



Latest IBM Z Innovations – “Take a Look Inside the z14!”

IBM Competitive Project Office

Fehmina Merchant, Ph.D.

Acknowledgements: Special thanks to the CPO and Z teams for their valuable inputs.

July 2017

Contents

- Introduction 3
- Differentiated value at the core 3
- Optimized for traditional and new applications 5
- Designed for data protection..... 6
- Designed to deliver data insights with machine learning 7

Introduction

Over the years, companies have invested in the IBM Z for its highly optimized and secure platform for transaction processing and data serving. And today, IBM Z is at the center of the global economy, supporting well over 30 billion transactions a day, and managing more than 80% of the most valuable corporate data.¹ With the new IBM z14, we're raising the bar and delivering a core infrastructure that is even more powerful and secure, unlike any available enterprise system in the marketplace.

The newest IBM z14 is built for the digital economy to support a massive scale of transactions and data while providing unprecedented levels of security. In this paper, we take a closer look at what IBM z14 has to offer over its previous generation IBM z13, and how customers can be sure to increase their operational efficiency and maximize their bottom line by upgrading to this latest Z technology.

Differentiated value at the core

IBM z14 is designed to support massive capacity and performance while delivering on service level agreements for an exceptional digital customer experience. To deliver outstanding performance, the entire system up and down the stack - cores/processors, cache, memory, I/O – has been optimized as outlined in Figure 1.

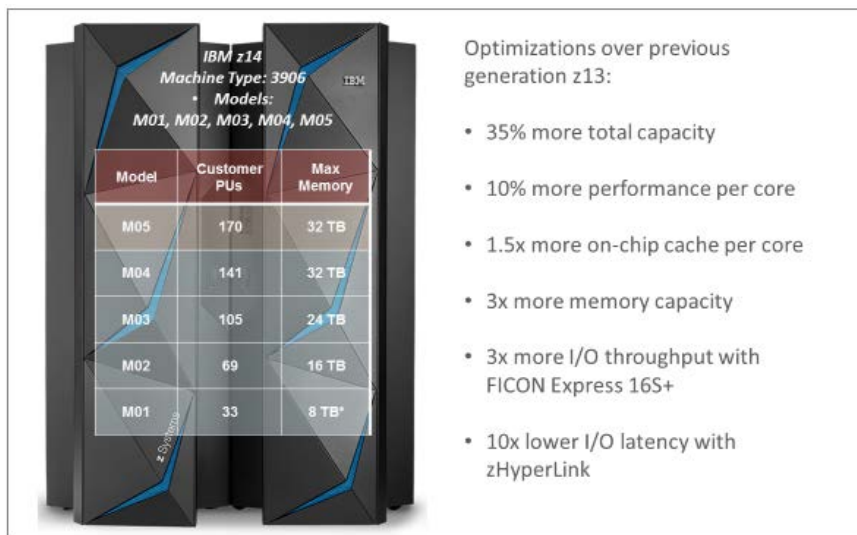


Figure 1. IBM z14 is optimized to deliver exceptional performance

A fully configured IBM z14 can have up to 170 cores which are expected to offer up to 35% more total capacity in a single footprint compared to the largest previous generation IBM z13.² It includes a new industry-leading, faster 5.2 GHz 10-core processor chip, that leverages the density and efficiency of 14nm silicon-on-insulator technology to deliver increased performance and capacity for a wide range of workloads. A single processor of z14 for equal n-way at common client configurations has

¹ Source: IBM.

² Based on preliminary internal measurements and projections and compared to the z13. Official performance data will be available upon announce. Results may vary by customer based on individual workload, configuration and software levels. Visit LSPR website for more details at: <https://www-304.ibm.com/servers/resourcelink/lib03060.nsf/pages/lsprindex>.

approximately 10% greater capacity than previous generation z13.³ IBM z14 makes use of these high-performing processors to offer five different models scaling from 33 to 170 total configurable cores.

The cache architecture on the new z14 has also been redesigned with 1.5x more on-chip cache per core compared to the predecessor IBM z13. Specifically, the z14 core has 96 KB for L1, a 33% increase. The L2 caches have been doubled to 4 MB each and the L3 cache has also been doubled to 128 MB in addition to improvements in latency for L4 cache. Bigger and faster caches, in conjunction with processor performance improvements, can help z14 deliver consistently fast response times for transactional workloads for an improved customer experience.

