

An InterSystems Benchmark Report:

InterSystems EMPI



RELEASE:
NOVEMBER 2025

 **InterSystems**[®]
Creative data technology

InterSystems EMPI has demonstrated a sustained average bulk ingestion throughput of **over 5400 transactions per second**.

Executive Summary

InterSystems EMPI™ is a next-generation enterprise master person index – an automated, easily integrated solution for identity resolution. It is available as a standalone solution or can seamlessly pair with **InterSystems HealthShare Unified Care Record®**, merging multi-source data into a single, accurate longitudinal health record.

We have recently completed a performance and scalability benchmark of our latest version of InterSystems EMPI. The test profiles consisted of two distinct workloads – **Bulk Ingestion** and **Person Search**. Testing focused on several key areas including bulk data ingestion, large data volume, and a high rate of sustained transactions per second (TPS). This document describes the observed characteristics and also provides general configuration and sizing guidelines for InterSystems EMPI.

The benchmark simulates workloads that have been designed to match typical production-like deployments starting with massive bulk loading demographics of a 308-million persons data set generated using Synthea (the most popular synthetic data generator for FHIR). Once ingestion is completed, various person search workloads were conducted mixing concurrency levels to demonstrate scale and performance along with differences between deterministic and probabilistic search methods. Details of the workload simulations are described in the Workload Description and Methodology section of this report.

The 2025.2 version of InterSystems EMPI has demonstrated a sustained average **bulk ingestion throughput of over 5400 transactions per second**. This high pace of ingestion allowed for over 308 million persons to be loaded in ~13.5 hours. Person search round trip average latencies are as low as 100ms for deterministic and 400ms for probabilistic searches for an aggregate **average 263ms person search response time** demonstrating massive national-level scale of population volumes with consistent sub-second response time searches.

Workload Description and Methodology

Apache JMeter was used as the load generation tool. JMeter is an open-source Java-based tool used for load and performance testing of static and dynamic resources. It provides the ability to simulate heavy loads with a multi-threading framework, perform tests on various protocols, scenarios, and generate throughput and latency reports.

High-Level Benchmark Architecture

The benchmark test consisted of three JMeter driver servers that communicated to standard RESTful API endpoints on the InterSystems EMPI instance. The RESTful API endpoints were presented through Apache HTTPD 2.4 web server and integrated with InterSystems Web Gateway to communicate with InterSystems EMPI. Figure 1 illustrates the high-level diagram of both the ingestion and search workloads.

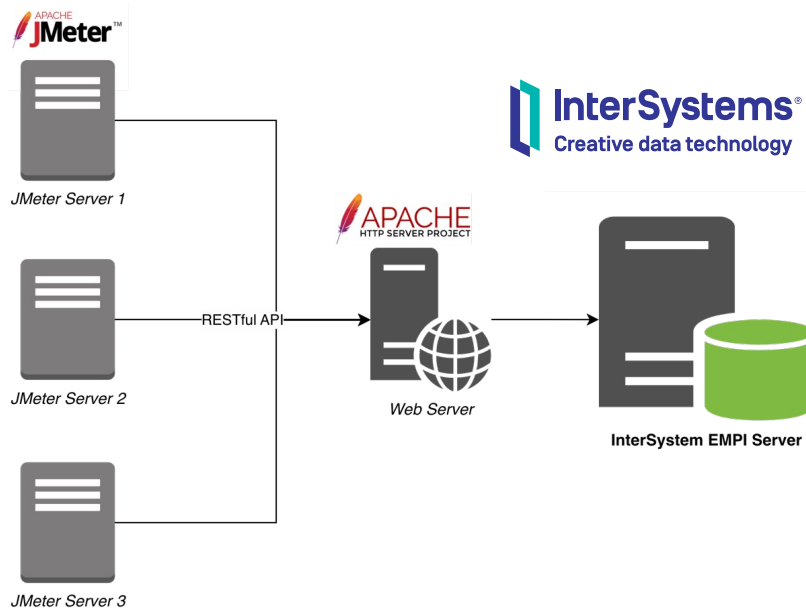


Figure 1: High-Level Diagram of Ingestion and Search Workloads

Bulk Ingestion

Synthea was used to generate over 300 million synthetic persons to be loaded into InterSystems EMPI. Multiple records were generated and distributed to each of the three JMeter driver servers used for the test. Each JMeter driver workflow used an **AddPerson(...)** RESTful API endpoint on the InterSystems EMPI instance. The ingestion process is more than just inserting records into a database. During the ingestion process, the identify matching engine performs operations such as linking or potential linking, marking records for review validation, as well as identifying candidates for referential matching.

