# IBM and InterSystems demonstrate impressive performance results

*POWER8 and Caché 2015.1 achieved high scalability during a recent Epic benchmark*

## Contents



## Introduction

### Purpose of this paper

This paper describes the methods and results obtained when testing an electronic medical recordkeeping (EMR) system based on technology from Epic and InterSystems. Epic provides software that is widely used by the general healthcare community, such as hospitals, clinics, medical research facilities and other organizations that are responsible for providing medical care to patients. Epic software uses InterSystems Caché to manage the large repository of patient-related data accessed by the application.

Caché is an advanced database management system and rapid application development environment that offers lightning-fast performance, massive scalability, and robust reliability. It is a new generation of database technology providing multiple modes of data access. Data is only described once in a single integrated data dictionary and is instantly available using object access, high-performance SQL, and powerful multidimensional access—all of which can simultaneously access the same data. Caché comes with several built-in scripting languages, and is compatible with the most popular development tools.

### Purpose of the Epic/Caché testing

The purpose of the testing is twofold: customer sizing and scalability.

#### Sizing

Epic uses the information from the testing to provide customized hardware sizing information for each of their customers. This allows the company to give each customer recommendations for requirements such as number of cores, amount of memory, and quantity of disk storage. The size of hospitals and other health-related facilities can vary significantly. Therefore, the hardware requirements can vary as well.

#### Scalability

Scalability results provide information about a server's upper transaction limits while still meeting response time requirements. The scalability limits can be due to multiple factors such as core count, CPU clock speed, inter-core communication and storage response times. A component of the testing process allows test engineers to observe and analyze these limits and identify ways to scale higher. As healthcare organizations grow rapidly, scalability becomes important in accommodating increased user loads.

### The testing process

The testing process requires multiple components, which will be described later in this paper. Aside from the extensive hardware requirements, engineers from Epic, InterSystems and IBM typically gather at the same location to exchange information and ideas for improving scaling, as well as addressing other issues during the testing.

### What this paper covers

This paper will cover the overall testing process and methodology by describing the conceptual and physical layouts for the test environment; the test methodologies; and the data that is collected for analysis.

Results from the latest round of testing will be presented and compared to prior tests to demonstrate the improvements made to the Epic application, InterSystems Caché and the IBM® POWER8® hardware, which have contributed to the higher scalability achieved.

## The Caché environment

Caché is an SQL/NoSQL database product developed by InterSystems. Caché maintains as many active pages as possible in buffers. The buffers consist of a large shared memory region in RAM containing a copy of the most actively used data.

### Caché terminology

This section will provide a brief description of terminology associated with the test environment.

#### GREF

A global reference (GREF) is a measure of work being done by the Caché database engine. As requests for data reads and updates increase, GREFs per second (GREFs/s) will also increase. GREFs/s is used as a metric to determine the level of throughput per unit of time completed in environment being tested.

#### Symmetric multiprocessing (SMP)

The Caché SMP architecture consists of a single Caché DB server that serves two roles: it runs the Caché DB engine, and it hosts multiple users. With SMP, Caché allows organizations to quickly scale up by adding new CPU cores to their systems. While the scalability that can be achieved using SMP is lower than that offered by ECP, it is still sufficient for the majority of customers.

### Enterprise Cache Protocol (ECP)

The Caché ECP architecture offers a way to share data, and access locks and executable code among multiple Caché systems. Data and code are stored remotely, but cached locally to provide efficient access with minimal network traffic. This arrangement allows Caché customers to quickly scale out by adding new application servers, achieving levels of scalability far higher than would be possible with SMP.

## Epic test methodology

The Epic test methodology is designed to provide both sizing and scalability information for a specific hardware component, such as an IBM Power® server. Test results are used to determine the number of compute cores and memory size that a specific customer will require. A small customer may only need 16 cores, whereas a large customer may need 32 cores or more. Typically, the memory and storage requirements will also increase as the core requirements increase.

Information about scaling is also gathered, to understand the behavior of the systems at the high end. This information is used to ensure that a customer will not exceed the capabilities of a given system when it has its full complement of cores, memory, communication adapters, and storage components.

### The Epic test environment

The principal components needed to test any system are user load and a realistic patient database. It is obviously not practical to engage several thousand concurrent users to generate sufficient load to apply any meaningful stress to the system under test. Instead, a simulation that represents multiple diverse end-user workflows is generated using several "test engine drivers." These drivers are pre-loaded with workflows, which represent a set of different roles within a medical organization.