Each z14 can now have up to 32 TB which is three times the memory available compared to previous generation z13. The additional memory can help support larger local buffer pools, larger in-memory databases, as well as being able to efficiently process huge amounts of data for faster business insights.

IBM z14 is specifically built with I/O options to deliver speed, scale, and reduced latency in getting to data which is critical to achieving exceptional transaction throughput. The new FICON Express16S+ channels are designed to boost I/O rates and reduce single stream latency. Compared to the previous generation FICON Express16S on z13, the new FICON Express16S+ on z14 can deliver three times more IOs/sec and can drive 25% more throughput for reads and writes.⁴

For customers with existing FICON SAN infrastructure, IBM zHyperLink Express introduces a new I/O paradigm on z14. zHyperLink Express is a direct connect short distance I/O adapter which offers extremely low latency connectivity to FICON storage systems. It can improve response times for I/O sensitive workloads by as much as 50% without requiring application changes.⁵ zHyperLink has been designed to deliver up to 10x lower read latency compared to High Performance FICON.⁶

Like the previous generations, z14 can also be split into up to 85 logical partitions (LPARs). These partitions can allow for workloads to be run in highly secure isolation on a single box, obviating the need for having separate physical systems as would be the case on other competitive platforms. It provides capabilities to dynamically scale with capacity on demand for non-disruptively adding resources to support varying workload demands while controlling costs. Further, with superior workload management capabilities, z14 can enable resources to be effectively used by multiple workloads, thereby achieving the highest possible utilization of the platform unlike any other system out there in the marketplace.

³ Based on preliminary internal measurements and projections and compared to the z13. Official performance data will be available upon announce. Results may vary by customer based on individual workload, configuration and software levels. Visit LSPR website for more details at: <https://www-304.ibm.com/servers/resourcelink/lib03060.nsf/pages/lsprindex>

⁴ Results were obtained in a laboratory environment running an I/O driver program under z/OS utilizing the zHPF protocol and small data transfer operations, FICON Express16S+ operating at 16 Gbps, and FICON Express16S operating at 16 Gbps. The actual throughput or performance that any user will experience may vary.

⁵ The response time estimate is based on IBM internal measurements and projections that assume 75% or more of the workload response time is associated with read DASD I/O and the storage system random read cache hit ratio is above 80%. The actual performance that any user will experience may vary.

⁶ Projections of read I/O latency are based on I/O service times and CPU queueing delays from IBM internal measurements (z14 and zHyperLink results with DS8886 and z13 measurements). The actual performance that any user will experience may vary.

For years, IBM Z has been designed to operate with the highest levels of reliability and availability, and the new IBM z14 is no different.⁷ At the hardware level, it ships with redundant cores, which are automatically activated in the rare event of hardware failure to guarantee continuous availability. Further, redundancy in the memory subsystem is built-in using the highly robust RAIM technology that can provide protection from failures at DRAM, DIMM and memory channel level. IBM Z's memory subsystem sets the gold standard for being the most resilient solution in the industry.

Optimized for traditional and new applications

On IBM z14, new instructions have been added to further boost the performance of various workloads. New instructions in the SIMD (single instruction, multiple data) facility can speed up traditional workloads that make use of decimal operations (i.e. COBOL 6.2, PL/I 5.2) and modern applications like analytics (i.e. Apache Spark for z/OS). The compression co-processor in each core has also been improved to use fewer CPU cycles for compression and expansion. DB2 plans to take advantage of the compression co-processor to support index compression and enhanced data compression. These compression features can enable further improvements in DB2 memory usage, data transfer, and storage efficiency.

IBM z14 also offers several enhancements to support Java which has become a language of choice for clients as they extend and modernize business logic to further drive CICS, IMS, and DB2 transactions. These enhancements include 50+ new instructions to support Java and improved throughput of running Java on z Integrated Information Processor (zIIP) and Linux IFL with next generation simultaneous multithreading (SMT). It also encompasses enhancements to the Crypto Express6S, exploitation of the new Galois Counter Mode (GCM) encryption, and exploitation of SIMD instruction enhancements that can help improve performance for Java-based applications.

