

z/OS Performance “HOT” Topics

Kathy Walsh
IBM

Monday, August 12, 2013
Session Number: 14022



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

*BladeCenter®, DB2®, e business (logo)®, DataPower®, ESCON, eServer, FICON, IBM®, IBM (logo)®, MVS, OS/390®, POWER6®, POWER6+, POWER7®, Power Architecture®, PowerVM®, S/390®, Sysplex Timer®, System p®, System p5, System x®, System z®, System z9®, System z10®, Tivoli®, WebSphere®, X-Architecture®, zEnterprise®, z9®, z10, z/Architecture®, z/OS®, z/VM®, z/VSE®, 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 registered 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 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.

Notice Regarding Specialty Engines (e.g., 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 (e.g., 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 only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Agenda

- Processor Information
 - New zBC12 Processor and new zEC12 GA2 Support
 - PR/SM Absolute Capping
 - Instructions Counts
 - Warning Track
 - LPAR Interrupt Delay Time
 - zEnterprise Data Compression (zEDC)
 - Flash Express
- Performance and Capacity Planning Topics
 - New XES Support
 - CPU MF and HIS Support
 - zPCR Latest Status
 - zIIP Capacity Planning in DB2
 - New GRS Support
 - New
- ▶ Addendum
 - Older APARs or Performance Information

www.ibm.com/support/techdocs

Techdocs - the Technical Sales Library

This site provides access to the Technical Sales Support organization's technical information databases. It gives you access to the most current installation, planning and technical support information available from IBM pre-sales support, and is constantly updated. You can browse or search these databases by date, document number, product, platform, keywords, etc.

New to Techdocs? Take a look at our [detailed introduction](#), which describes the document categories available (those listed on the navigation area on the left side of this page).

Rather than browse these categories, as a convenience you may enter a search of the full **Techdocs** database, or of any category you wish, here:

Search: Allow word variants
 "Fuzzy" search
 for:
 Hits: Order by:
 Include docs updated: [Help for Search](#)

Also available: our [Advanced search](#), where you can select documents based on various assigned document attributes.

New to Techdocs?
 Is this your first visit to **Techdocs** (the Technical Sales Library)?
 → [Learn more](#)

Returning to Techdocs?
 Looking for what's new in the **Techdocs Library**?
 → [Latest updates](#)

Need Technical Support?
 Looking for support resources or other documents and tools?
 → [Support & downloads](#)

zBC12 Overview

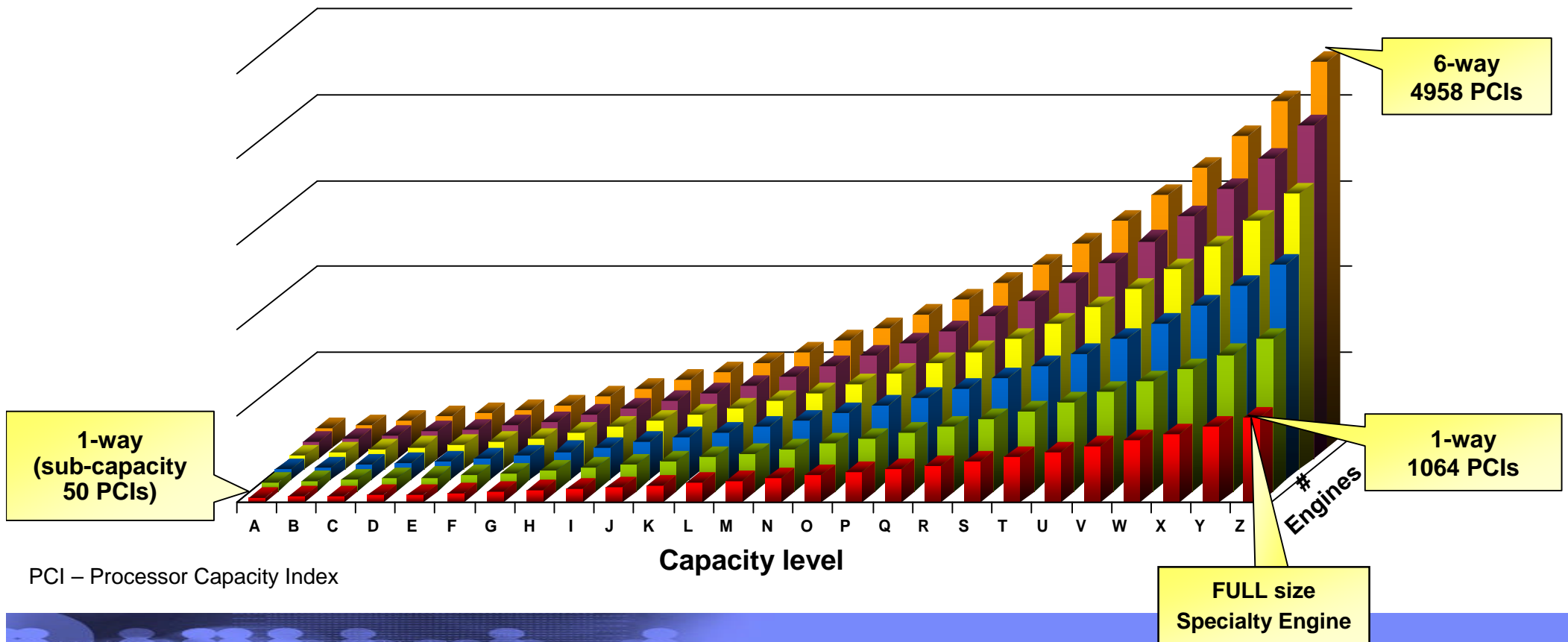


- **Machine Type**
 - 2828
- **2 Models**
 - H06 and H13
 - Single frame, air cooled
 - Non-raised floor option available
 - Overhead Cabling and DC Power Options
- **Processor Units (PUs)**
 - 9 PU cores (using 4 and 5 core PU SCMs) per processor drawer (One for H06 and two for H13)
 - 2 SAPs per system, standard
 - 2 spares designated for Model H13
 - 1 Integrated firmware processor
 - Dependant on the H/W model - up to 6 or 13 PU cores available for characterization
 - 156 capacity settings
- **Memory**
 - Up to 512 GB for System including HSA
 - System minimum = 8 GB (Model H06), 16 GB (Model H13)
 - 16 GB fixed HSA separately managed
 - RAIM standard
 - Maximum for customer use 496 GB (Model H13)
 - Increments of 8 or 32 GB
 - Flash Express
- **I/O**
 - Support for non-PCIe Channel features
 - PCIe Gen2 channel subsystem
 - Up to 64 PCIe Channel features
 - Up to 2 Logical Channel Subsystems (LCSSs)
- **STP - optional** (No ETR)

zBC12 Sub-capacity Processor Granularity

- The zBC12 has 26 CP capacity levels (26 x 6 = 156)
 - Up to 6 CPs at any capacity level
 - All CPs must be the same capacity level
 - All specialty engines run at full speed
 - Processor Value Unit (PVU) for IFL = 100