### General hardware layout and requirements

The test engine drivers run a set of predefined scripts that generate workflows. Test engineers are able to control the quantity of workflows at any given time, thus allowing for data collection and analysis at a specific combined user load.

### SMP

The SMP hardware layout will consist of the test engine drivers, which are directly connected via an Ethernet-based LAN connection to the server, in this case the POWER8 E870 system. The E870 is an 80-core SMP system with 2 TB of memory. The software running on the E870 handles new user logins as well and ongoing requests from all users.

### ECP

The ECP environment introduces an application server layer. The application server manages the user logins and will host the actual user session. The application server sends requests from multiple users to the main Caché DB engine, running on the POWER8 E870. Data that is returned to the application servers is cached in the application server's memory, so that the data can be reused. This reduces the work that needs to be done by Caché on the main DB server by eliminating the management of user sessions and limiting the number of application servers submitting requests for data, which results in better efficiencies within the main DB engine. In addition to the 80-core POWER8 E870, the test ECP environment included a total of 150 application server cores.

SMP testing and ECP testing were conducted using IBM DS8000® and IBM FlashSystem® storage technology. The DS8000 and FlashSystem storage systems were configured such that response times from the storage were well within acceptable ranges. This greatly reduced the possibility of the storage component becoming a bottleneck for response time.

### The testing process

The testing process consists of a "ramp-up" of workload. The ramp-up is increased until a certain number of simulated users is attained. The increase in users is suspended in order to collect a "scenario." The scenario allows engineers to gather GREF information, as well as response time information at fixed load level. Once this data is collected (typically over a 25 minute interval), the ramp-up continues until a new load level is reached. The process is then repeated.

### The testing criteria

Once the load has reached a point where the response time exceeds an established criterion, the ramp-up process is suspended. Additional data is collected and the system is analyzed to determine what is causing the increased response times. There are multiple factors that can trigger increased response times. Aside from simple resource exhaustion (maximum CPU utilization), such things as high lock contention, storage related latencies, and communication latencies (in ECP mode) can cause higher response times. Again, the criterion for scalability limits is response time as load (GREFs) is increased.

### Data collected

Data is collected at multiple levels.

IBM collects system data utilizing several tools. These include system traces, network monitors, application code profiling, hardware counter data, memory utilization, and CPU utilization statistics. The data assists IBM in identifying what system or OS resources are being over utilized, or are not performing as expected. The data is also passed to engineers from InterSystems to help them identify "hot spots" within their code.

InterSystems collects detailed information about how data is being managed and accessed by the Caché kernel code. This data provides information about specific data structures that are being accessed by multiple users, resulting in processes waiting excessively for access to data. Other data, including spin counts while waiting for a resource and how frequently a process will sleep while waiting for a resource, is also collected. A thread that sleeps while waiting for a resource requires a context switch, which is a time-consuming activity.

Epic examines how various workflows are being completed, and whether there are code segments that handle those workflow requests that could be improved in order to reduce lock contention between concurrent users.

## Test results

For the latest testing, significant improvements were made by all three companies to their respective software and (in IBM's case) hardware and OS products. These changes led to improvements in scalability compared to past tests.

This section presents results of tests from both the latest testing as well as previous tests. Previous tests used older versions of both hardware and software.

The prior and current test configurations consisted of:

| | | |
|---|---|---|
| **IBM** | Prior: | POWER7+™ P780 32 cores, AIX® 7.1 |
| | New: | POWER8 E870 56 cores, AIX 7.1 |
| **InterSystems** | Prior: | Caché 2013.1 |
| | New: | Caché 2015.1 |
| **Epic** | Prior: | Epic 2014 |
| | New: | Epic 2014 |

### Comparisons of SMP results between Caché 2013.1 and 2015.1

Testing showed that while using an SMP architecture, InterSystems Caché 2015.1 running on POWER8 and AIX 7.1 was able to scale to about 8.6 million GREFs, achieving a level of scalability that was more than double that offered by Caché 2013.1 running on IBM POWER7+.

The test results were impressive for two reasons: the platform was able to double transaction volume, while keeping response times static. This combination of near-linear scalability and predictable response times gives healthcare organizations the ability to handle growing numbers of users, with no perceived degradation in performance.