Additionally, z14 introduces Guarded Storage Facility for Java-based applications which can help reduce program pauses during Java garbage collection. With this new pause-less feature, clients running z/OS Java applications can expect to achieve as much as three times better throughput for response-time constrained service level agreements (SLAs) using the new IBM Java 8 SR5 pause-less garbage collection feature on z14 compared to using Java 8 SR3 on z13.⁸

⁷ <http://itic-corp.com/blog/2017/06/ibm-lenovo-servers-deliver-top-reliability-cisco-ucs-hpe-integrity-gain/>

⁸ Performance results based on IBM internal tests running Java Store Inventory and Point-of-Sale in COMPOSITE mode application benchmark on stand-alone dedicated IBM z14 and z13 machines using z/OS 2.2.2 with APAR OA51643 and no other workloads running in the LPAR under test. Both z14 and z13 were configured with 1CP and 8 SMT zIIPs with total 17 hardware threads. Capacity projections were done to estimate the benefit of moving from z13 z/OS 2.2 Java 8 SR3 to z14 z/OS 2.2 Java 8 SR5 with Pause-less garbage collection enabled by java option -Xgc:concurrentScavenge. The response-time constrained Service Level Agreements (SLAs) metric used for this claim was based on geometric mean of (throughput @ 10ms, 25ms, 50ms, 75ms and 100ms response time SLAs). Hardware instrumentation data was collected and analyzed on all benchmarks to verify performance results.

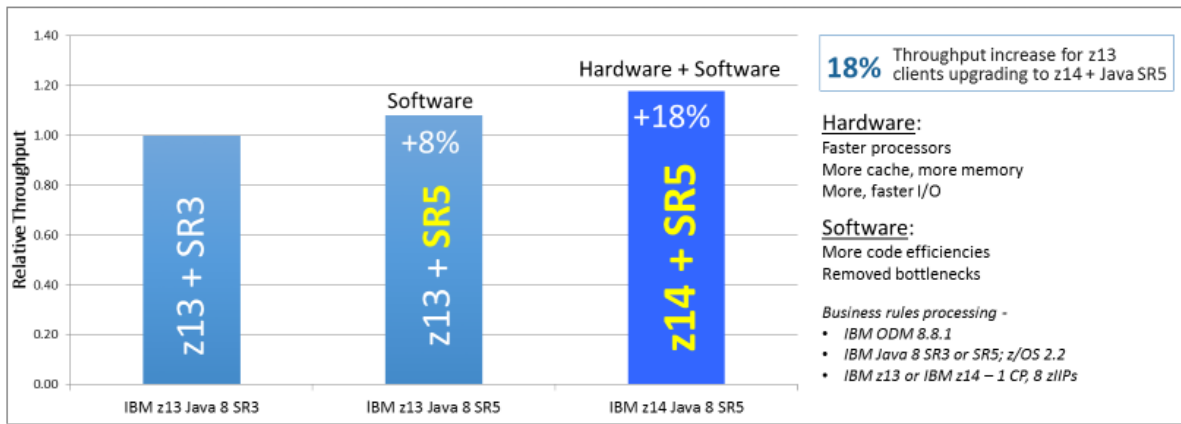


Figure 2. Achieve exceptional performance for Java-based applications on IBM z14

With these synergistic enhancements on hardware and software, z/OS Java-based applications when upgraded to the latest IBM z14 hardware and software stack can result in much as 18% throughput increase over the previous generation IBM z13 as shown in Figure 2.⁹

Designed for data protection

For ultimate protection for core corporate data and simplified compliance, IBM z14 features Pervasive Encryption where 100% encryption is the new standard, with no need for application changes, no downtime, and no impact to SLAs.

IBM Z provides a Pervasive Encryption capability for data both at rest and in flight. This capability begins with its hardware and follows through to the moment the data leaves the system as outlined in Figure 3.

