The IBM z14 Performance Brief

7 November 2017

David Hutton

STSM, IBM Z Performance and Design

hutton@us.ibm.com

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a more complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*BladeCenter®, CICS®, DataPower®, DB2®, e business(logo)®, ESCON, eServer, FICON®, IBM®, IBM (logo)®, IMS, MVS, OS/390®, POWER6®, POWER6+, POWER7®, Power Architecture®, PowerVM®, PureFlex, PureSystems, S/390®, ServerProven®, Sysplex Timer®, System p®, System p5, System x®, z Systems®, System z9®, System z10®, WebSphere®, X-Architecture®, z13™, z13s ™ , z Systems™, z9®, z10, z/Architecture®, z/OS®, z/VM®, z/VSE®, zEnterprise®, zSeries®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

*** All other products may be trademarks or registered trademarks of their respective companies.**

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured Sync new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained Sync the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

z14 Performance: Design Highlights – Primary Performance Drivers

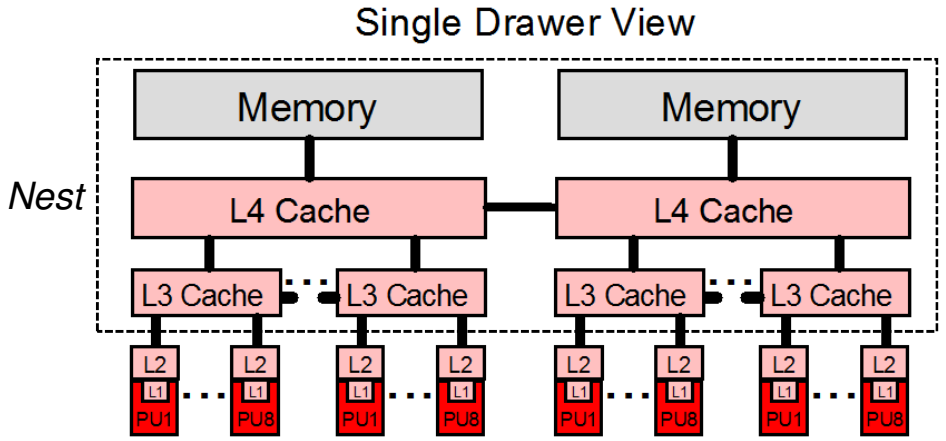
- Processor
  - Microarchitectural enhancements (branch prediction, resource turnaround, prefetching)
    - Improved IPC <-> Reduced CPI
  - Second generation SMT for zIIPs, IFLs and, new to z14, SAPs (always enabled)
    - Introduced on z13, improved on z14 – better sharing, multiple and faster translation engines
  - Second generation SIMD unit for analytics introduced on z13 – new decimal architecture introduced on z14
  - Up to 10 processor units per chip vs 8 on z13
  - Up to 170 configurable processor units per CEC on z14 vs 141 on z13
  - 4 different uni speeds (as always)
- Memory Subsystem
  - Continued focus on keeping data closer to the processor unit
    - Larger L1, L2, L3 caches. Individual L4s are larger but now shared by 2x L3s *so effectively smaller*
    - Improved IPC (Instructions Per Cycle) – aka Reduced CPI (Cycles per Instruction)
    - Second generation NUMA – reduced variability with single-node drawer
  - 3.2X configurable memory (8TBx4 drawers = **32TB on z14** vs 10TB on z13)
    - Primary use cases: supports large volume of Linux guests and in-memory analytics
    - New DB2 supports 16TB buffer pools however z/OS supports 4TB-max images; *give them a reason to grow!*
- PR/SM
  - 85 customer partitions (same as z13)
  - Improved algorithms
  - HiperDispatch
    - Exploits new chip configuration
    - Required for SMT on zIIPs

