



Advanced Technical Skills (ATS) North America

# CPU MF - the “Lucky” 113s - z196 Update and WSC Experiences

**SHARE** Session 7717

August 4, 2010

**John Burg**

[jpburg@us.ibm.com](mailto:jpburg@us.ibm.com)

**IBM**



# Trademarks

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

AlphaBlox*	GDPS*	RACF*	Tivoli*
APPN*	HiperSockets	Redbooks*	Tivoli Storage Manager
CICS*	HyperSwap	Resource Link	TotalStorage*
CICS/VSE*	IBM*	RETAIN*	VSE/ESA
Cool Blue	IBM eServer	REXX	VTAM*
DB2*	IBM logo*	RMF	WebSphere*
DFSMS	IMS	S/390*	zEnterprise
DFSMSHsm	Language Environment*	Scalable Architecture for Financial Reporting	xSeries*
DFSMSrmm	Lotus*	Sysplex Timer*	z9*
DirMaint	Large System Performance Reference™ (LSPR™)	Systems Director Active Energy Manager	z10
DRDA*	Multiprise*	System/370	z10 BC
DS6000	MVS	System p*	z10 EC
DS8000	OMEGAMON*	System Storage	z/Architecture*
ECKD	Parallel Sysplex*	System x*	z/OS*
ESCON*	Performance Toolkit for VM	System z	z/VM*
FICON*	PowerPC*	System z9*	z/VSE
FlashCopy*	PR/SM	System z10	zSeries*
	Processor Resource/Systems Manager		

\* 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.

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 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.

## Other Related Presentations

- **zPCR Capacity Sizing Lab - Part 1 Introduction & Overview - Wed 1:30 PM**
- **zPCR Capacity Sizing Lab – Part 2 Hands on Lab – Wed 3 PM**
- **To MIPS or Not to MIPS – Thursday 9:30**
- **Lunch and Learn: [The All New LSPR and z196](#) - Thursday 12:15**
- **Framework For Doing Capacity Sizing on System z – Thursday 1:30 PM**
- **IBM Smart Analytics Optimizer – Thursday 3 PM**

# Topics

- **CPU MF Introduction**
  - What it is, how to enable,
  - New support for z10s and zEnterprise 196 (z196)
    - Including Sync Interval and Identification of Processor Type (e.g. GCP, zAAP, zIIP)
- **Workload Characterization Update**
  - Step 1 completed
- **Key Performance Metrics for z10s and z196s**
  - CPI, Problem State, Cache / Memory Hierarchy
  - New metrics and formulas
- **WSC Customer Experiences with SMF 113s**
  - Lessons Learned Summary
  - HiperDispatch=No/Yes
  - DB2 10 for z/OS Beta 1MB Page Buffer pools
- **Summary**

# CPU Measurement Facility Introduction

# What is the z10 CPU Measurement Facility

- **New hardware instrumentation facility “CPU Measurement Facility” (CPU MF)**
  - Available on System z10 GA2 (EC and BC) and z196
  - Supported by a new z/OS component (Instrumentation), Hardware Instrumentation Services (HIS)
  
- **Potential Uses – for this new “cool” virtualization technology**
  - COUNTERS
    - Supplement Current Performance Metrics
    - Workload characterization
  - SAMPLING
    - ISV product improvement
    - Application Tuning
  
- **IBM Research article**
  - *“IBM System z10 performance improvements with software & hardware synergy”*
  - <http://www.research.ibm.com/journal/rd/531/jackson.pdf>
  - Contact IBM team for copy of the article

# Requirements and Steps to utilize z10 and z196 CPU MF

## ■ Requirements for CPU MF

- z196 or System z10 machine
  - z10 must be at GA2 Driver 76D – Bundle #20 or higher
- z10 z/OS LPAR being measured must be at z/OS 1.8 or higher with APARs:
  - OA25755, OA25750, and OA25773 – also **OA30486 for z/OS 1.10 and higher for new functionality**
  - OA27623 also recommended to add “CPU Speed” to SMF 113s and HIS COUNTERS output
  - Not **currently** supported for z/OS running as a z/VM guest – **z/VM native prototype support in process**
- **z196** z/OS LPARs being measured at z/OS 1.9 or higher require **APAR OA30486**
  - z/OS 1.8 requires OA33052

## ■ Steps to utilize CPU MF

- Configure the z10 or z196 to collect CPU MF Data
  - Update LPAR Security Tabs (See appendix)

**Operationally CPU MF works the same on z196**

**CPU MF has a very low overhead to run, is easy to implement, and is a very small SMF record**

- Configure HIS on z/OS to collect CPU MF Data

- Set up HIS Proc
- Set up OMVS Directory
- Collect SMF 113s via SMFPRMxx

```
//HIS PROC
//HIS EXEC PGM=HISINIT,REGION=0K,TIME=NOLIMIT
//SYSPRINT DD SYSOUT=*
```

- Collect CPU MF Data

- Start HIS – Modify with Begin/End – for COUNTERS or SAMPLING
- “F HIS,B,TT='Text',PATH='/his/',**CTRONLY,CTR=ALL**

**Remember CTR=ALL to get Extended Counters!**

- Analyze the CPU MF Data
  - SMF 113s

## New HIS support for Sync Interval, PU Type and STATECHANGE

- **APAR OA30486 with z/OS 1.12 GA – will be rolled down to z/OS V1R11 and z/OS V1R10**
  - **Applicable for z10s and z196s for new functionality**
  - **New CPU MF capability to sync SMF 113s with other SMF records**
    - SMFINTVAL=SYNC
      - Synchronize records with the SMF global recording interval
    - ..or choose Interval time 1-60
    - **Recommendation is “SYNC”:**
      - “F HIS,B,TT='Text',PATH='/his/',**CTRONLY,CTR=ALL,SMFINTVAL=SYNC**”
  - **Identification of PU Type (GCP, zIIP or zAAP) in SMF 113 record**
    - SMF113\_2\_CpuProcClass '0' - GCP / '2' - zAAP / '4' - zIIP
  - **STATECHANGE**
  - Both SMFINTERVAL and STATECHANGE can be abbreviated, e.g, SI=SYNC, SC=SAVE
    - “F HIS,B,TT='Text',PATH='/his/',**CTRONLY,CTR=ALL,SI=SYNC,SC=SAVE**”
- **In SMF 113s the z196 processor is identified by**
  - SMF113\_2\_CTRVN2 = **'2' for z196**, '1' for z10