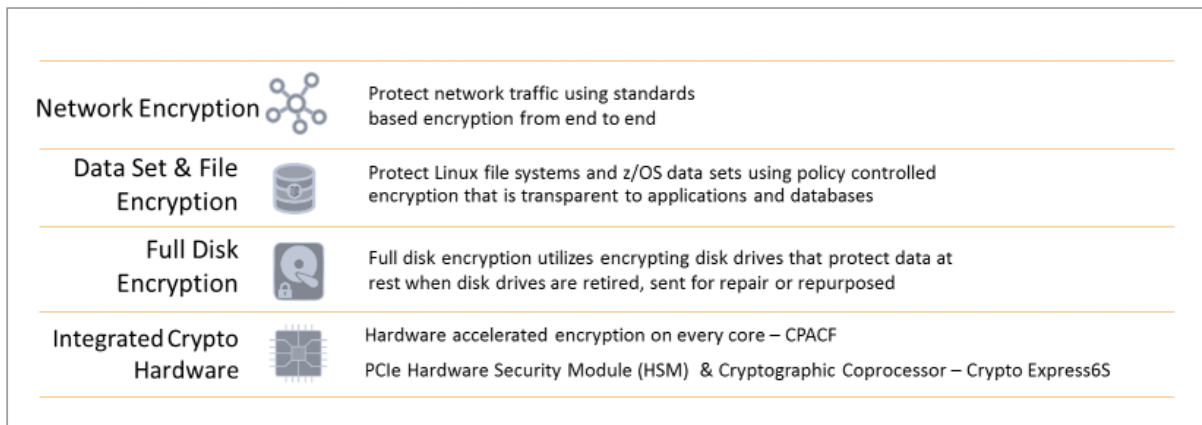


Figure 3. Achieve ultimate protection for data with Pervasive Encryption on IBM z14

Integrated Crypto Hardware: Each processor in IBM Z has a coprocessor called a CPACF (Central Processor Assist for Cryptographic Functions). It performs a set of full symmetric cryptographic and hashing functions. CPACF on z14 can provide seven times more throughput than on its predecessor z13

⁹ Performance results based on IBM internal tests.

and more than two times better throughput compared to the x86 platform.¹⁰ IBM also offers a Hardware Security Module (HSM) called Crypto Express. This card is certified to FIPS 140-2 level 4 – its highest level. It provides cryptographic operations above and beyond CPACF as well as key storage.

Full disk encryption: IBM Z can serve keys from the HSM to unlock disks in DS8000 storage systems.

Dataset and file encryption: IBM Z can encrypt datasets as a matter of policy, requiring no further attention from administrators or developers.

Network encryption: Traffic between nodes and the coordinating coupling facility is encrypted. Also, a readiness tool, zERT, is provided to enable administrators to determine if network traffic meets specified policy.

Because an auditor can independently verify the IBM Z data and infrastructure is protected and encrypted according to established enterprise security policy, any breach can be determined to be non-reportable in near real time. Comparing to what is available in the marketplace on commodity servers, pervasive encryption capabilities on IBM z14 are easier to enable, easier to manage, more secure, and more likely to achieve an environment capable of near zero reportable data breaches.¹¹ To achieve a similar protection without using this new Pervasive Encryption would require considerable additional labor by the customer.

Designed to deliver data insights with machine learning

Machine learning is one of the applications that can take advantage of the huge treasure trove of high-value customer and transactional data that sits on IBM Z today and deliver insights like never before.