z14 Performance: Design Highlights – More Details

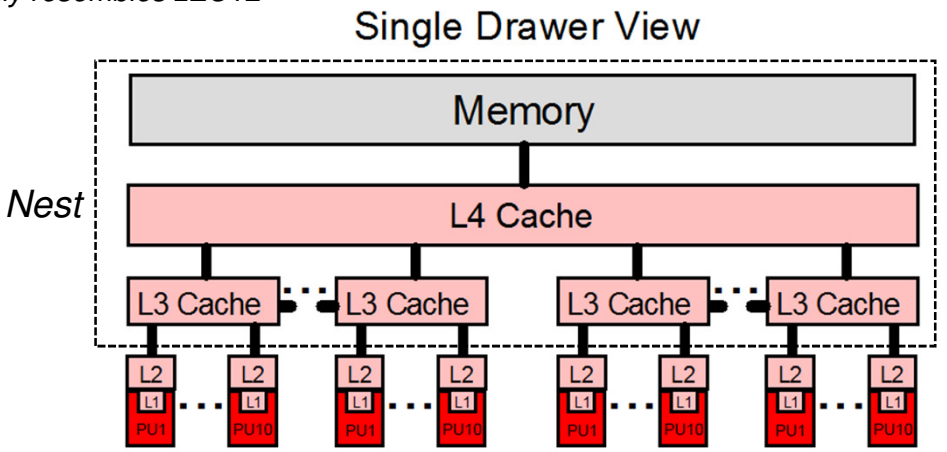
- **Processor uArch Improvements**
  - **5.2 GHz z14 vs 5.0 GHz z13 (+4%)**
  - Merged L1/TLB1 – eliminates TLB1 miss penalty, inlined TLB2 access on L1 miss mitigates TLB2 access penalty
  - Four HW-implemented translation engines on z14 vs one picocoded engine on z13
  - 2x CRSTE (combined region segment table entry) and 1.25x PTE (page table entry) growth
  - Branch prediction improvements; 33% BTB1-and-2 growth, new perceptron predictor and simple call-return stack
  - Pipeline optimization; improved instruction delivery, faster branch wakeup, reduced execution latency, improved OSC prediction, …
  - Optimized 2nd generation SMT; improved thread balancing, multiple outstanding translations, optimized hang avoidance mechanisms
  - Improved Hot Cache Line handling; dynamic throttling – e.g., “XI strong-arming” and a Hot Line Table
- **Cache Improvements**
  - L1 I-Cache increased from 96K to 128K per Core (1.33x)
  - L2 D-Cache increased from 2MB to 4MB per Core (2x)
  - L3 Cache increased from 64MB to 128MB per CP (2x)
  - L4 Cache changed from 480MB L4 + 224MB NIC to 672MB L4, *however only one L4 per drawer on z14 vs two on z13*
  - L4 sequential prefetch
- **Storage Hierarchy Improvements**
  - Single SC Topology (Reduced system latencies)
  - Redesigned System protocols (Reduce contention points)
  - New L3 fetch miss resumption on L4 sequential pre-fetch conflict
  - Bus Feeds and Speeds
  - Wider processor store bus
  - Improved Hot Cache Line handling; contention affinity, single SC
- **Software / Firmware Improvements**
  - PR/SM: Improved Memory Affinity, improved logical partition placement algorithms based on z13 experience
  - New Translation Management Facility (avoid expensive ops)
  - New per-work-unit dispatch cache footprint metrics (cache efficiency)
  - Improved Hot Cache Line handling (new instructions)

z14 vs z13 Hardware and Topology Comparison

- z13
  - CPU
    - 5.0 GHz
    - Major pipeline enhancements
    - 1 picocoded translation engine
  - Caches
    - L1 private 96k i, 128k d
    - L2 private 2 MB i, 2 MB d
    - L3 shared 64 MB / chip
    - L4 shared 480 MB / node
      - Plus 224 MB NIC



- z14 – *L3 clustering and cache sizes aside, topology strongly resembles zEC12*
  - CPU
    - 5.2 GHz
    - Logical directory w/ inclusive TLB
    - 4 HW-implemented translation engines
  - Caches
    - L1 private 128k i, 128k d
    - L2 private 2 MB i, 4 MB d
    - L3 shared 128 MB / chip
    - L4 shared 672 MB / ~~node~~ drawer



z14 Performance: Capacity Highlights

- Full speed single-thread capacity ratios relative to z13
  - **Average 1.10x z/OS at equal Nway**; *range 1.06 to 1.17 based on workload type and Nway*
  - **Average 1.31x max box capacity**; *range 1.26 to 1.35 based on workload type*
    - 170w z14 vs 141w z13
- Subcapacity models
  - Uniprocessor capacity ratio to full speed z14

model	z13 mult	z13 MIPS	**z14 mult**	**z14 MIPS**
401	0.15	250	**0.14**	**256**
501	0.44	746	**0.41**	**751**
601	0.63	1068	**0.59**	**1081**
701	1.00	1695	**1.00**	**1832**