Person Search

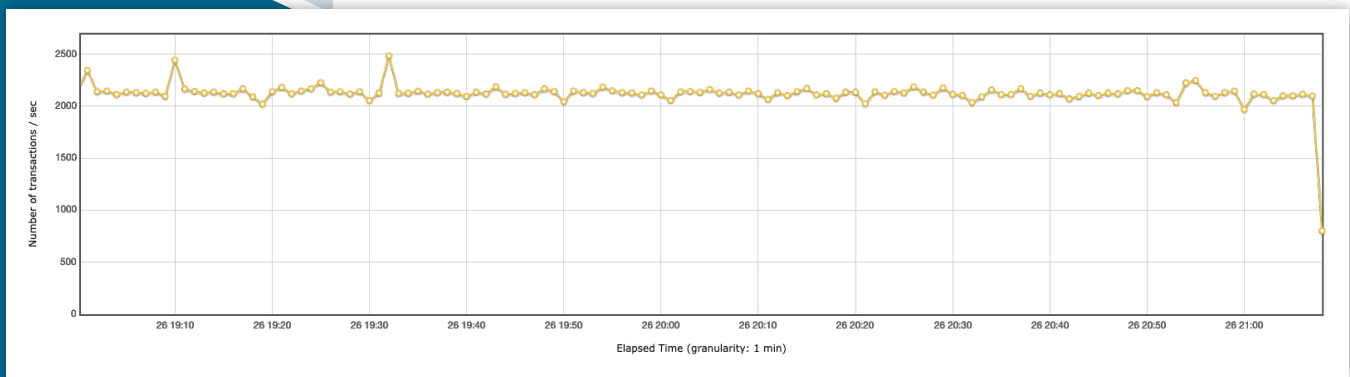
InterSystems EMPI leverages deterministic and probabilistic algorithms to match records. These algorithms have different workflows and complexity. The **PersonSearch(...)** benchmark workload used a 50/50 split ratio between the two distinct types of search algorithms to demonstrate a more realistic mixed workload.

Benchmark Results

Bulk Ingestion Workload

As previously described, three JMeter driver servers were used to simulate a high rate of add/update persons to InterSystems EMPI. Each workload driver had more than 100 million person records to process, for an aggregate of 308 million person records loaded into InterSystems EMPI. The sustained aggregate average throughput was ~6400 TPS (~2200 per driver) with an average latency of ~60ms with a 100% success rate. The graphs of Figure 2 below, provide an example of the sustained TPS and predictable latency from one of the workload drivers.

TPS



Latency

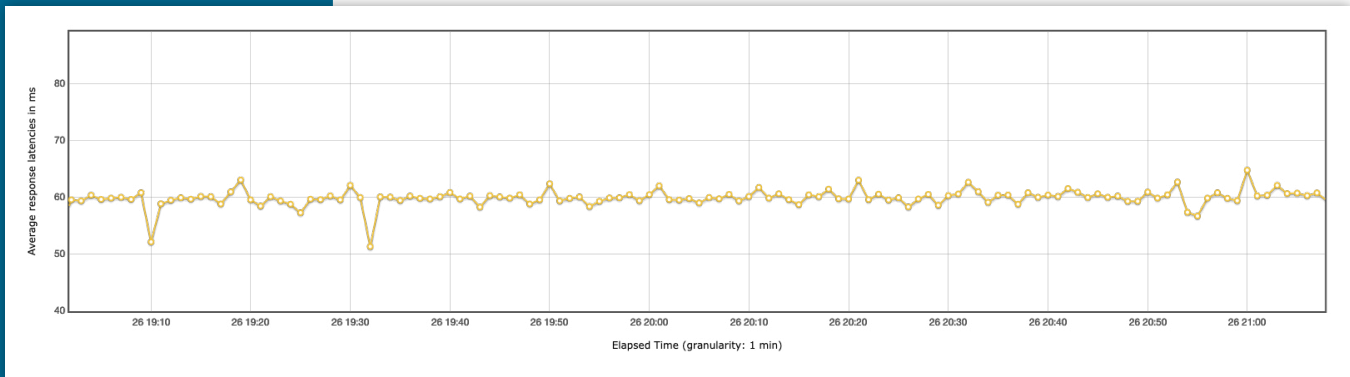


Figure 2: TPS and Latency per Workload Driver Over Time

During ingestion, server CPU utilization was efficient and consistent, allowing for linear scaling of ingestion TPS rates. Figure 3 illustrates the CPU utilization over time showing the expected flat high rate of utilization for a bulk ingestion workload.