Number of zBC12 CPs	Base Ratio	Ratio z114 To zBC12
1 CP	z114 Z01	1.36
2 CPs	z114 Z02	1.37
3 CPs	z114 Z03	1.37
4 CPs	z114 Z04	1.36
5 CPs	z114 Z05	1.36
6 CPs	z114 Z05	1.58



System z Cache Topology – z114 vs. zBC12 Comparison

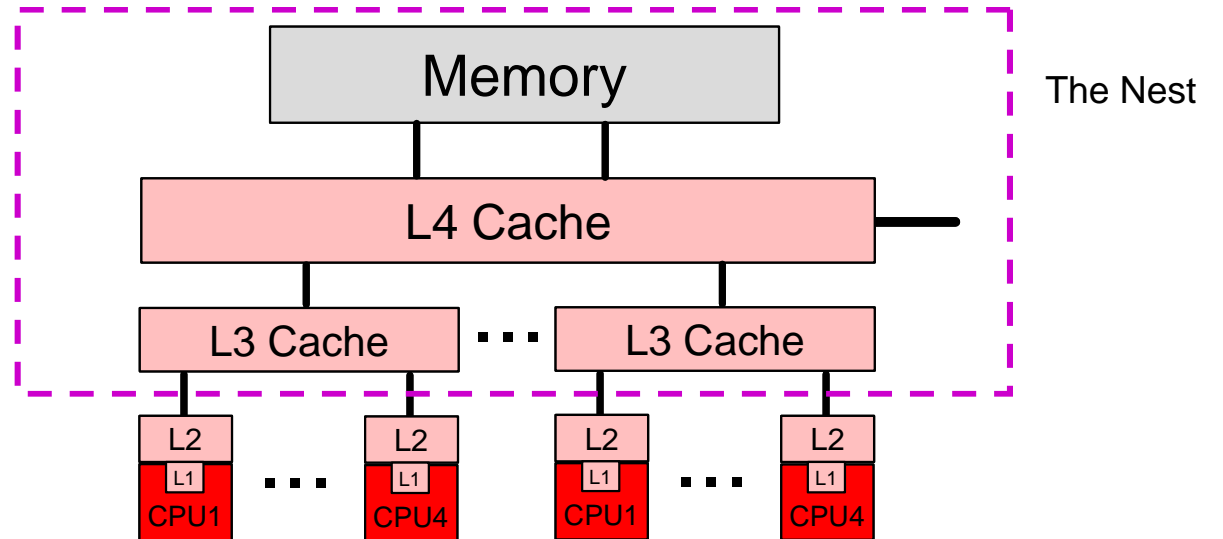
■ z114

▶ CPU

- 3.8 GHz
- Out-Of-Order execution

▶ Caches

- L1 private 64k i, 128k d
- L2 private 1.5 MB
- L3 shared 12 MB / chip
- L4 shared 96 MB / book
 - 24 MB to each core



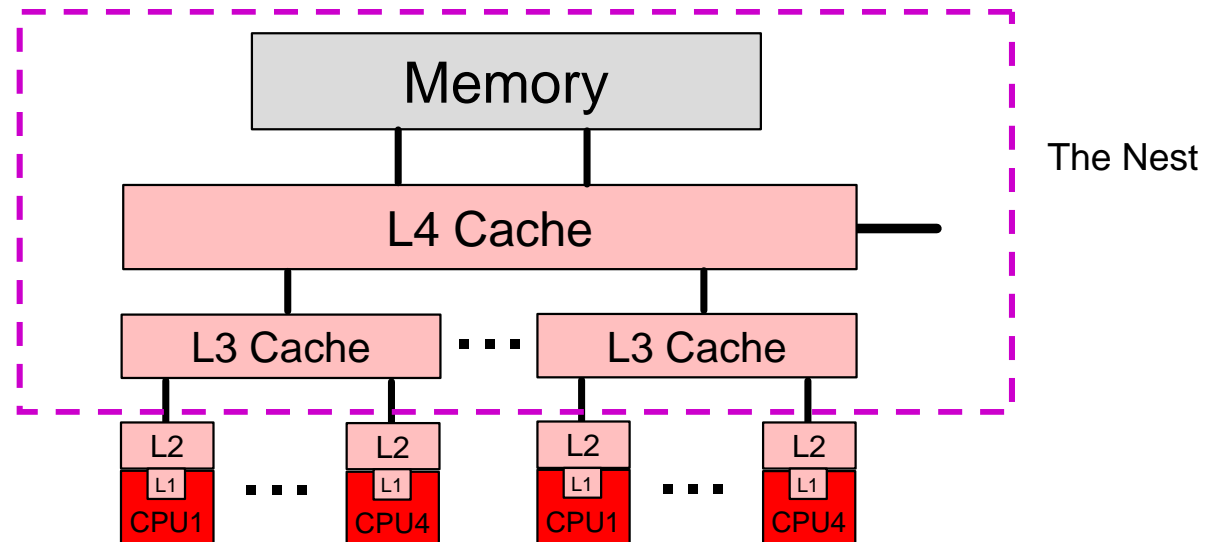
■ zBC12

▶ CPU

- 4.2 GHz
- Enhanced Out-Of-Order

▶ Caches

- L1 private 64k i, 96k d
- L2 private 1 MB i + 1 MB d
- L3 shared 24 MB / chip
- L4 shared 192 MB / book
 - 32 MB to each core



Specialty CP Update

■ 2012: zEC12 SOD

- With the zEC12 announcement August 28th, 2012: The IBM zEnterprise EC12 is planned to be the last high-end System z server to offer support for zAAP specialty engine processors
- IBM recommends users with zAAPs to consider planning for migration of zAAPs to zIIPs using zAAP on zIIP support

■ 2013: New zIIP and zAAP ratios

- zEC12 and zBC12 servers only
 - Ratio is now 2:1; for every GCP, may optionally purchase either two zIIPs and/or two zAAPs
- For servers before the zEC12 and zBC12
 - Ratio remains 1:1; for every GCP, may optionally purchase one zIIP and/or one zAAP

zPCR Latest Status

- Version 8.4 (7/2013)
 - IBM zEnterprise BC12 (zBC12) has been added
 - LSPR data continues to be based on z/OS 1.13
 - zAware can be configured as an SCP type in either the GCP or IFL pool
- Support for zIIP/zAAP 2:1 ratio on zEC12 / zBC12 targeted for 4Q13
 - In interim configured additional specialty CPs and when warning box launches click on “I am upgrading from a previous ...”