  - Up to 33 CPs (general purpose processors) for each subcap model; z13 had up to 30 CPs
- SMT capacity option
  - Same as z13, customers can choose to run two HW threads per core on IFLs and zIIPs
    - Controlled by OS parm at the LPAR level
    - Added HW threads appear as additional processors to the OSes
  - New to z14, SMT will additionally be employed on SAPs – not user-controllable
  - Capacity improvements per IFL and zIIP core:
    - z14 ST -> z14 SMT: 10 .. 40%, average 25%
    - z13 SMT -> z14 SMT: ~15% z/VM Guest 2, ~10% others

LSPR: What's New For z14

- Workload updates
  - Upleveled software: z/OS 2.2, DB2 11, CICS 5.3, IMS 14, WAS 8.5.5.9, COBOL 6.1
  - Minor tweaks to three (LOW, AVG and HIGH RNI) hardware-characteristic-based workload categories
    - Based on CPU MF data from customers' zEC12 to z13 migrations
- HiperDispatch continues to be turned on for all measurements
- LSPR will continue to publish only single HW thread capacity in the multi-image table
  - Multi-image (MI) table
    - Median LPAR configuration for each model based on customer profile
      - Including effect of average number of ICFs and IFLs
    - Most representative for vast majority of customers
    - Basis for single-number metrics MIPS, MSUs and SRM constants
- zPCR continues to allow any configuration to be modeled
  - Customized LPAR configurations and workloads
  - Utilizes CPU MF (SMF 113) from the EDF file to determine LSPR Workload Match (LOW, AVERAGE, and HIGH) from L1MP and numeric-RNI
  - SMT capacity effect will be included via a user controlled "dial"
    - Set dial to reflect the estimated capacity increase of two threads over one thread
    - If no SMT history, pre-install guidance is to set the dial to the default 25% across the board; previously z13 had defaults of 20% for z/VM Guest 2 and 25% for zIIPs
    - Post-install guidance in setting dial from metering data available in RMF and the z/VM Performance Report

Variability Among Workloads on z14

- Performance variability is generally related to fast clock speed and physics
  - Increasing memory hierarchy latencies relative to microprocessor speed
  - Increasing sensitivity to frequency of "missing" each level of processor cache
  - Workload characteristics – RNI in particular – are determining factor, not application type
- zEC12 demarked the end of an era while z13 ushered in a new one
  - Substantial frequency gains from generation to generation are no more
  - Greater reliance on performance driven by improved IPC in core and nest (e.g., “uarch enhancements”) vs frequency
  - Workloads do not all react the same to these changes
  - Microbenchmarks are particularly susceptible to this effect
- Moving away from the MCM to a NUMA topology also created variability
  - Greater reliance on PR/SM to do the right thing in placing LPARs
  - Enabling HiperDispatch maximizes cache reuse potential and minimizes cache line migration traffic across the topology
- z14 is a second-generation z13
  - Mitigated some of the NUMA variability by largely removing the nodal boundary within a drawer
  - More cores per drawer and each with higher capacity than z13 means more work can “fit” on a single drawer
  - Added architecture and updated z/OS to more smartly manage locks
- **We expect migrations to z14 from z13 to be stable**
- **Workloads migrating to z14 from zEC12 and prior can expect to see similar albeit slightly less variability than the typical z13 experience**

z14 Performance: Summary

- Full speed single-thread capacity ratios relative to z13
  - **Average 1.10x z/OS at equal Nway**; *range 1.06 to 1.17 based on workload type and Nway*
  - **Average 1.31x max box capacity**; *range 1.26 to 1.35 based on workload type*
- Capacity improvements per IFL and zIIP core:
  - z14 ST -> z14 SMT: 10 .. 40%, average 25%
  - z13 SMT -> z14 SMT: ~15% z/VM Guest 2, ~10% others
- 3.2X configurable memory (8TBx4 drawers = **32TB on z14** vs 10TB on z13)
  - Primary use cases: supports large volume of Linux guests and in-memory analytics
  - New DB2 supports 16TB buffer pools however z/OS supports 4TB-max images; *give them a reason to grow!*
- Optimized 2nd generation SMT
  - Improved thread balancing, multiple outstanding translations, optimized hang avoidance mechanisms
- **Variability**
  - **We expect migrations to z14 from z13 to be stable**
  - **Workloads migrating to z14 from zEC12 and prior can expect to see similar albeit slightly less variability than the typical z13 experience**

Thanks for coming! hutton@us.ibm.com