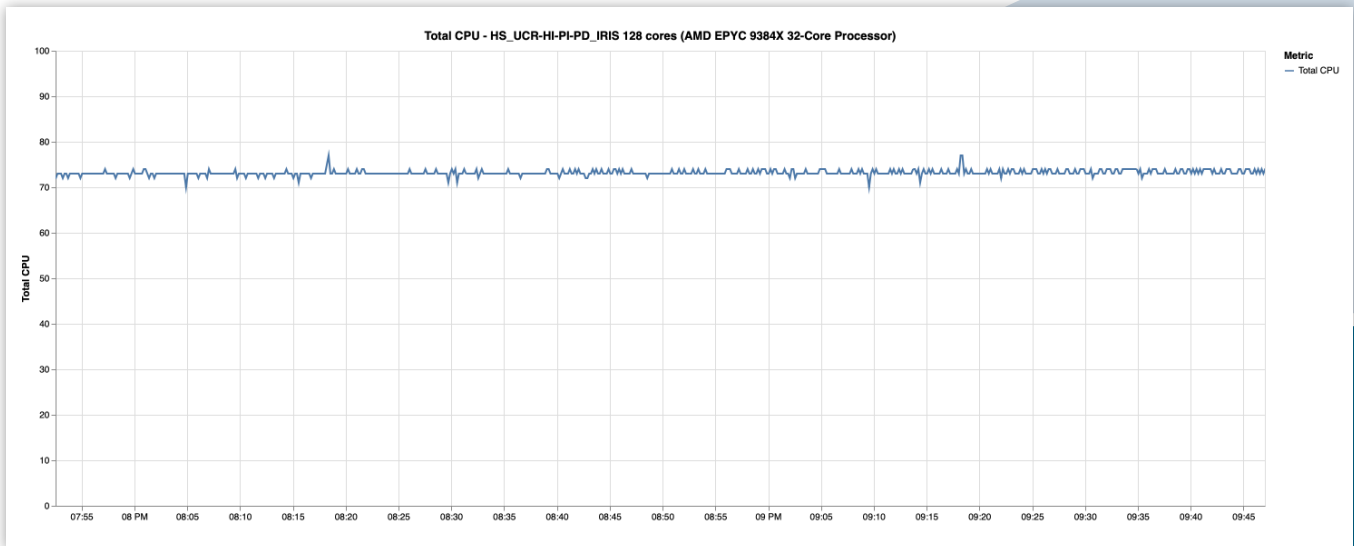


Figure 3: InterSystems EMPI Server CPU Utilization - Ingestion

Person Search Workload

The same three workload drivers were used to simulate a high rate of Person Search transactions. Both deterministic and probabilistic algorithms were used in a 50/50 split ratio to provide a more realistic use case. The ratio varies from customer to customer depending on use cases, so a 50/50 split was deemed a fair representation for this benchmark.

For this Person Search test, a long runtime was used to show the performance over time demonstrating that a short up-front burst of activity was not simply pulling from database cache but actually sustaining extremely low latency while doing a high rate of database physical disk reads.