Notice

The LPAR Host processor specified has a nonstandard zAAP or zIIP configuration. For a new processor, the number of zAAP and/or zIIP CPs cannot exceed the number of General Purpose CPs. For certain processor upgrade scenarios the number of zAAP and zIIP CPs allowed is based on the number previously installed, and therefore, could exceed the number of General Purpose CPs.

Click the checkbox below if the statement applies.
Otherwise only a standard configuration will be allowed.

I am upgrading from a previous System z family that included zAAP or zIIP CPs.

14219: zPCR Capacity Sizing Lab - Part 1: Intro and Overview, Thursday, 3:00 PM

13954: zPCR Capacity Sizing Lab - Part 2: Hands-on Lab, Thursday, 6:00 PM PM

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

PR/SM Absolute Capping

- zEC12 GA2 and zBC12 allows specification of an “**absolute capping limit**”
 - Provided originally for non z/OS images like zLinux
 - Expressed in terms of 1/100ths of a processor (0.01 to 255.0)
 - Specified independently from the LPAR weight
 - The shared partition's processing weight still dictates the logical partition priority compared to other shared logical partitions
 - Most effective for absolute caps higher than the capacity the partition's relative weight would deliver
 - Absolute capping is not recommended to be set below the capacity the logical partition's weight would deliver
 - Insensitive to capacity changes or LPAR (de)activations
 - Specified per processor type in image profile and partition controls panel
- Unlike initial capping may be used concurrently with defined capacity and/or group capacity management
 - The respective absolute capacity becomes effective before other capping controls
 - WLM/SRM recognizes new cap, e.g. for routing decisions

PR/SM Absolute Hardware Capping

Customize Image Profiles: SSYS : SOSP15 : Processor

SSYS
SOSP15
General
Processor
Security
Storage
Options
Load
Crypto

Group Name <Not Assigned>

Logical Processor Assignments

Dedicated processors

Select	Processor Type	Initial	Reserved
<input checked="" type="checkbox"/>	Central processors (CPs)	3	1
<input type="checkbox"/>	System z application assist processors (zAAPs)	0	0
<input checked="" type="checkbox"/>	System z integrated information processors (zIIPs)	2	0

Not Dedicated Processor Details for :

CPs zIIPs

CP Details

Initial processing weight: 100 (1 to 999) Initial capping

Enable workload manager

Minimum processing weight: 0

Maximum processing weight: 0

Absolute Capping: None Number of processors (0.01 to 255.0) 1.23

Cancel Save Copy Profile Paste Profile Help

z/OS 2.1 - Instruction Counts

- Provide a more consistent metric which does not see as much CPU variability due to:
 - Impacts of hardware caching
 - LPAR configurations
 - Software stack
 - Workload interactions – dispatch rate
- New support provides Instruction Counts in SMF 30 records
 - Supported on z10 processors and later which supports CPUMF
 - Similar to the SMF 30 CPU time fields
 - SMFPRMxx new keyword `SMF30COUNT|NOSMFCOUNT`
 - Requires CPUMF basic and extended counters in HIS be active

14220: WSC Short Stories and Tall Tales: Wed, Aug 14, 2013, 1:30 PM

New Function - Warning Track

- Requires zEC12 or zBC12
- APARs OA37186 and OA37803
- PR/SM recognizes a logical CP has to be undispached from a physical CP and issues a Warning Track Interrupt (WTI, aka EXT 1007) and sets a grace period for z/OS to return the logical CP to PR/SM
 - If grace period expires before z/OS returns logical CP to PR/SM, PR/SM undispaches the logical CP and will redispach it later
- Once z/OS receives a WTI
 - Saves status for the work running (makes work available to dispatch on another CP)
 - Issues a DIAG 49C to return the CP to PR/SM
 - Becomes PR/SM's responsibility to redispach the logical CP when able and resume execution at instruction after DIAG 49C
 - z/OS keeps track of the following statistics:
 - How many DIAG 49Cs were successful / unsuccessful in returning CP to PR/SM before the end of the grace period
 - How long PR/SM undispached the logical CP for successful DIAG 49Cs

LPAR I/O Interrupt Delay Processing

- APAR OA37160 – New Function
 - z/OS 1.12 and above
 - Requires zEC12 or zBC12 processor
- New Interrupt Delay Time
 - Keeps track of the time from when a subchannel is made status pending with primary status to when the status is cleared by TSCH
 - Tracking the accumulated delay encountered due to PR/SM needing to dispatch z/OS processing the interrupt as well as any z/OS delay
- APAR OA39993 - RMF support
 - RMF 74.1 record (device) and RMF 79.9
 - Enhanced RMF Post-processor device report

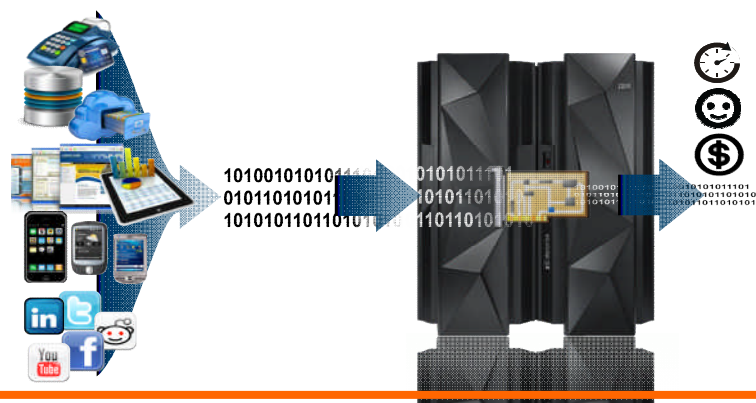
STORAGE	DEV	DEVICE	NUMBER	VOLUME	PAV	LCU	DEVICE	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG
GROUP	NUM	TYPE	OF	CYL	SERIAL		ACTIVITY	RESP	IOSQ	CMR	DB	INT	PEND	DISC	CONN	%	%	%	AVG	%	
							RATE	TIME	TIME	DLY	DLY	DLY	TIME	TIME	TIME	CONN	UTIL	RESV	ALLOC	ALLOC	
XTEST	2208	33903	3339	TRXSX9	1	0032	0.001	.384	.000	.128	.000	.123	.256	.000	.128	0.00	0.00	0.0	0.0	100.0	
XTEST	2209	33903	3339	TRXSXA	1	0032	0.001	.256	.000	.000	.000	.135	.256	.000	.000	0.00	0.00	0.0	0.0	100.0	
	220A	33909	10017	TRXT01	1	0032	0.000	.000	.000	.000	.000	.000	.000	.000	.000	0.00	0.00	0.0	0.0	100.0	
	220B	33909	10017	TRXT02	1	0032	0.000	.000	.000	.000	.000	.000	.000	.000	.000	0.00	0.00	0.0	0.0	100.0	
	220C	33909	10017	TRXT03	1	0032	0.000	.000	.000	.000	.000	.000	.000	.000	.000	0.00	0.00	0.0	0.0	100.0	
	220D	33909	10017	TRXT04	1	0032	0.000	.000	.000	.000	.000	.000	.000	.000	.000	0.00	0.00	0.0	0.0	100.0	