Recommend SMFINTVAL=SYNC or  
SI=SYNC

z196 Extended Counters have changed,  
use CTRVN2 to determine if z10 or z196



# HIS STATECHANGE

- **HIS detects and handles significant hardware events (state change)**

- Replacement Capacity (Customer Initiated Upgrade)
- On/Off Capacity on demand

Verify with SMF 113s that “CPU Speed” or “Effective GHz” changed as expected

- **How HIS reacts depends on the STATECHANGE parameter specified**

- STATECHANGE=STOP
  - Stop the collection run when the event was detected
- STATECHANGE=IGNORE
  - Continue the collection run as if the event never happened
- STATECHANGE=SAVE (Default)
  - Record the previous state of the system (Save all data)
    - Write and close the .CNT file
    - Close all .SMP files (1 per CPU)
    - Cut SMF Type 113 Records (1 per CPU)
  - Continue the collection run with the new state
    - Create new .SMP files (1 per CPU)
    - Cut SMF Type 113 Records (1 per CPU)

Recommend STATECHANGE=SAVE, (the default) so don't need to specify

- **STATECHANGE information not directly reported in the SMF 113**

- You will see additional record(s) and an increase/decrease in CPIDs or “CPU Speed”

# New HIS APAR OA30486 support for z196 – WSC Example

15:33:13.24 JPBURG 00000200 F HIS,B,TT='Z196 w/ TEST',CTRONLY,CTR=ALL,SMFINTVAL=SYNC

15:33:14.22 STC01226 00000000 HIS011I HIS DATA COLLECTION STARTED

Time Stamp	CPU #	CpuProcClass	CTNVN1	CTNVN2	CPSP	
10 JUL 22 15:33:14.22	0	0	1	2	5208	← '5.2' GHz z196
10 JUL 22 15:33:14.22	1	0	1	2	5208	
10 JUL 22 15:33:14.22	4	4	1	2	5208	
10 JUL 22 15:33:14.22	5	4	1	2	5208	
10 JUL 22 15:35:00.00	0	'0' GCP	0	2	5208	← '2' z196
10 JUL 22 15:35:00.00	1	→ 0	1	2	5208	
10 JUL 22 15:35:00.00	4	'4'-zIIP → 4	1	2	5208	
10 JUL 22 15:35:00.00	5	4	1	2	5208	

**SMF 113 Synched with SMF Global Recording Interval - 5 Minutes**

# New HIS APAR OA30486 support for z196 – WSC Example

12:19:11.82 JPBURG 00000200 F HIS,B,TT='Z196 TEST', CTRONLY,CTR=ALL,SMFINTVAL=1

12:19:11.82 STC32434 00000000 HIS011I HIS DATA COLLECTION STARTED

Time Stamp	CPU #
10 JUL 16 12:19:11.82	0
10 JUL 16 12:19:11.82	1
10 JUL 16 12:19:11.82	2
10 JUL 16 12:19:11.82	3
10 JUL 16 12:20:11.82	0
10 JUL 16 12:20:11.82	1
10 JUL 16 12:20:11.82	2
10 JUL 16 12:20:11.82	3