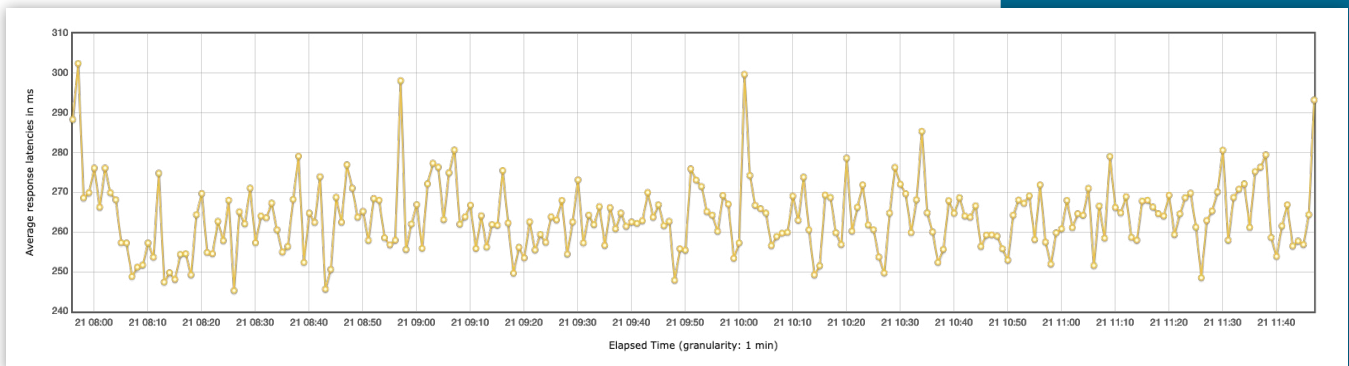


Figure 4: Round Trip Latency - Person Search

As the graph shows above in Figure 4, the average round trip latency is ~263ms throughout the test run for a duration of nearly 4 hours. The following graph in Figure 5 shows the physical disk read rates sustaining an average ~80,000 disk reads per second also for the several hours of the test run. It is important to show this metric over time to show the high TPS rate and round-trip low latency are not from a short period of only pre-cached data in database cache.

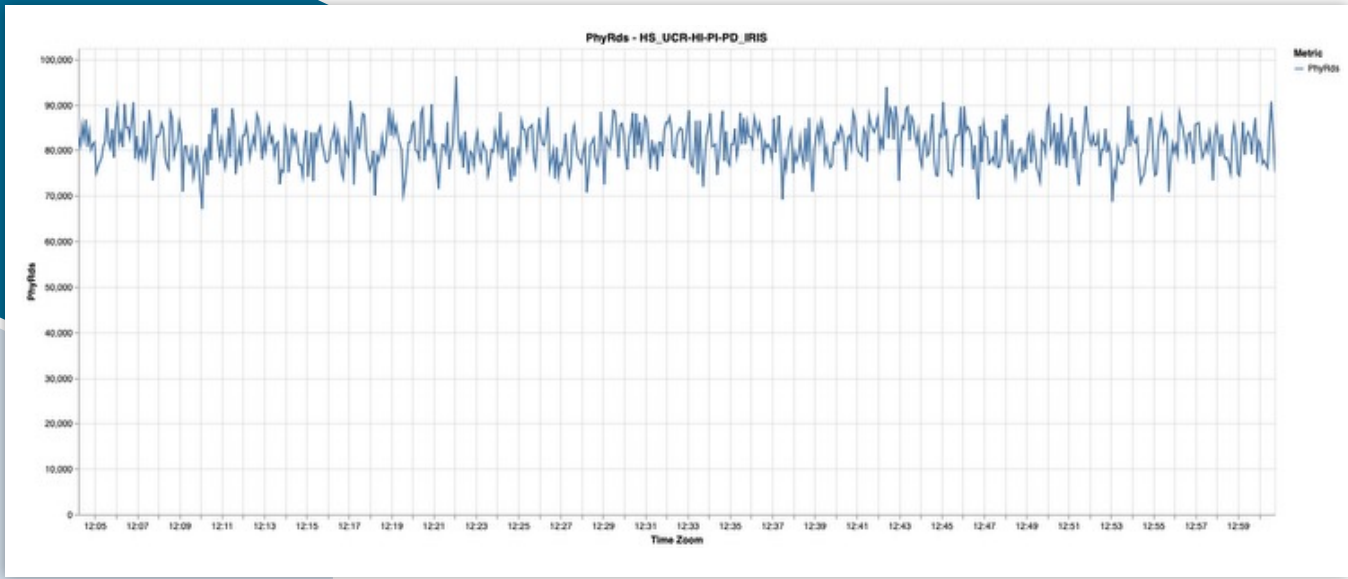


Figure 5: Physical Disk Reads - Person Search

In addition to a realistic physical disk read rate, the actual number of database accesses per second, known as “Glorefs,” is also a consistent sustained rate of over 8 million database accesses per second.