IBM zEnterprise Data Compression (zEDC)

New data compression offering that can reduce resource usage

What is it?

- ✓ *A combined software (z/OS V2.1) and hardware (zEC12 / zBC12 and zEDC Express) solution designed to help reduce resource consumption, disk utilization and optimize cross platform exchange of data*



How is it different

- **Performance:** Efficient alternative for larger files. Reduced CPU overhead for SMF diagnostic jobs
- **Efficient:** Optimized algorithms scan text to locate the re-use of phrases and refers back to earlier references
- **Industry Standard:** Compatible with open zlib based compression – widely used across all platforms
- **Economical:** Reduced DASD space requirements and improved effective bandwidth without significant CPU overhead***

15% reduction in elapsed time for SMF extraction with up to **40%** reduction for CPU time*
 Logger overhead reduced by up to **30%** **

* When running an SMF extraction/dump against an SMF logstream with records compressed by zEDC
 ** The amount of data sent to an SMF logstream can be reduced by up to 75% using zEDC compression – reducing logger overhead
 *** SOD for BSAM/QSAM access methods
 All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

z/OS and z/VM V6.3 Exploitation of zEDC Express

- Introductory Use:
 - SMF Log Data - Using SMF logger, either DASD-only or structure based, looking to reduce the logger overhead or collect additional data
 - Installations sending and receiving large files

- Planned Use (SOD):*
 - Systems with large volumes of extended format BSAM/QSAM sequential data*
 - Systems with data on DS8870 with flash drives who want to use it more efficiently when storing extended format BSAM/QSAM sequential data*
 - Clients using Java where they create a stream of compressed data*
 - z/VM V6.3 guest exploitation support*

*All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

RMF Support for zEDC Express

- New support in RMF will provide information on PCI Express based functions
 - zEnterprise Data Compression (zEDC) capability using zEDC Express
 - RDMA (Remote Direct Memory Access) over Converged Enhanced Ethernet
- Information added via new subtype to SMF 74 (74.9) called PCI Express Based Function Activity
- Written by RMF Monitor III
 - General PCIE Activity – both zEDC and RDMA activity
 - Hardware Accelerator Activity
 - Hardware Accelerator Compression Activity
 - Types of data provided:
 - I/O Queue Time
 - I/O Execution Time
 - Number of compressed bytes in and out
 - Number of decompressed bytes in an out
 - Device drive buffer statistics

SMF Logstream Exploitation of zEDC Express (OA41156)

- Writing to SMF Logstream:
 - Controlled by new SMFPRMxx keyword – COMPRESS
- Reading via IFASMF DL utility:
 - Uncompressed records will be read as done today
 - Compressed blocks encountered, decompression will be attempted
 - New keyword, SOFTINFLATE, provided to allows blocks to be read on down level systems or on z/OS V2.1 systems running without the zEDC Express hardware
 - Due to CPU usage, SOFTINFLATE is off by default
 - If SOFTINFLATE NOT specified on IFASMF DL jobs running on down level systems or on z/OS V2.1 running on hardware without zEDC enabled, then an error will occur
 - No records will be deleted from the SMF logstream
 - IFASEXIT will return an error when a compressed block is found

Hardware	Software	Accelerators	Decompression	Compression
zEC12 GA2 or zBC12	z/OS V2.1	Active	Hardware	Hardware
zEC12 GA2 or zBC12	z/OS V2.1	Not Active	Software (via SOFTINFLATE)	None
Pre-zEC12 GA2	z/OS V2.1	N/A	Software (via SOFTINFLATE)	None
All	z/OS V1R13 or V1R12	N/A	Software (via SOFTINFLATE and PTFs)	None

Java Exploitation via zlib*

- The Open Source zlib library is a highly used cross platform library
- IBM is providing in z/OS V2.1 an enhanced zlib library as part of the z/OS USS Application Service Base (FMID HOT7790)
 - The Java Runtime will provide zEDC compression access via the *java.util.zip* package
 - This package will replace the standard zlib library which is currently used with the IBM provided zlib library
 - May see reduced disk space or network bandwidth requirements using zEDC compression in Java with minimal CPU overhead
 - CPU cost of compression may be reduced using zEDC compression compared to software based zlib compression

*All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

■ System z Capacity Planning Opportunities:

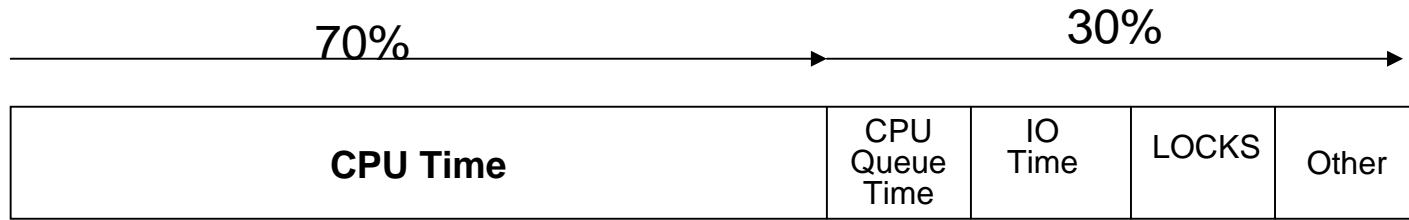
- Per thread (engine) speed improvements for CMOS CPs is slowing dramatically
 - Every CMOS platform is facing this issue
 - Future capacity gains will be by adding more CPs rather than much faster CPs
 - Enhances need for parallel operation and more reliance on parallel sysplex