**SMF 113 Written every 1 Minute**

# Workload Characterization Update

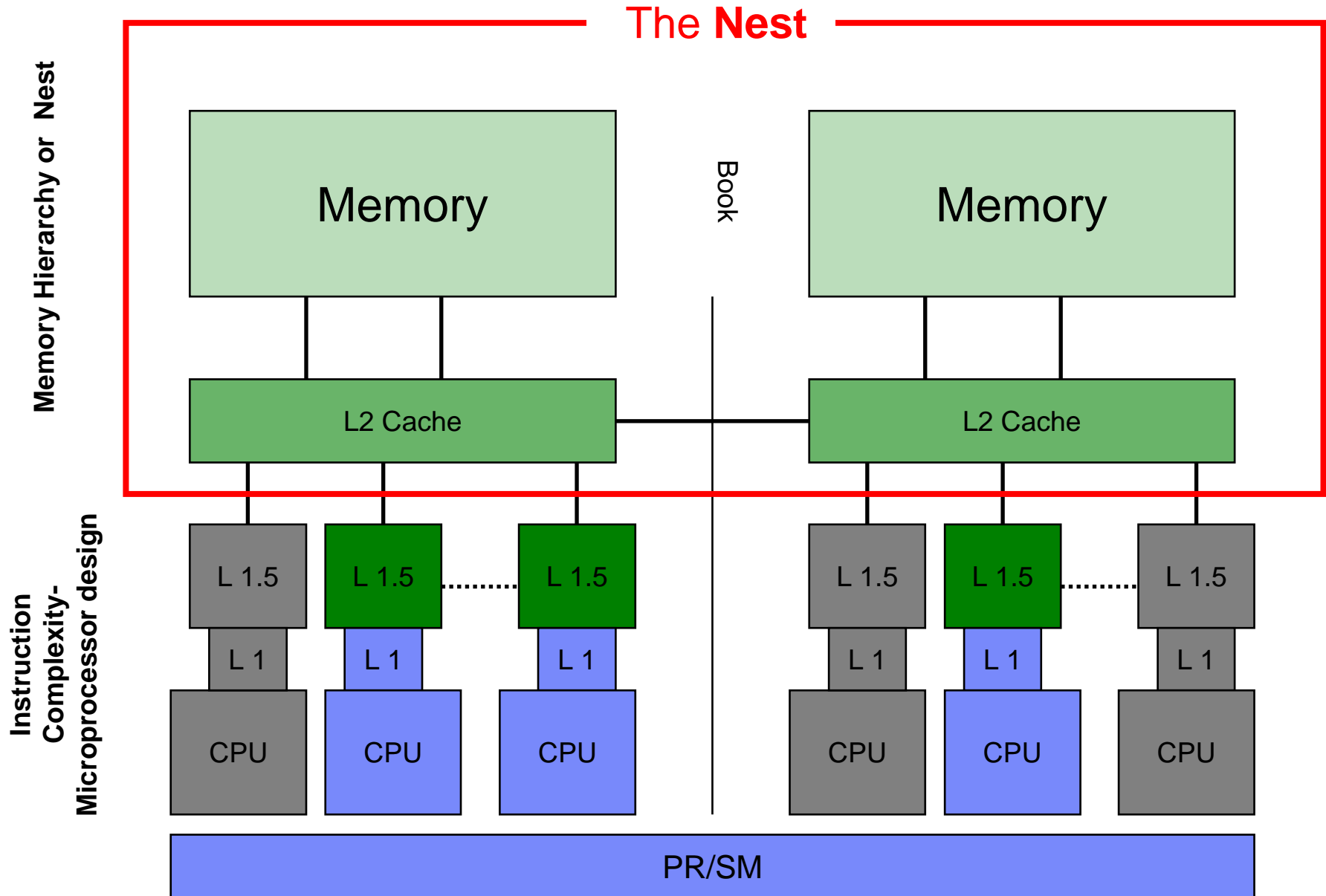
- **Historically, LSPR workload capacity curves (primitives and mixes) have had application names or been identified by a “software” captured characteristic**
  - For example, CICS, IMS, OLTP-T, CB-L, LoLO-mix, TI-mix, etc
- **However, capacity performance is more closely associated with how a workload is using and interacting with a processor “hardware” design**
- **With the availability of CPU MF (SMF 113) data on z10, the ability to gain insight into the interaction of workload and hardware has arrived**
- **The knowledge gained is still evolving, but the first step in the process is to produce LSPR workload capacity curves based on the underlying hardware sensitivities**
- **Thus, the LSPR for z196 will introduce three new workload categories which replace all prior primitives and mixes**
  - Based on new hardware defined metric called [Relative Nest Intensity](#)
  - Low, Average, High (Relative Nest Intensity)
- **To simplify the transition, an easy and automatic translation of old names to new categories will be supplied in zPCR**
  - For example, if you have been using LoLO-mix in your studies, you will simply use the new “Average” workload in the future

## ▪ **Instruction Complexity (Micro processor design)**

- Many design alternatives
  - Cycle time (GHz), instruction architecture, pipeline, superscalar, Out-Of-Order, branch prediction and more
- Workload effect
  - May be different with each processor design
  - But once established for a workload on a processor, doesn't change very much

## ▪ **Memory Hierarchy or “Nest”**

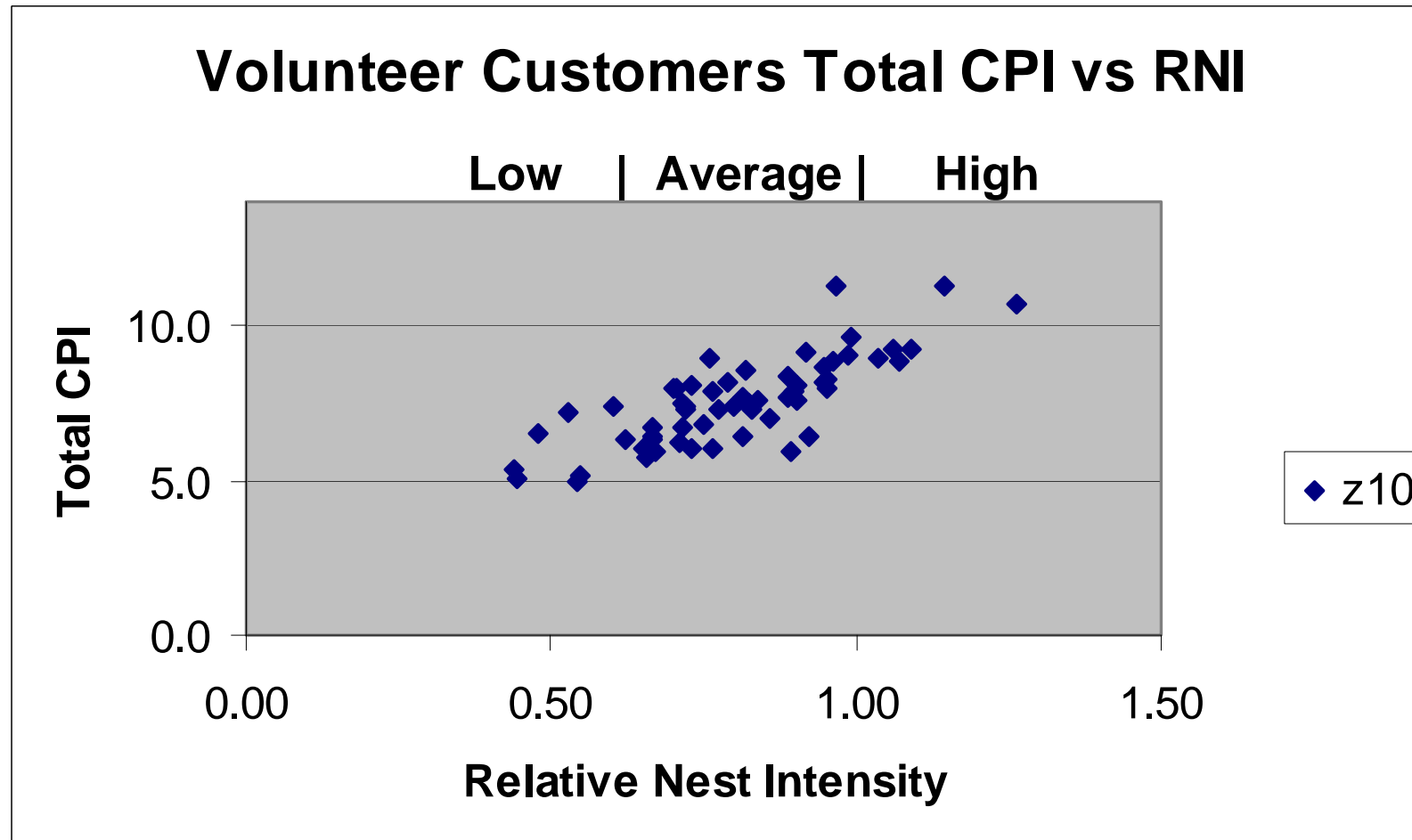
- Many design alternatives
  - Cache (levels, size, private, shared, latency, MESI protocol), controller, data buses
- Workload effect
  - Quite variable
  - Sensitive to many factors: locality of reference, dispatch rate, IO rate, competition with other applications and/or LPARs, and more
  - Net effect of these factors represented in “Relative Nest Intensity”
- Relative Nest Intensity (RNI)
  - Activity beyond private-on-chip cache(s) is the most sensitive area
  - Reflects distribution and latency of sourcing from shared caches and memory
  - Level 1 cache miss percentage also important
  - Data for calculation available from CPU MF (SMF 113) starting with z10



