# z/VM 7.1 Performance Update

Version 2019-08-15.1

*Brian K. Wade, Ph.D.*
*IBM z/VM Performance*
*bkw@us.ibm.com*

# Trademarks

**The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.**

BladeCenter*
DB2*
DS6000*
DS8000*
ECKD
FICON*
GDPS*
HiperSockets
HyperSwap
IBM LinuxONE Emperor
IBM LinuxONE Rockhopper
IBM Z*
IBM z13*
IBM z14
OMEGAMON*
Performance Toolkit for VM
Power*
PowerVM
PR/SM
RACF*
Storwize*
System Storage*
System x*
System z*
System z9*
System z10*
Tivoli*
zEnterprise*
z/OS*
zSecure
z/VM*
z Systems*

* Registered trademarks of IBM Corporation

## 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.
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.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.
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.
Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and
Linux is a registered trademark of Linus Torvalds 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.
OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.
TEALEAF is a registered trademark of Tealeaf, an IBM Company.
Windows Server and the Windows logo are trademarks of the Microsoft group of countries.
Worklight is a trademark or registered trademark of Worklight, an IBM Company.
UNIX is a registered trademark of The Open Group in the United States and other countries.

* Other product and service names might be trademarks of IBM or other 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 from 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 from 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.

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the “Authorized Use Table for IBM Machines” provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html (“AUT”). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

# Notice Regarding Specialty Engines (*i.e.*, zIIPs, zAAPs, and IFLs):

Any information contained in this document regarding Specialty Engines ("SEs") and SE-eligible workloads provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines *(i.e.,* zIIPs, zAAPs, and IFLs). IBM authorizes customers to use IBM SEs only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the “Authorized Use Table for IBM Machines” provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html (“AUT”).

No other workload processing is authorized for execution on an SE.

IBM offers SEs at a lower price than General Processors/Central Processors because customers are authorized to use SEs to process only certain types and/or amounts of workloads as specified by IBM in the AUT.

# Credits

- Your z/VM 7.1 Performance team:
  - Bill Guzior
  - Steve Jones
  - Dave Spencer
  - Xenia Tkatschow
  - Brian Wade
  - Dave Wierbowski
- The z/VM 7.1 Performance Report:
  - John Franciscovich
  - Bob Neill
  - Patty Rando
  - Xenia Tkatschow
  - Brian Wade
  - Dave Wierbowski
- Your z/VM 7.1 Performance Toolkit team:
  - John Franciscovich
  - Don McGlynn
- This chart deck:
  - Xenia Tkatschow
  - Brian Wade
  - Dave Wierbowski
- Thanks also to anyone we inadvertently failed to mention

# Agenda

- z/VM 7.1 regression performance
- Performance of new functions
  - Dump enhancements
- z/VM 6.4 APARs in the base of z/VM 7.1
- Small performance fixes in z/VM 7.1
- z/VM 7.1 performance APARs
- Changes to monitor records
- z/VM Performance Toolkit changes
- z/VM 7.1 4Q18 updates
  - TLS/SSL elliptic curve cryptography
- z/VM 7.1 2Q19 updates
  - Paging to EAV volumes
  - Vswitch priority queueing
  - Eighty logical processors
- Summary

# Regression Performance

# z/VM 7.1 Regression Performance

- We ran about 120 scenarios:
  - Some non-SMT, some SMT-2
  - Some using Apache static file web serving in various ways
  - Some using our VIRSTOR load generator
  - Some using DayTrader (a WAS and DB/2 workload)
  - Some storage-rich, some storage-constrained
  - Some 1-core, some 3-core, some mid-sized, and some as large as 64-core
  - All on z14

- z/VM levels we used:
  - Base runs were done on z/VM 6.4 plus all closed PTFs as of February 1, 2018
  - Comparison runs were done on the z/VM 7.1 *code freeze* driver of May 24, 2018

- Typical measures of accomplishment:
  - ETR (external transaction rate): units of application work per second
  - ITR (internal transaction rate): what ETR would scale to if the LPAR could run this workload completely busy