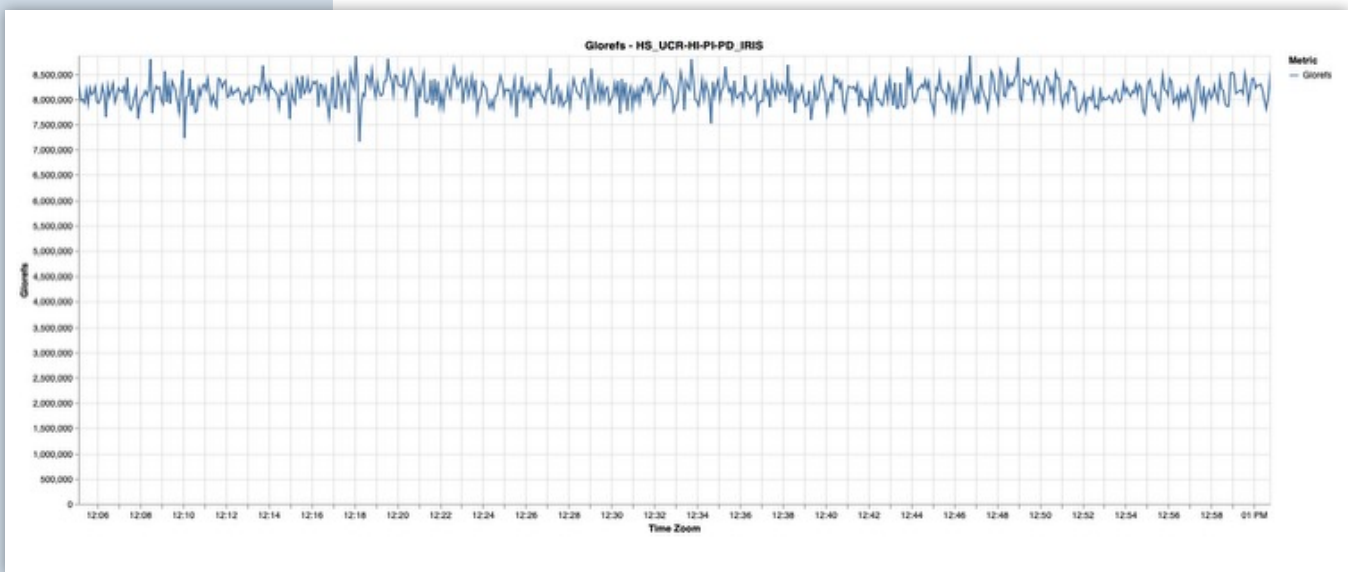


Figure 6: Database Accesses per Second - Person Search

Hardware Used

Server Configuration

The InterSystems EMPI test utilized a server with 2 x AMD EPYC 9384X 32-Core Processors providing a total of 64 cores / 128 threads at a base clock speed of 3.1GHz, 1.5TB RAM, and 10GbE network interface. This test used an Ubuntu 24.04.1 LTS operating system.

Disk Configuration

The server leveraged 6 x 7.6TB locally attached NVMe SSD with isolation of the InterSystems IRIS Data Platform database and transaction journal volumes onto separate disk devices following InterSystems best practices. After ingestion workload completed all 308M persons, the database was ~622GB in size.

Summary

The following table provides summarizes the throughput and latency of the workloads:

Workload	# Driver Threads	Data Volume	JMeter Driver Transactions per Second (TPS)	Average Latency (ms)
Ingestion	384	308-million-person bulk load	6400 TPS	60 ms
Search	192	44,000 searches per minute	732 TPS	263 ms

Conclusion

This benchmarking report demonstrates that InterSystems EMPI delivers exceptional scalability and performance for enterprise-level identity resolution. With sustained bulk ingestion rates exceeding 6,400 transactions per second and the ability to load 308 million records in under 14 hours, the solution proves its readiness for national-scale deployments. Person search workloads maintained an average latency of 263 ms, even under prolonged high-volume conditions, validating the platform's ability to provide rapid, accurate identity matching across massive datasets.

These results confirm that InterSystems EMPI is engineered for speed, reliability, and efficiency—empowering healthcare organizations to unify patient data and accelerate initiatives such as clinical trials and population health management at scale.

Learn more about [InterSystems EMPI](#).