Caché 2015.1 running on POWER8 enabled significantly greater scalability than Caché 2013.1 running on POWER7+, while also maintaining excellent response times. In end user facing applications, these response times are key. For instance, in a healthcare setting, an excellent response time means that physicians would be able to pull up records with no delays, allowing them to quickly answer patient questions and create an informed treatment plan.

### Comparisons of ECP results between Caché 2013.1 and 2015.1

From an ECP standpoint, the combination of the POWER8 processor technology and the Caché 2015.1 code drove scalability up to 14.5 million GREFs. This represents a level of scalability that is more than double that offered by Caché 2013.1 running on POWER7+, while still maintaining equivalent end-user response times.

By placing the end-user sessions on the application servers, the ECP architecture enables the main Caché database server to manage transactions more efficiently. As a result, even small improvements made to the Caché data server—through updates to the hardware or the software—will translate to large overall improvements in the performance offered by the multiple application servers.

## Improvements from POWER7+ to POWER8

### Improvements to bus fabric

Changes were made to the bus fabric to manage lock contention timing.

### Increased bandwidth between cores, sockets and CECs

More connections between central electronics complexes (CECs) were added to increase data throughput between cores, thereby reducing cache-to-cache coherence latencies.

### Lwarx and stwcx instructions and timing fix

A bug that caused the timing between load word and reserve indexed (lwarx) and store word conditional indexed (stwxc) reservations to reduce the ability for stores to complete as often as they could was corrected.

### Changing spin priority instructions

In the POWER8 instruction set, pseudo instructions were introduced that allowed applications to change the instruction priorities for segments of machine code. By reducing the priority of instructions which were used to spin on a lock, SMT threads sharing the same CPU were able to complete "real" work more efficiently. This allowed processes holding a lock to get out of the way in less time and allow processes waiting for the same resource to be dispatched sooner.

### Semaphore locking: 64 locks per set

A tunable parameter which allows up to 64 locks per semaphore set was introduced to the most recent version of the AIX 7.1 OS. Prior versions were limited to one lock per semaphore set. This caused delays in semaphore operations, since thousands of semaphores were associated within a semaphore set.

### Increase processor cache

Increased L2 and L3 caches, as well as the introduction of an L4 cache in POWER8 allows for more data to be available to the processor. This helps reduce the likelihood of cache misses when lock data is being propagated to multiple CPUs at a rapid rate.

## Improvements to Caché

For the Caché 2015.1 release, InterSystems engaged in a large optimization effort to provide more massive scalability than ever before. The scalability improvements come both in vertical scaling, by more efficiently utilizing very large processor core counts in a single OS partition, and in horizontal scaling, with concurrency improvements within ECP, InterSystems' technology for distributing database accesses to multiple application servers. This effort helped support InterSystems' partners such as Epic with large-scale deployments.

The optimization effort included innovative algorithmic improvements in the core of the database engine to alleviate bottlenecks and contention points observed in very high-scale systems. By eliminating contention, more CPU cores can be utilized simultaneously for database accesses. Similarly, eliminating points of contention within ECP allows more CPU cores to be utilized per application server, and more linear scaling across more application servers.

InterSystems worked with IBM and Epic to validate and fine-tune the changes made to Caché. Caché 2015.1 demonstrated a 2x scalability improvement over Caché 2013.1 on a single system, with near-linear vertical scalability through 56 cores. In an ECP configuration, Caché 2015.1 also demonstrated greater than 2x scalability improvement over Caché 2013.1.

---

*"Caché version 2015.1 continues to provide groundbreaking performance improvements. Nearly doubling scalability, this version is a strong platform for our user organizations as they engage to meet the high throughput demands of complex analytics and aggressive growth in the era of volume-to-value transformation in healthcare."*

— Carl Dvorak, President, Epic

---

## Conclusions

Extensive collaboration between Epic, InterSystems and IBM has resulted in significant improvements in scalability and performance. Modifications in both the hardware and software layers has greatly reduced latencies which, in prior versions of the environment, limited the ability to generate the higher GREFs/s levels that are now attainable. All three companies continue to work together to further increase performance and scalability of the Epic EMR environment.

© Copyright IBM Corporation 2016

IBM Corporation
IBM Systems
Route 100
Somers, NY 10589

Produced in the United States of America
June 2016

IBM, the IBM logo, ibm.com, POWER8, Power, DS8000, POWER7+, FlashSystem, and AIX are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF

MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.


Please Recycle

POW03176-USEN-00