- Our findings:
  - ETR ratios, comparison/base: mean ($\mu$) = 1.001, standard deviation ($\sigma$) = 0.023
  - ITR ratios, comparison/base: $\mu$ = 1.010, $\sigma$ = 0.051
  - These are within our routinely observed run variation

# New Function

# Some Words About New Function

- With the change to continuous delivery, a release isn't as rich in debuting function as it used to be
  - A release is now more like a rollup of recent new-function PTFs
- But we are still shipping plenty of new function!
  - http://www.vm.ibm.com/newfunction/
- We update our z/VM Performance Report http://www.vm.ibm.com/perf/reports/zvm/html/ concurrently with the appearance of new function
- And now let's visit the function that made its first appearance in z/VM 7.1

# Dump Enhancements

- As central storage has grown, we have had to enhance dumping

- Two new operands on SNAPDUMP and SET DUMP:
  - PGMBKS NONE to omit page management blocks (PGMBKs) from the dump
  - FRMTBL NO to dump the frame table (the map of real storage) in a more efficient way

- CPU efficiency improvement: dumper now uses Prefetch Data (PFD) to have the CPU prefetch cache lines (rows) of the real frame table

- Net effect in our workload: with all enhancements in play, compared to z/VM 6.4,
  - Size of the dump was decreased by 99%
  - Elapsed time to dump was decreased by 97%

<- WOW

- Read the report: http://www.vm.ibm.com/perf/reports/zvm/html/710dmp.html

# Monitor Record Changes

- Some small upward-compatible changes
- Read the article: http://www.vm.ibm.com/perf/reports/zvm/html/710man.html

# And By The Way…

- One can no longer dedicate a logical processor to a virtual processor of a guest
- Compared to z/VM 6.4, there are command or output changes related to:
  - EAV minidisks
  - Crypto
  - Encrypted paging
  - TLS/SSL CRYPTO APVIRT
  - Dumping
  - HiperDispatch
  - Resource pools
- Read more about it: http://www.vm.ibm.com/perf/reports/zvm/html/710man.html

# APARs and Small Fixes

# Performance-related Small-enhancement APARs Against z/VM 6.4

- PI72106: TCP/IP SSL exploit CRYPTO APVIRT
- PI73016: TCP/IP exploit OSA Express6S
- VM65846: Can help reduce the need to intercept out of SIE to enforce a guest's virtual architecture level
- VM65929: Support concurrent I/O on XIV EDEVs
- VM65942: Support for the z14. Includes per-VCPU priv op tracking in monitor. (Went PE: also apply VM66071)
- VM65987: Lets guests use the Guarded Storage Facility.
- VM65988: Improvements to the CP spin lock manager
- VM65989: CP dump enhancements
- VM66026: Monitor enhancements for zHPF.
- VM66063: New unparking models can decrease PR/SM overhead
- VM66090: Improves performance of simulation of I/O to PCI functions
- VM66095: Improve Monitor for FCP chpids and devices
- VM66098: Support Extent-Space-Efficient storage.

These are all in the base of z/VM 7.1.

# Performance-related Repair APARs Against z/VM 6.4

- VM65644 (1701): SCSI monitor fields not filled
- VM65741 (1701): make all 3390-A eligible for MDC
- VM65885: Perfkit needs deprecated HPF monitor fields
- VM65886 (1601): CCW fast-trans incorrectly marked minidisk I/O as ineligible for HyperPAV aliases
- VM65916: HiperSockets Guest LAN NIC lost initiative
- VM65946: SECUSER output is slow
- VM65979: Removed unnecessary MDC purge done during HyperSwap
- VM65992 (1701): HiperSockets performance issue on short busy
- VM65998 (1701): crypto polling too frequent
- VM66016: Abend during zHPF paging error recovery
- VM66026 : Monitor enhancements for HyperPAV and PAV aliases (went PE: also apply VM66036)

(xxxx) is RSU number

These are all in the base of z/VM 7.1.