# CPU MF

## z10 Customer Workload Characterization Summary

---





# Workload Characterization Future Vision – Step 1 is Complete

- **Future vision to help identify workload characteristics and to provide better input for capacity planning and performance**
  - **Step 1 – Created Workload Categories from SMF 113s - complete**
    - **Over 150 z10 Customer/Partitions have participated thru 8/1. Thank You!**
    - Measured LSPR with these new Categories
  - Step 2 – Refine Workload Selection Process
    - **As you move to z196 from z10, looking for “Before” and “ After volunteers**

Still Looking for “Volunteers” – (3 days, 24 hours/day, SMF 70s, 71s, 72s, 113s per LPAR)  
“Before z10” and “After z196”

If interested send note to [jpburg@us.ibm.com](mailto:jpburg@us.ibm.com), No deliverable will be returned

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

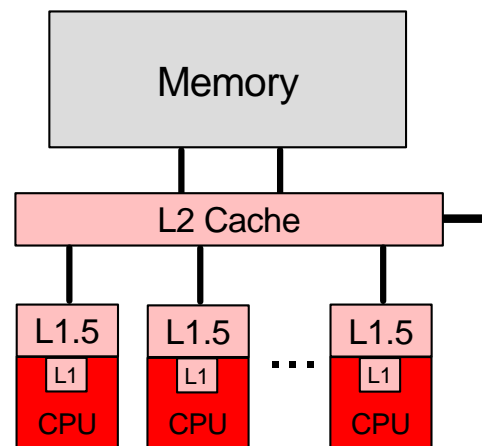
- What is CPU MF?
  - A new z10 and z196 capability to measure cache / memory hierarchy characteristics
- How can it be used today?
  - To supplement current performance metrics (e.g. from SMF, RMF, DB2, CICS)
  - As a secondary data source to understand **why** performance may have changed
- What can it be used for in z196 capacity planning?
  - Capacity Sizing process is the same as today with zPCR
    - Based on DASD I/Os per MSU consumed
    - And optionally use a new Relative Nest Intensity “Hint”
    - CP3KEXTR will process SMF 113s and include in EDF file for zPCR
  - SMF 113 data may prove useful in support of an installation of z196
- What CPU MF is not
  - It is **Not** a substitute for traditional performance nor capacity metrics
  - It does **Not** indicate the capacity being achieved by the LPAR or processor
- Recommend Enabling CPU MF COUNTERS on key z10 production partitions
  - See ***CPU MF Overview and WSC Experiences Techdoc TC000041***
    - <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TC000041>
    - Overview presentation and a white paper on how to enable CPU MF COUNTERS

# Key Performance Metrics for z10s and z196s

## z196 versus z10 hardware comparison

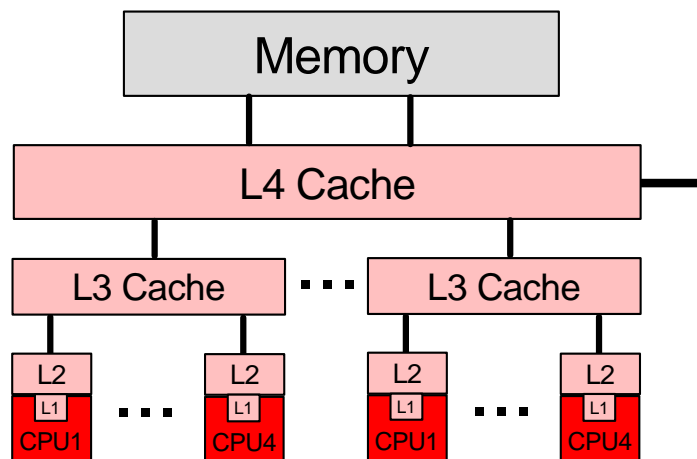
### ■ z10 EC

- ▶ CPU
  - 4.4 GHz
- ▶ Caches
  - L1 private 64k i, 128k d
  - L1.5 private 3 MB
  - L2 shared 48 MB / book
  - book interconnect: star