- Availability of subcapacity models continues to grow
 - Provide capacity as more, slower processors increasing parallelism
 - Especially useful in environment with large number of LPARs
 - Additional capacity can be acquired in smaller increments
 - Receive benefit since Specialty CPs run at full n-way speed

**Impact of these trends will most likely be seen first in the
Batch Window**

Workload Considerations

Online Transaction



Processor	CPU Time	Other Time	Total
z196-708	.028	.012	.040
zEC12-707	.023	.012	.035
zEC12-611	.036	.012	.048

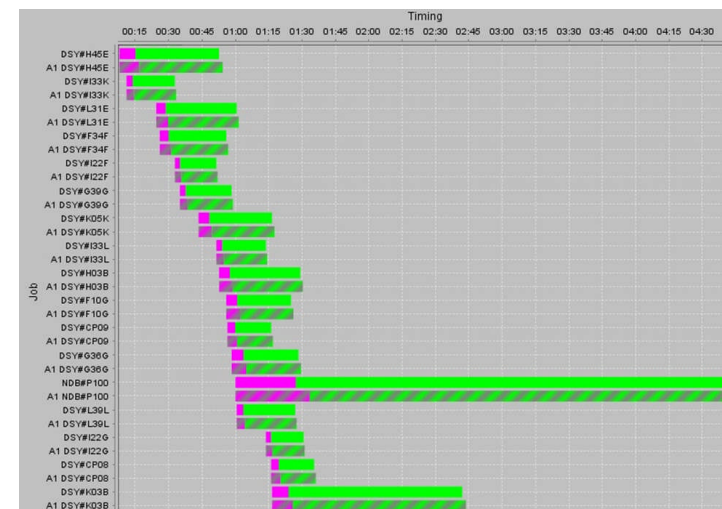
The real issue is in the batch window where CPU time can be significant, and CP speed issues can impact elapsed time and job network time

Enhanced SMF 30 Reporting

- APAR OA39629 – New Function
- New SMF 30 function to provide enhanced reporting in the CPU Accounting section
 - z/OS 1.12 and above
- Highest percent of CPU time used by a single task in the address space in the interval, step, or job
 - SMF30_Highest_Task_CPU_Percent
- Program name associated with the task with the highest percentage of CPU time in the address space
 - SMF30_Highest_Task_CPU_Program

IBM System z Batch Network Analyzer (zBNA)

- IBM System z Batch Network Analyzer
 - A free, “as is” tool to analyze batch windows
 - Available to Customers, Business Partners and IBMers
 - PC based, providing graphical and text reports
 - Includes Gantt charts and support for Alternate Processors (what if scenarios)
 - Next release* targeted for 4Q2013
 - Support for SMF 42 records to understand the “Life of a Data Set”
 - Support for SMF 14/15 records to understand zEDC compression candidates



- Available Now on Techdocs

<https://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS5132>

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

Why use zBNA?

- Perform "what if" analysis and estimate the CPU upgrade effect on batch window
- Identify job time sequences based on a graphical view
- Filter jobs by attributes like CPU time / intensity, job class, service class, etc.
- Review the resource consumption of all the batch jobs
- Drill down to the individual steps to see the resource usage
- Identify candidate jobs for running on different processors
- Identify jobs with speed of engine concerns (top tasks %)

Flash Express – What is it?

FLASH Express

- Flash Express is a PCIe IO adapter with NAND Flash SSDs
- Physically comprised of internal storage on Flash SSDs
- Used to deliver a new tier of memory- storage class memory
- Uses **PCIe I/O drawer**

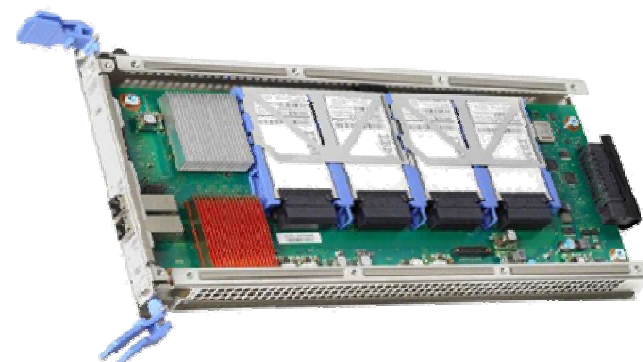
- Sized to accommodate *all LPAR paging*
 - Each **card pair** provides **1.4 TB** usable storage (2.8 TB total)
 - Maximum 4 card pairs (4 X1.4=5.6 TB)
- Supported on z/OS V1.13 plus web deliverable

Designed for continuous availability

- Concurrent Firmware update for service
- RAID 10 design
- **Immediately usable**
 - No capacity planning needed
 - No intelligent data placement needed

- **Secured**

- Flash Express adapter is protected with 128-bit AES encryption
- Key Management provided based on a Smart Card
- Secure Cryptographic Erase meets audit requirements