# Small Performance Fixes in z/VM 7.1

- **Relative max-share:** the arithmetic for calculating the entitlement associated with a relative max-share was repaired.
- **Cache behavior:** in virtual networking, certain control blocks were reorganized to reduce cache contention.
- **Vswitch load balancing:** The vswitch load balancing algorithm was improved.
- **CP use of Diag x'9C':** CP no longer issues unnecessary or ill-advised Diag x'9C's to PR/SM.
- **Skipped monitor intervals:** On low-utilization systems there was risk for skipping monitor intervals. This is repaired.
- **Sleeping sibling CPU:** In SMT-2, when a CPU queued work to its own DV and its sibling CPU was asleep, the sleeping sibling did not get awakened. This is repaired.

# Performance-related Repair APARs against z/VM 7.1

- As of Aug 14, 2019
- VM65690: z/VM hang due to errors in machine check recovery (6.4 and 7.1)
- VM65858: relative share for guests in CPU pool not respected (6.4 and 7.1)

# Some Words About VM65858 and CPU Pools

- The problem found was that we failed to put a guest onto the limit list when we should have done so; in other words, there was a legitimate bug in enforcing the pool limit
- But we also found many people do not understand what CPU pools really do
- Share settings mediate CPU power in the system as a whole, *not within a CPU pool*
- CPU pool limiting works like this:
  - When the pool reaches its limit, all members get put onto the limit list for “a while”
  - After “a while” passes, all members get removed from the limit list
- Again, share settings do not serve as a way to parcel out the power of a CPU pool
- BTW, this is the same as how PR/SM’s LPAR group capping works
- There is a fixed-if-next documentation update sitting on our desks back home

# z/VM Performance Toolkit

- VM66085: HyperPAV Paging
  - (New) Use of HyperPAV aliases: HPALIAS, HPSHARE
  - (New) Per-volume reports: VOLUME, VOLLOG
  - (Changed)
    - Pretty much every disk-related report: CACHELOG, CACHEXT, CPOWNLOG, CTLUNIT, DEVICE, DEVICE CPOWNED, DEVICE HPF, DEVLOG, DEVMENU, HPFLOG, IOCHANGE
    - Paging-related reports: AGELLOG, STORAGE
    - Others: BENCHMRK

- VM65959: formatted output collector can now handle metrics for CPU pools, type capping, multithreading depth, and group capping
- VM65959: formatted output collector now contains correct count of CPUs
- VM66088: LPAR and LSHARACT display CPU counts correctly
- VM66164: formatted output collector now contains data for the KVLDEVICE group

- More info: http://www.vm.ibm.com/perf/reports/zvm/html/710man.html

# Fourth-Quarter 2018

# 4Q18 Regression Behavior

- We checked it on a z14
- Same basic procedure as for z/VM 7.1 base
- ETRR $\mu = 0.992$, $\sigma = 0.097$
- ITRR $\mu = 0.994$, $\sigma = 0.097$
- These are within our routinely observed run variation

# TLS/SSL Elliptic Curve Cryptography

- z/VM V7.1 provides **stronger** and **faster** security ciphers for the TLS/SSL server with elliptic curve (EC) cryptography
- A Telnet Connection rampup workload using an EC cipher showed a 77% reduction in CPU/tx when compared back to an equivalent non-EC cipher
- PI99184 (TCP/IP)
- More info: http://www.vm.ibm.com/perf/reports/zvm/html/4q8qk.html

# Second-Quarter 2019

# 2Q19 Regression Behavior

- We checked it on a z14
- Same basic procedure as for z/VM 7.1 base
- ETRR $\mu = 0.998$, $\sigma = 0.020$
- ITRR $\mu = 0.997$, $\sigma = 0.019$
- These are within our routinely observed run variation

# Paging to EAV Volumes

- z/VM can now page to all cylinders of an EAV volume
  - The new maximum size of a paging volume is 1182006 cylinders (811 GB)

- Our performance evaluation was very light
  - We just made sure an I/O to an EAV volume took the same amount of time as an equivalent I/O to a non-EAV volume

- Exploitation considerations
  - For equal paging *space*, you can now have fewer paging *volumes*
  - Be careful not to decrease paging *I/O concurrency*
    - Maybe replace N small volumes with K EAVs and N-K HyperPAV aliases
    - Over time you might find even fewer HyperPAV aliases would be needed
  - Interesting Perfkit reports:
    - HPALIAS – do I have enough aliases?
    - DEVICE CPOWNED – how is my paging subsystem doing?