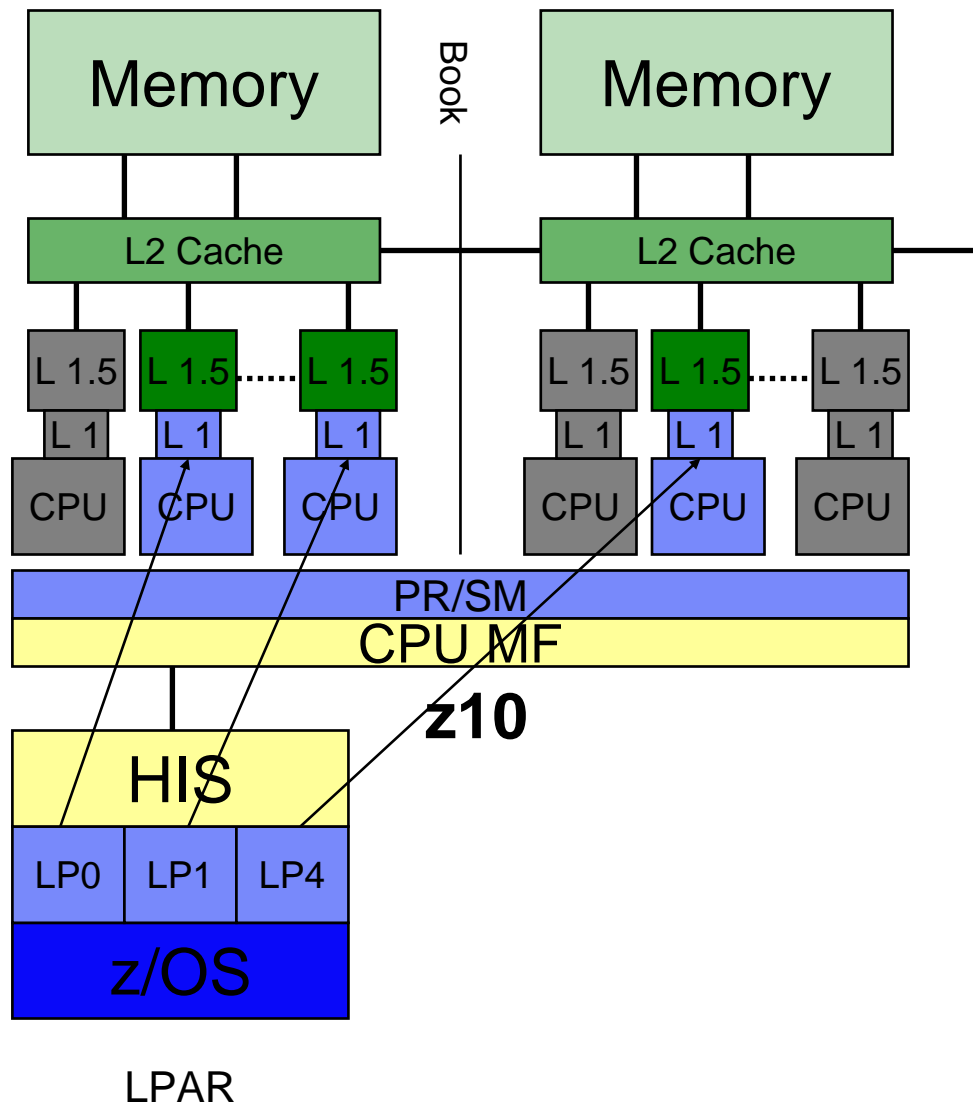


### ■ z196

- ▶ CPU
  - 5.2 GHz
  - Out-Of-Order execution
- ▶ Caches
  - L1 private 64k i, 128k d
  - L2 private 1.5 MB
  - L3 shared 24 MB / chip
  - L4 shared 192 MB / book
  - book interconnect: star



# CPU MF and HIS provide a z/OS logical view Resource Usage and Cache Hierarchy Sourcing



## LPAR / Logical CP view:

- Memory Accesses
- Cache
  - L 2 / (L4 z196) Accesses (local and remote)
  - L3 Accesses on z196
  - L1.5 / (L2 z196) Accesses
  - L1 Sourced from Hierarchy
- Instructions and Cycles
- Crypto function