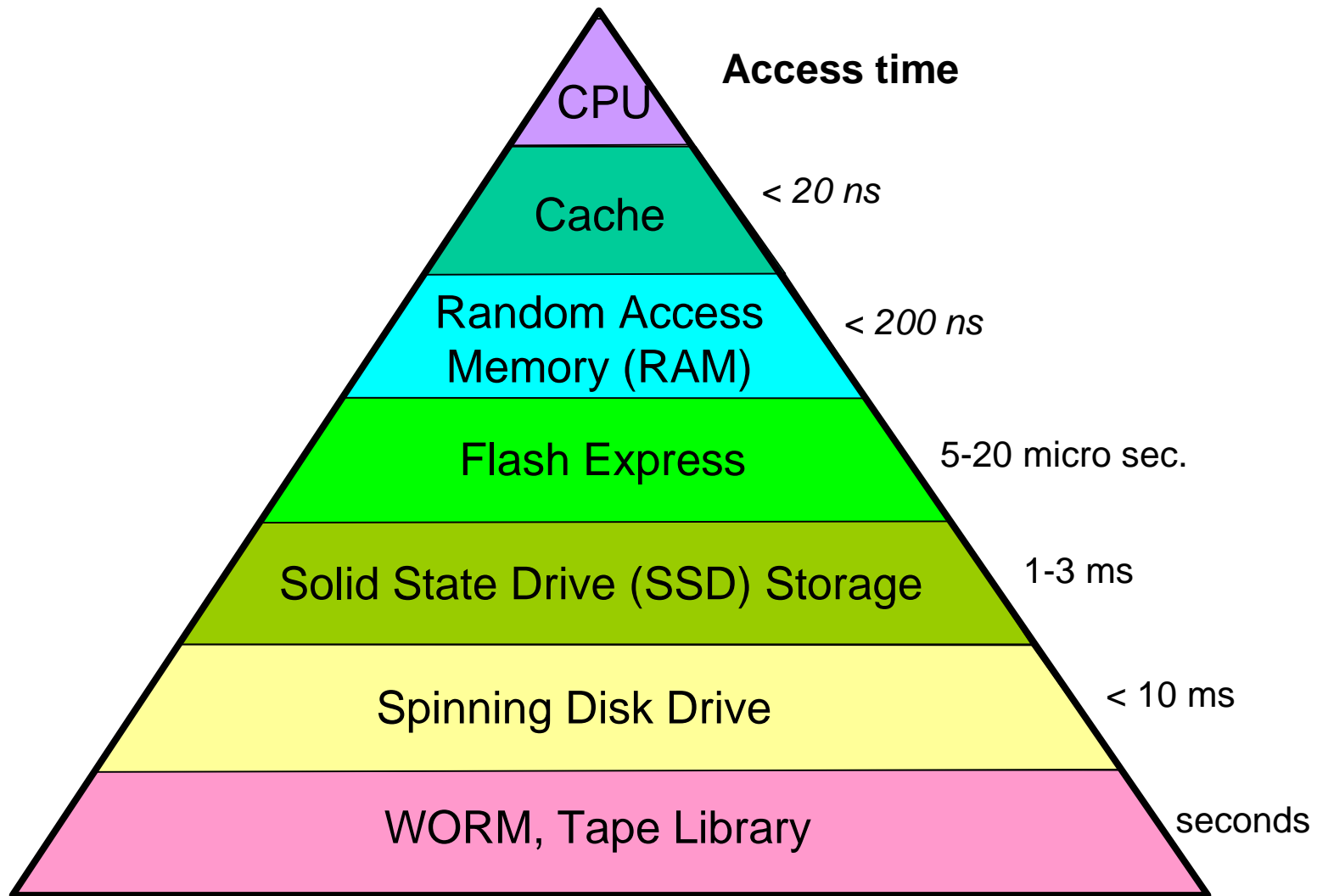


One Flash Express Card

Session 14119: Experiences with Using IBM zEC12 Processor Flash Memory

Tue, Aug 13: 1:30 PM-2:30 PM

Relative Access Times for different technologies



Session 13464: Processor architecture and the importance of the storage hierarchy -
Tue Aug 13th - 9:30 AM-10:30 AM

OA41661 – New XES Sync/Async Support

- CF ASYNC requests have longer service time and consume more system resources than CF SYNCH requests
 - Long-running SYNC requests eventually cost more than the same operation done ASYNC due to the CPU spinning on the SYNC request
 - There is a crossover point where issuing an operation asynchronously costs less in terms of CPU than issuing it synchronously
 - XES compares actual observed synch service times against dynamically calculated thresholds to determine if it's more efficient to process the request asynchronously
 - Separate conversion thresholds for different request types
- Changes to the thresholds:
 - Reducing a threshold makes it more likely a request of the corresponding type will be processed asynchronously
 - Increasing a threshold makes it more likely the request will complete synchronously
 - For example, you may choose to increase a threshold to try to achieve a greater throughput rate, accepting a greater CPU cost to do so

OA41661 – New XES Sync/Async Support

- New support lets the installation set sync/async conversion thresholds
 - COUPLExx parmlib member supports a new SYNCASYNC statement
 - SIMPLEX - for simplex list and cache requests
 - DUPLEX - for duplexed list and cache requests
 - LOCKSIMPLEX - for simplex lock requests
 - LOCKDUPLEX - for duplexed lock requests
- When modifying the thresholds IBM recommends:
 - Do not depart greatly from the system-determined default values
 - Large variations from the default thresholds in either direction can significantly affect the performance of coupling facility requests
 - Can determine the current default values from the output of the DISPLAY XCF,COUPLE command
- PTFs provided for z/OS 1.9 and above

z/OS 2.1 – CPU MF Summary

- New HISSERV Service
 - New Programming Interface for real time CPU MF access
- ZOS SOFTWARE Counters
 - New Counter Set (in addition to Basic, Problem, Crypto, and Extended)
 - Only recorded in SMF 113 subtype 1
- Command Changes to improve operability and data recording
- SMF 113 Record Changes
 - SMF 113 Subtype 2
 - Interval Start and End Time
 - Machine Sequence Code (e.g. Processor Serial Number)
- SMF 113 Subtype 1 – New
 - Same as SMF 113 Subtype 2 but **only Delta values**
 - Includes ZOS Counter Set

Recommendation remains to continuously run CPU MF Counters (Basic and Extended) – collecting SMF 113 Subtype 2 records

“F HIS,B,TT='Text',PATH='/his',CTRONLY,CTR=(B,E),SI=SYNC”

Looking for zEC12 / zBC12 Migration “Volunteers” to send SMF data

- Want to validate / refine Workload selection metrics

Looking for “Volunteers”

(3 days, 24 hours/day, SMF 70s, 72s, 113s per LPAR)

“Before” and “After”

