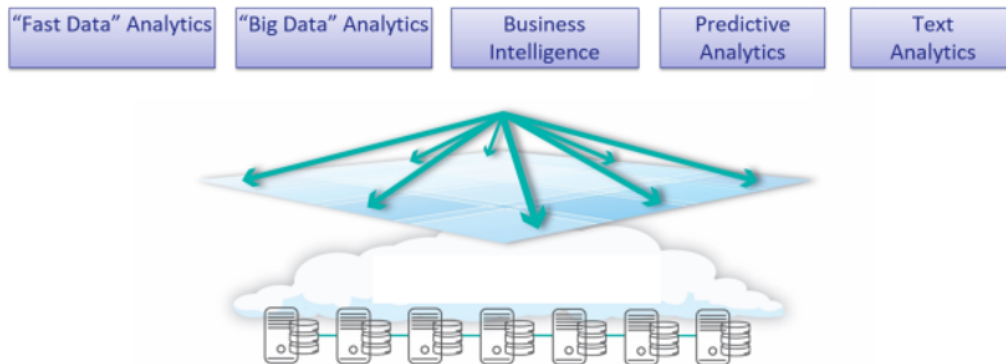
InterSystems IRIS is built on a multiple model database. This means the data is stored once and can be accessed via multiple models including relational and object models that are always synchronized. This eliminates the need to duplicate data or provide mappings between different representations (object-to-relational 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 less complexity. The following figure illustrates how you can access data in multiple modes:



3 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, business intelligence, and natural language processing, and can incorporate a wide range of third party and open source analytics packages as needed.

Advanced analytics technologies for extracting information from larger and more diverse data sets 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. InterSystems IRIS has multiple analytics options as illustrated:



3.1 Apache Spark Integration

Apache Spark is a high performance open source cluster computing framework, 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 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 (as less data needs to be passed through each connection), and support high-speed data ingestion to the sharded cluster.

3.2 Business Intelligence

InterSystems IRIS provides fully integrated support for business intelligence (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 is based directly on transactional data and any other data that might be needed. A fully automated synchronization option avoids the need for ETL processing. Drag and drop analysis capabilities enable non-technical users to examine the data at any level, performing complex queries with ease. InterSystems IRIS dashboards provide a way to display live business metrics and give restricted analysis options to other users.

3.3 Predictive Model Markup Language Support

By providing embedded support for Predictive Model Markup Language (PMML), InterSystems IRIS allows you to incorporate predictive models created by data mining and machine learning algorithms using external tools and applications. 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 in InterSystems IRIS, native code is generated to allow execution of the model in real-time, without requiring any external tool or 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.

3.4 Natural Language Processing

InterSystems IRIS provides natural language processing (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 natural language processing 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 natural language processing 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 natural language processing capabilities in InterSystems IRIS.

4 Interoperability

InterSystems IRIS provides a complete set of embedded integration and interoperability capabilities. 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 resolution of differences in semantics and data schemas 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. Concerns about application interfaces, adapters, or middleware mechanisms are minimized. The graphical models facilitate 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.

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, REST architectures and web services (e.g., JSON, XML, XPATH, XSLT, SOAP, DTDs).

5 Unified Development Environment

InterSystems IRIS provides a unified graphical and code-based environment that simplifies and accelerates development and maintenance of real-time, data-rich solutions. It provides a consistent representation of diverse programming models, programming interfaces, and data formats, providing a single development environment across all functionality.

6 Flexible Deployment Options

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

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

7 Summary

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.