# Current CPU MF Key Performance Metrics:

CPI	PRBSTATE	L1MP	L15P	L2LP	L2RP	MEMP	LPARCPU
-----	----------	------	------	------	------	------	---------

**CPI – Cycles per Instruction**

**PRBSTATE - % Problem State**

**L1MP – Level 1 Miss %**

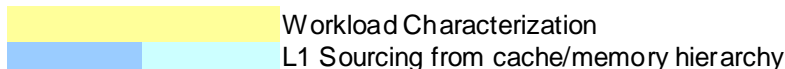
**L15P – % sourced from L1.5 cache**

**L2LP – % sourced from Level 2 Local cache (on same book)**

**L2RP – % sourced from Level 2 Remote cache (on different book)**

**MEMP - % sourced from Memory**

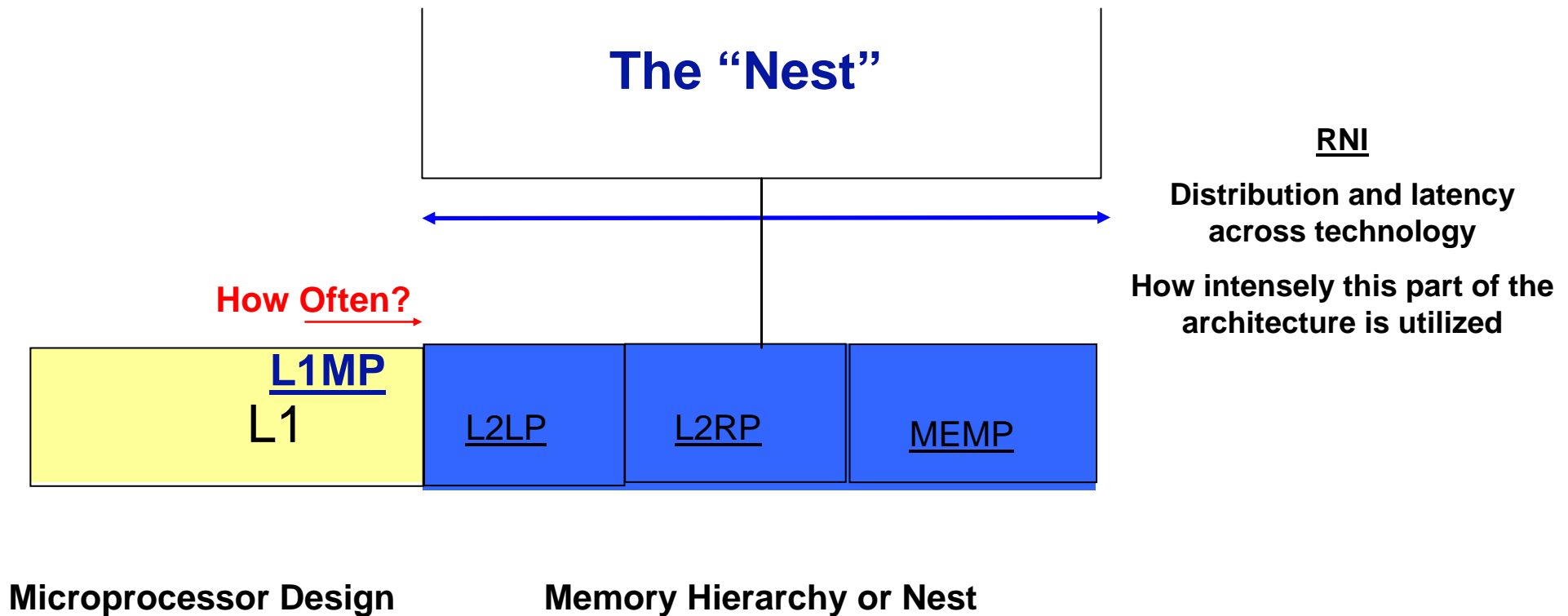
**LPARCPU - APPL% (GCPs, zAAPs, zIIPs) captured and uncaptured**



# Introducing the **new** Relative Nest Intensity (RNI) metric

- Relative Nest Intensity reflects the distribution and latency of sourcing from shared caches and memory
  - For z10 Technology the Relative Nest Intensity =  $(L2LP * 1 + L2RP * 2.4 + MEMP * 7.5) / 100$

## Relative Nest Intensity







# Updated z10 CPU MF Workload Characterization Summary

Customer	SYSID	MON	DAY	CPI	PRBSTATE	Est Instr Cmplx	Est Finite CPI	Est SCPL1M	L1MP	L15P		L2LP	L2RP	MEMP	Rel Nest Intensity	LPARCPU	Eff GHz
All Volunteers		Minimum		3.1	1.1	2.1	0.9	59.6	1.3	48.6		5.6	0.0	2.2	0.4	14.4	
All Volunteers		Average		<b>7.2</b>	<b>31.2</b>	<b>3.2</b>	<b>3.9</b>	<b>101.4</b>	<b>3.9</b>	<b>68.9</b>		<b>21.2</b>	<b>1.6</b>	<b>8.3</b>	<b>0.9</b>	<b>376.3</b>	
All Volunteers		Maximum		12.0	67.1	5.6	8.6	194.9	6.9	82.8		32.9	6.9	20.2	1.8	1442.3	4.40

## New z10 columns are

1. **Est Instr Cmplx CPI**
2. **Est Finite CPI**
3. **Est SCPL1M**
4. **Rel Nest Intensity**
5. **Eff GHz**

CPI – Cycles per Instruction

Prb State - % Problem State

Est Instr Cmplx CPI – Estimated Instruction Complexity CPI (infinite L1)

Est Finite CPI – Estimated CPI from Finite cache/memory

Est SCPL1M – Estimated Sourcing Cycles per Level 1 Miss

L1MP – Level 1 Miss %

L15P – % sourced from Level 2 cache

L2LP – % sourced from Level 2 Local cache (on same book)

L2RP – % sourced from Level 2 Remote cache (on different book)

MEMP - % sourced from Memory

Rel Nest Intensity – Reflects distribution and latency of sourcing from shared caches and memory

LPARCPU - APPL% (GCPs, zAAPs, zIIPs) captured and uncaptured


Eff GHz – Effective gigahertz for GCPs, cycles per nanosecond

Workload Characterization

L1 Sourcing from cache/memory hierarchy

# WSC z196 Sample CPU MF from July – 5 Minute Synched Intervals

**z196**



SYSID	Mon	Day	SH	Hour	CPI	Prb State	Est Instr Cmplx CPI	Est Finite CPI	Est SCPL1M	L1MP	L2P	L3P	L4LP	L4RP	MEMP	Rel Nest Intensity	LPARCPU	Eff GHz
SYSD	JUL	22	N	17.25	3.65	2.3	2.70	0.95	26	3.7	77.8	20.5	0.9	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.33	3.68	2.3	2.73	0.95	26	3.6	77.4	20.8	0.9	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.42	3.67	2.3	2.72	0.95	26	3.7	78.0	20.3	0.9	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.50	3.64	2.3	2.71	0.93	26	3.6	77.8	20.5	0.9	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.58	3.66	2.3	2.72	0.94	26	3.6	77.9	20.4	0.8	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.67	3.65	2.3	2.72	0.94	26	3.6	77.0	21.1	0.9	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.75	3.66	2.3	2.72	0.94	26	3.6	77.4	20.8	0.9	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.83	3.64	2.3	2.70	0.94	26	3.6	77.1	21.0	0.9	0.2	0.7	0.24	0.8	5.2
SYSD	JUL	22	N	17.92	2.78	49.2	2.06	0.72	34	2.1	76.7	18.3	1.8	1.4	1.9	0.42	1.5	5.2
SYSD	JUL	22	N	18.00	3.65	3.2	2.71	0.94	26	3.6	77.0	21.1	1.0	0.2	0.7	0.25	0.8	5.2
SYSD	JUL	22	N	18.08	5.00	0.8	3.46	1.53	27	5.7	86.1	11.9	0.3	0.1	1.7	0.28	9.7	5.2
SYSD	JUL	22	N	18.17	3.72	3.2	2.76	0.96	27	3.6	76.8	21.0	1.1	0.2	0.8	0.26	0.9	5.2
SYSD	JUL	22	N	18.25	3.82	3.7	2.76	1.06	28	3.7	77.4	19.8	1.2	0.6	1.1	0.30	0.9	5.2