Production partitions preferred

If interested send note to jpburg@us.ibm.com,

No deliverable will be returned

Benefit: Opportunity to ensure your data is used to influence analysis

GRS Support to monitor GQSCAN and ISGQUERY

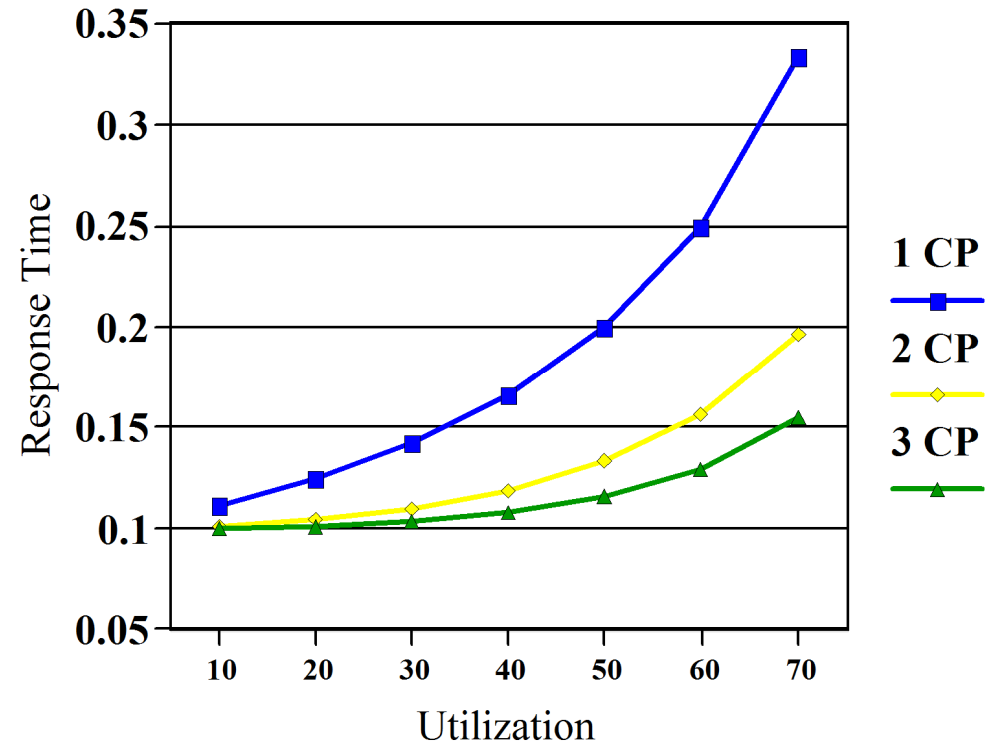
- New Function APAR OA42221
 - Supports z/OS 1.13 and above
 - New SMF 87 record provides monitoring of global, generic queue scans to help identify programs that issue global generic queue scans
 - GQSCAN and ISGQUERY REQINFO=QSCAN services can cause spikes in GRS CPU usage and GRS private storage when invoked many times for global, generic queue scans
 - Potential CPU Impact is greater in GRS STAR mode
 - SMF 87 records contain information about the caller such as TCB, ASID, PSW, and details about the queue scan service invocation
 - Setup:
 - The SMF 87 records are mapped by macro ISGYSMFR
 - Specify MONITOR(YES) in your GRSCNFxx PARMLIB member, or issue the SETGRS MONITOR=YES command

z/OS 2.1 UNIX I/O Count Enhancements

- Provide a better way of knowing which files or directories are heavily accessed
 - Current SMF 92.11 (close) records tend to produce too much output and do not give a good indication of which files are heavily accessed
- New support:
 - New SMF 92.16 record created to contain sockets and character special files (will no longer appear in subtype 11)
 - Clears 92.11 records of “noise” caused by sockets and character special files
 - New SMF 92.17 record created for accesses to regular files and directories
 - Records the amount of times a regular file or directory is accessed and the SMF record is written on two occasions:
 - When the internal representation of the file is deleted
 - At the end of the SMF global recording interval
(Note: count of accesses to the file is cleared after writing the record)
- Update SMFPRMxx to record the desired subtypes

Queuing Impacts of Server Busy

	1 CP	2 CP	3 CP	4 CP
1	0.1010	0.1000	0.1000	0.1000
10	0.1111	0.1010	0.1001	0.1000
20	0.1250	0.1042	0.1010	0.1003
30	0.1429	0.1099	0.1033	0.1013
40	0.1667	0.1190	0.1078	0.1038
50	0.2000	0.1333	0.1158	0.1087
60	0.2500	0.1563	0.1296	0.1179
70	0.3333	0.1961	0.1547	0.1357
80	0.5000	0.2778	0.2079	0.1746
90	1.0000	0.5263	0.3724	0.2969
99	10.0000	5.0251	3.3706	2.5448



- Arrival rates and zIIP busy will influence ‘Needs Help’
- Can run zIIPs very busy IF there are multiple classes of work with different response time objectives, but watch IIPCP time
- Recommendation for online workloads:
 - 1-2 Specialty CPs: Keep zIIPs 40-50% busy
 - >= 3 Specialty CPs: Keep zIIPs 60-70% busy

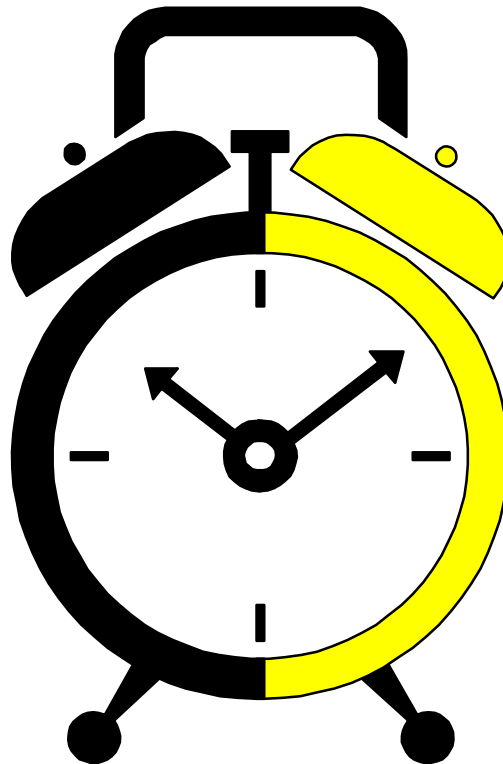
Predicting DB2 10 zIIP Eligibility for Prefetch and Deferred Writes

- In DB2 V9 workloads executing under 'non-preemptable' SRBs in DBM1 include:
 - Prefetch
 - Deferred write
 - Castout
 - Pseudo close
 - Write scheduling
- In DB2 10, prefetch and deferred write are executed under 'preemptable' SRB processing in DBM1
 - These workloads are marked as zIIP eligible
 - Represents the vast majority of the DB2 9 'non-preemptable' SRB processing in DBM1 address space
 - Will depend on workload mix

- Be sure you have sufficient zIIP capacity and DB2 address spaces are appropriately defined to support time sensitive activities
- If you have 5 GCP and 1 zIIP you have 1/5th the opportunity to get dispatched
- May need to plan increased zIIP capacity when moving to DB2 10

Addendum

- Older information which should still be understood, or make you go Hmmm.
- APARs which are still causing issues, even though they are old.