- New-function APARs (z/VM 7.1 only):
  - CP:  VM66263 (nucleus and CPFMTXA)
  - CMS:  VM66297 (Documentation for a SMAPI entry point)
  - Perfkit:  VM66293 (more later in this presentation)

# Vswitch Priority Queueing

- Exploits the priority queueing capabilities of the OSA-Express adapter
  - Provides a means to direct a VNIC's outbound packets to a low-priority, normal-priority, or high-priority output queue on a vswitch's uplink port
    - Priority is relevant only when the OSA is fully saturated
  - Different VNICs can be provided with different grades of uplink service
- Evaluated using streaming workloads, request/response workloads, and mixed workloads
- Results
  - CPU efficiency benefit observed in all the workloads
    - Even when the prioritization function was not exploited
  - The effect of the prioritization was:
    - Observed when the OSA was heavily utilized
    - Not observed when the OSA was lightly utilized
    - Effective at keeping a heavy streaming workload from oppressing a light request-response workload.
- New-function APARs (z/VM 7.1 only)
  - VM66219 (CP)
  - PH04703 (TCP/IP)
  - VM66223 (DirMaint)
- More info: http://www.vm.ibm.com/perf/reports/zvm/html/2q9sr.html

# Eighty Logical Processors

- On z14, the support limit is now 80 logical processors
  - Non-SMT: 80 logical processors
  - SMT-1: 40 logical cores, and only even-numbered logical processors => 40 processors
  - SMT-2: 40 logical cores, and
    - IFL cores: per core, an even-numbered and an odd-numbered processor
    - All other core types: per core, only an even-numbered logical processor
- What did we run?
  - Memory-rich: Linux AWM -> Linux Apache, HTTP serving
  - Memory-constrained: Linux AWM -> Linux Apache, HTTP serving
- Bottom lines:
  - The system scales OK to 80 processors
  - If you use layer-3 vswitch, do not build a giant LPAR and put all your guests onto one giant layer-3 vswitch
- VM66265 (CP), VM66296 (standalone dump)
  - VM66301 (CP save area misuse) is a prerequisite
- More info: http://www.vm.ibm.com/perf/reports/zvm/html/2q9r2.html

# z/VM Performance Toolkit

- Support for EAV paging
  - z/VM 7.1: VM66293
  - Report on larger devices
  - Updated reports: FCX109 DEVICE CPOWNED, FCX146 AUXLOG, FCX170 CPOWNLOG
  - More information: https://www-01.ibm.com/support/docview.wss?uid=isg1VM66293

- Support for 80 logical processors
  - z/VM 7.1: VM66292
  - z/VM 6.4: VM65863
  - All processor IDs are now displayed in hexadecimal
  - Updated reports: FCX100 CPU, FCX126 LPAR, FCX144 PROCLOG, FCX174 UTRANDET, FCX180 SYSCONF, FCX232 IOPROCLG, FCX239 PROCSUM, FCX287 TOPOLOG, FCX298 PUORGLOG, FCX299 PUCFGLOG, FCX300 DSVCLOG, FCX301 DSVBKACT, FCX303 DSVSLOG, FCX304 PRCLOG
  - More information: https://www-01.ibm.com/support/docview.wss?uid=isg1VM66292

# Summary

# Summary

- z/VM 7.1 offers good regression behavior compared to z/VM 6.4
- Performance or capacity improvements:
  - Collecting dumps
  - Paging to EAVs
  - Elliptic curve cryptography
  - Vswitch priority queueing
  - Eighty logical processors
- Improvements in z/VM Performance Toolkit
  - Support for HyperPAV paging
  - Support for EAV paging
  - Support for 80 logical processors
- There are a few new or changed monitor records
- Visit us on the web at http://www.vm.ibm.com/perf/reports/zvm/html/

# Thank you!

Send feedback to:
Brian Wade, bkw@us.ibm.com

Also, visit our z/VM Performance Report:
http://www.vm.ibm.com/perf/reports/zvm/html/