CPI – Cycles per Instruction

Prb State - % Problem State

Est Instr Cmplx CPI – Estimated Instruction Complexity CPI (infinite L1)

Est Finite CPI – Estimated CPI from Finite cache/memory

Est SCPL1M – Estimated Sourcing Cycles per Level 1 Miss

L1MP – Level 1 Miss %

L2P – % sourced from Level 2 cache

L3P – % sourced from Level 3 on same Chip cache

L4LP – % sourced from Level 4 Local cache (on same book)

L4RP – % sourced from Level 4 Remote cache (on different book)

MEMP - % sourced from Memory

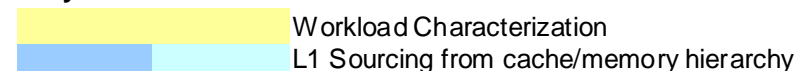
Rel Nest Intensity – Reflects distribution and latency of sourcing from shared caches and memory

LPARCPU - APPL% (GCPs, zAAPs, zIIPs) captured and uncaptured

Eff GHz – Effective gigahertz for GCPs, cycles per nanosecond

**CPU MF provides  
measurement of the z196  
Level 3 shared cache**

**These numbers come from a  
synthetic Benchmark and do not  
represent a production workload**



# Formulas – z10

Workload Characterization  
L1 Sourcing from cache/memory hierarchy

Metric	Calculation – <i>note all fields are <b>deltas</b> between intervals</i>
CPI	$B0 / B1$
PRBSTATE	$(P33 / B1) * 100$
L1MP	$((B2+B4) / B1) * 100$
L15P	$((E128+E129) / (B2+B4)) * 100$
L2LP	$((E130+E131) / (B2+B4)) * 100$
L2RP	$((E132+E133) / (B2+B4)) * 100$
MEMP	$((E134+E135) + (B2+B4-E128-E129-E130-E131-E132-E133-E134-E135)) / (B2+B4)) * 100$
LPARCPU	$( ((1/CPSP/1,000,000) * B0) / \text{Interval in Seconds} ) * 100$

CPI – Cycles per Instruction

PRBSTATE - % Problem State

L1MP – Level 1 Miss %

L15P – % sourced from L1.5 cache

L2LP – % sourced from Level 2 Local cache (on same book)

L2RP – % sourced from Level 2 Remote cache (on different book)

MEMP - % sourced from Memory

LPARCPU - APPL% (GCPs, zAAPs, zIIPs) captured and uncaptured

B\* - Basic Counter Set - Counter Number

P\* - Problem-State Counter Set - Counter Number

See “The Set-Program-Parameter and CPU-Measurement Facilities”  
SA23-2260-0 for full description

E\* - Extended Counters - Counter Number

See “IBM The CPU-Measurement Facility Extended Counters Definition  
for z10” SA23-2261-0 for full description

CPSP - SMF113\_2\_CPSP “CPU Speed”

# Formulas – z10 Additional

Metric	Calculation – <i>note all fields are <b>deltas</b> between intervals</i>
Est Instr Cmplx CPI	CPI – Estimated Finite CPI
Est Finite CPI	$((B3+B5) / B1) * .84$
Est SCPL1M	$((B3+B5) / (B2+B4)) * .84$
Rel Nest Intensity	$(1.0*L2LP + 2.4*L2RP + 7.5*MEMP) / 100$
Eff GHz	CPSP / 1000

Note these Formulas may change in the future

Est Instr Cmplx CPI – Estimated Instruction Complexity CPI (infinite L1)

Est Finite CPI – Estimated CPI from Finite cache/memory

Est SCPL1M – Estimated Sourcing Cycles per Level 1 Miss

Rel Nest Intensity – Reflects distribution and latency of sourcing from shared caches and memory

Eff GHz – Effective gigahertz for GCPs, cycles per nanosecond

B\* - Basic Counter Set - Counter Number

P\* - Problem-State Counter Set - Counter Number

See “The Set-Program-Parameter and CPU-Measurement Facilities”  
SA23-2260-0 for full description

CPSP - SMF113\_2\_CPSP “CPU Speed”



# Formulas – z196

Workload Characterization

L1 Sourcing from cache/memory hierarchy

Metric	Calculation – <i>note all fields are <b>deltas</b> between intervals</i>
CPI	$B0 / B1$
PRBSTATE	$(P33 / B1) * 100$
L1MP	$((B2+B4) / B1) * 100$
L2P	$((E128+E129) / (B2+B4)) * 100$
L3P	$((E150+E153) / (B2+B4)) * 100$
L4LP	$((E135+E136+E152+E155) / (B2+B4)) * 100$
L4RP	$((E138+E139+E134+E143) / (B2+B4)) * 100$
MEMP	$((E141+E142) + (B2+B4-E128-E129-E150-E153-E135-E136-E152-E155-E138-E139-E134-E143-E141-E142)) / (B2+B4)) * 100$
LPARCPU	$((1/CPSP/1,000,000) * B0) / \text{Interval in Seconds} * 100$

CPI – Cycles per Instruction

Prb State - % Problem State

L1MP – Level 1 Miss %

L2P – % sourced from Level 2 cache

L3P – % sourced from Level 3 on same Chip cache

L4LP – % sourced from Level 4 Local cache (on same book)

L4RP – % sourced from Level 4 Remote cache (on different book)

MEMP - % sourced from Memory

LPARCPU - APPL% (GCPs, zAAPs, zIIPs) captured and uncaptured

B\* - Basic Counter Set - Counter Number

P\* - Problem-State Counter Set - Counter Number

See “The Set-Program-Parameter and CPU-Measurement Facilities”  
SA23-2260-0 for full description

E\* - Extended Counters - Counter Number

See expected “The CPU-Measurement Facility Extended Counters  
Definition for z10 and z196” SA23-2261-01 for full description