Flash Express Support on zEC12/zBC12

- z/OS V2.1 and z/OS V1.13 with PTFs and the z/OS V1.13 RSM Enablement Offering Web deliverable
 - A zEC12 or zBC12 server with Flash Express PCIe cards
 - Flash Dynamic Reconfiguration and optional PLPA and COMMON Page data sets in enabling PTFs
- z/OS designed to use Flash for:
 - Pageable large (1MB) pages
 - Paging, when performance would be improved vs disk-based paging
 - SVC and Standalone Dump
 - Speculative page-ins to help buffer workload spikes (such as market open)
- Exploiters of fixed large (1MB) pages
 - JAVA 6 SR1 and later and its exploiters (WAS)
 - z/OS R11 and later C/C++ programs using LE
 - z/OS in z/OS R12 and later
 - IBM DB2 10
- Exploiters of pageable large (1MB) pages
 - z/OS 1.13 and z/OS 2.1 Language Environment
 - Maintenance roll-up of IBM 31-bit and 64-bit SDK7 for z/OS Java Technology Edition, V7
 - DB2 11 (planned*)
 - IMS 12 Common Queue Server with APAR PM66866

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

New DB2 / WLM Goal Setting Documentation

- Updated WLM classification information for DB2 started tasks
 - New section published June, 2012

- Recommends:
 - IRLMPROC in SYSSTC
 - ***ssnm*MSTR, *ssnm*DBM1, *ssnm*DIST and WLM-managed stored procedure address spaces** should be in a service class with a high importance and an aggressive velocity goal
 - May need to be marked CPU Critical
 - Recommends LPARs which run DB2 work at low dispatch priorities in environment with high CPU utilization to be more aggressive with blocked workload support
 - Protects environment by ensuring work holding DB2 locks gets CPU service
 - Recommends setting BLWLINTHD in IEAOPTxx from default of 20 to 3-5 secs

HiperDispatch APARs

- OA37736 – New Function
 - WLM uses free processor capacity of the CEC to determine if an LPAR should unpark one or more vertical lows
 - Apportionment didn't take into account relative weight of LPARs on CEC
 - Small LPARs could unpark and take free processor capacity leaving a larger LPAR unable to unpark vertical lows
 - LPARs may then not have proportional access, based on weight, to whitespace
 - Changes:
 - Now unpark an additional vertical low if CPU consumption of the partition is below its current weight **plus** newly introduced unused capacity share
 - Unparking only happens when no free processor capacity and the LPAR has CPU demand
 - SMF 99.12 and 99.14 record mappings will be published
 - 99.12 – Hiperdispatch Interval data
 - 99.14 – Hiperdispatch Topology data

V9 and V10 Sample CPU Times from SMF 100

■ V9 example:

CPU,TIMES	TCB TIME	PREEMPT SRB	NONPREEMPT SRB	TOTAL TIME	PREEMPT IIP SRB	/COMMIT
SYSTEM SERVICES ADDRESS SPACE	6.186504	0.001054	13.465251	19.652809	N/A	0.001272
DATABASE SERVICES ADDRESS SPACE(DBM1)	0.292939	0.013103	41.596518	41.902560	0.000000	0.002712
IRLM	0.000128	0.000000	0.420866	0.420994	N/A	0.000027
DDF ADDRESS SPACE	0.003241	0.000000	0.005809	0.009051	0.000000	0.000001

■ V10 example:

CPU,TIMES	TCB TIME	PREEMPT SRB	NONPREEMPT SRB	TOTAL TIME	PREEMPT IIP SRB	/COMMIT
SYSTEM SERVICES ADDRESS SPACE	5.388198	9.265621	1.226727	15.880545	N/A	0.001083
DATABASE SERVICES ADDRESS SPACE	0.144509	0.838230	0.022309	1.005048	25.877390	0.000069
IRLM	0.000060	0.000000	0.266525	0.266585	N/A	0.000018
DDF ADDRESS SPACE	0.086311	0.000138	0.004407	0.090856	0.000000	0.000006
TOTAL	5.619079	10.103988	1.519967	17.243034	25.877390	0.001176

Example of Potential Offload

- Eight Data Sharing Members during Online Peak

Total DB2 Started Task Avg. GCPs used minus DDF Threads	DBM1 NonPreempt Avg. GCPs	Total DBM1 Avg. GCPs	Estimated Average zIIP CPs Offloaded in V10	Estimated zIIP Offload as % of Total DB2 Started Tasks
1.08	0.49	0.57	0.49	46%
1.56	0.66	0.80	0.66	42%
1.01	0.41	0.50	0.41	41%
0.62	0.52	0.52	0.52	82%
0.30	0.26	0.26	0.26	86%
0.10	0.08	0.08	0.08	77%
1.08	0.51	0.58	0.51	47%
0.84	0.29	0.49	0.29	34%

Workload Promotion – LCK – Local Lcok

- In HiperDispatch, when a work unit that held a local/CML lock is undispached from a CPU:
 - z/OS will temporarily promote it to x'FF' to give it a chance to give up the local/CML lock
 - If the work unit does not give up its lock during the temporary promotion, it will be demoted back to its original dispatch priority.
 - Once it runs for 1 dispatch at its original dispatch priority it is eligible to be promoted again

- Done so work will run at a high priority until it releases its Local/CML lock

- Monitors will not display the WLM designated Dispatch Priority

- May introduce CPU delay to high importance workloads

WORKLOAD ACTIVITY

--PROMOTED--

BLK	0.000
ENQ	0.000
CRM	0.000
LCK	0.275
SUP	0.000

Check and understand why there are CPU times in any service classes

zEC12 Architecture Extensions

- **Transactional Execution (a/k/a Transactional Memory)**
 - Software-defined sequence treated by hardware as atomic “transaction”
 - Enables significantly more efficient software
 - Highly-parallelized applications
 - Speculative code generation
 - Lock elision
- **2 GB page frames**
 - Increased efficiency for DB2 buffer pools, Java heap, others
- **Software directives to improve hardware performance**
 - Data usage intent improves cache management
 - Branch pre-load improves branch prediction effectiveness
 - Block prefetch moves data closer to processor earlier, reducing access latency
- **New Decimal-Floating-Point Zoned-Conversion Facility**
 - Improves performance of applications compiled with the new Enterprise PL/I compiler



12915: z/OS V1.13 - A Performance Update, Tuesday, 9:30 AM-10:30 AM

Transactional Execution Facility – OA38829

- Support for z/OS 1.13 running on a zEC12
- New Facility intended to be used by JAVA 7, installation code, ISV code
- Provides LOADxx external to disable use of transactional execution facility

MACHMIG – Identifies one or more facilities z/OS should not use

Column	Contents
1-7	MACHMIG
10-72	List of facilities NOT to use EDAT2 TX

- Example in LOADxx
MACHMIG TX,EDAT2

z/OS Performance **HOT** Topics

Session: 14022

Kathy Walsh
IBM Corporation