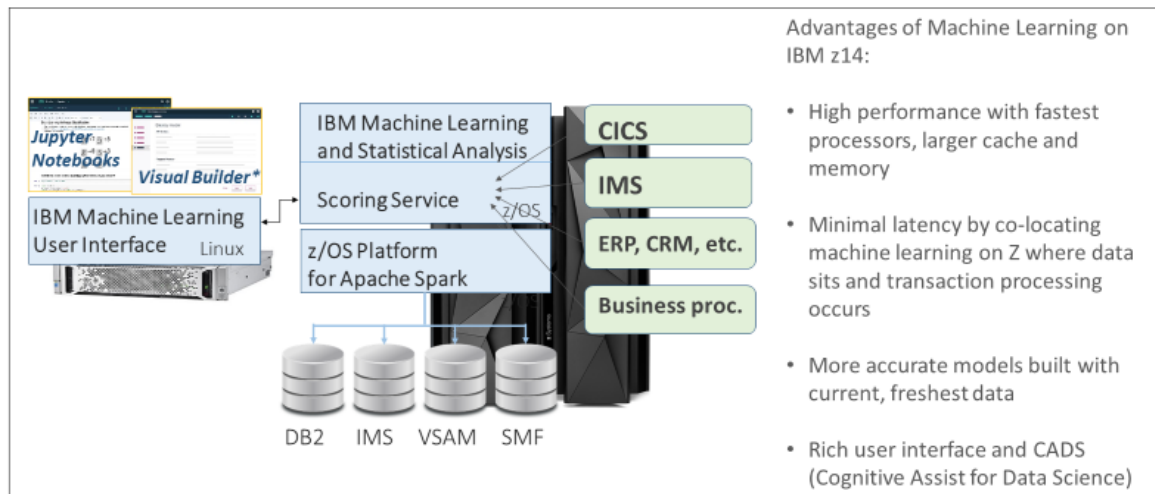


Figure 4. Deliver timely data insights with IBM Machine Learning for z/OS on IBM z14

¹⁰ Performance results based on IBM internal tests. Tests performed on – IBM z13 and z14 running RHEL with 8 IFLs w/SMT; x86 running RHEL with 8 cores w/SMT, E5-2667 v4 @ 3.20GHz.

¹¹ IBM internal study.

Because the z14 runs the fastest processors, has more cache and supports huge amounts of memory, it is particularly well suited for workloads like machine learning. And by co-locating machine learning on IBM Z where the transaction processing occurs (e.g., CICS, IMS), the network is eliminated, the calls go at memory-speed, and the latency is minimized.

Likewise, if you're moving data around to perform analytics, you're potentially limiting its business impact. But by running machine learning on IBM Z where the data originates, you can create more accurate models while eliminating unnecessary data movement costs and security risks. With IBM Z you gain real-time insights during transactions to best understand customers and grow your business.

IBM Machine Learning for z/OS (MLz) runs on top of the IBM z/OS Platform for Apache Spark, which includes a unique data layer optimized for ingesting all mainframe system of record data (DB2, IMS, VSAM, SMF, etc.). MLz enhances the existing SparkML libraries, adds authentication services, and includes model management and scoring services. Because MLz and Spark are Java based, the workload is almost completely offloadable to zIIPs.¹²

The rich and sophisticated MLz user interface includes both Jupyter Notebook, a favorite programming tool of advanced data scientists, and a Visual Model Builder, designed with a wizard-driven interface to walk non-data scientists through model development. Integrated into both is CADS (Cognitive Assist for Data Science), a unique and leading-edge technology from IBM Research that optimizes model building by automatically analyzing the data and making recommendations for best algorithms. The Hyper Parameter Optimization (HPO) feature helps model builders select the best parameters to use with the models.

IBM Z customers who are considering an investment in machine learning should look no further than IBM Machine Learning for z/OS, as only IBM MLz is designed specifically to take advantage of mainframe attributes, and only IBM MLz is optimized for z/OS data.

Closing remarks

IBM z14 delivers an optimal and secure core infrastructure that today's businesses demand for a trusted digital experience. With the new z14, IBM Z customers can be sure to increase their operational efficiency and maximize their bottom line like never before.

© Copyright IBM Corporation 2017

IBM Corporation
Route 100
Somers, NY 10589
USA

¹² zIIPs can also be run in SMT mode, effectively boosting performance by as much as 25%. Based on preliminary internal measurements and projections and compared to the z13.

Produced in the United States
June 2017
All Rights Reserved.

IBM, the IBM logo, IBM Z, z14, and Crypto Express are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both.

Other company, product or service names might be trademarks or service marks of others.

The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this documentation, it is provided "as is" without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

No other publication, distribution or use is permitted without the prior written consent of IBM. Customers who want a "deep drill down" on CPO Competitive Case Studies should be directed to the IBM Competitive Project Office under NDA. An NDA is required to explain CPO's methodologies, processes and competitive comparison. You can contact IBM Competitive Project Office by sending an email to ibmcpo@us.ibm.com

References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way.

IBM Competitive Project Office (CPO) Case Study results are based on IBM internal studies of IBM brand name solutions designed to replicate a typical IBM customer workload usage in the marketplace. The results were obtained under laboratory conditions, and not in an actual customer environment. IBM's internal workload studies are not benchmark applications, or industry standards, nor are they based on any benchmark standard.