CPSP - SMF113\_2\_CPSP “CPU Speed”

# Formulas – z196 Additional

Metric	Calculation – <i>note all fields are <b>deltas</b> between intervals</i>
Est Instr Cmplx CPI	CPI – Estimated Finite CPI
Est Finite CPI	$((B3+B5) / B1) * .63$
Est SCPL1M	$((B3+B5) / (B2+B4)) * .63$
Rel Nest Intensity	$1.6*(0.4*L3P + 1.0*L4LP + 2.4*L4RP + 7.5*MEMP) / 100$
Eff GHz	CPSP / 1000

Note these Formulas may change in the future

Est Instr Cmplx CPI – Estimated Instruction Complexity CPI (infinite L1)

Est Finite CPI – Estimated CPI from Finite cache/memory

Est SCPL1M – Estimated Sourcing Cycles per Level 1 Miss

Rel Nest Intensity – Reflects distribution and latency of sourcing from shared caches and memory

Eff GHz – Effective gigahertz for GCPs, cycles per nanosecond  
Workload Characterization

L1 Sourcing from cache/memory hierarchy

B\* - Basic Counter Set - Counter Number

P\* - Problem-State Counter Set - Counter Number

See “The Set-Program-Parameter and CPU-Measurement Facilities”  
SA23-2260-0 for full description

CPSP - SMF113\_2\_CPSP “CPU Speed”

## WSC Experiences

Lessons Learned since March 2010

Customer HiperDispatch Measurement

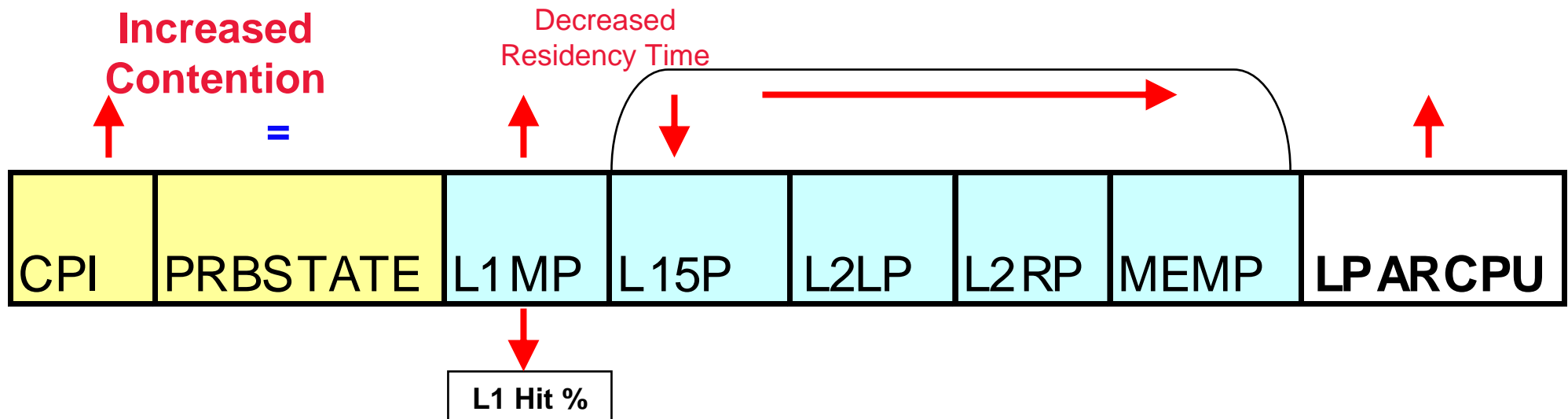
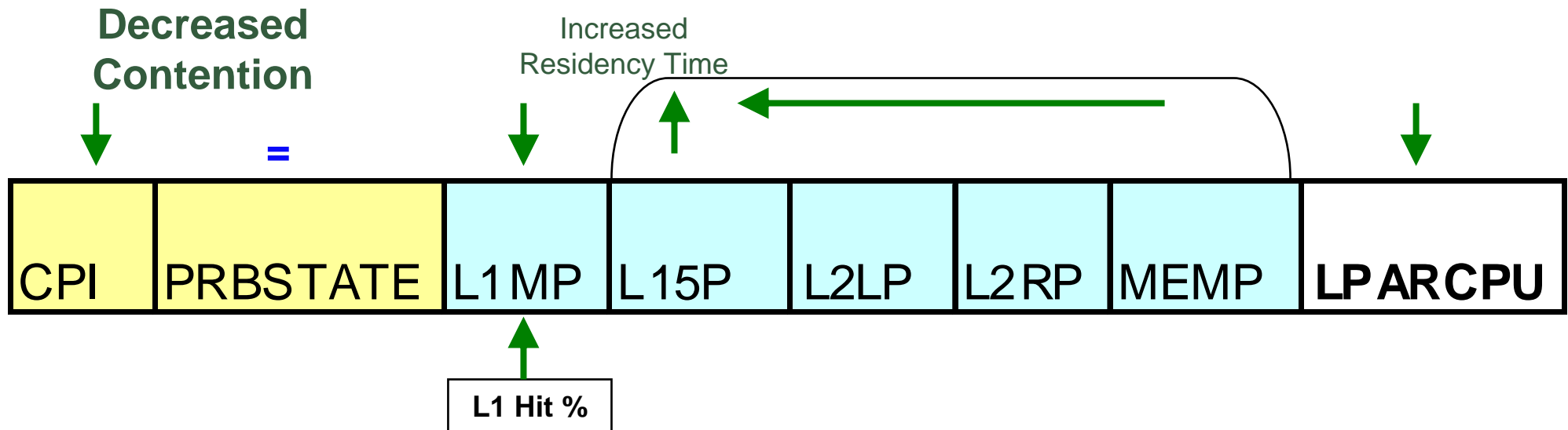
Customer 1 MB Page Measurement

# CPU MF – Lessons Learned since March 2010

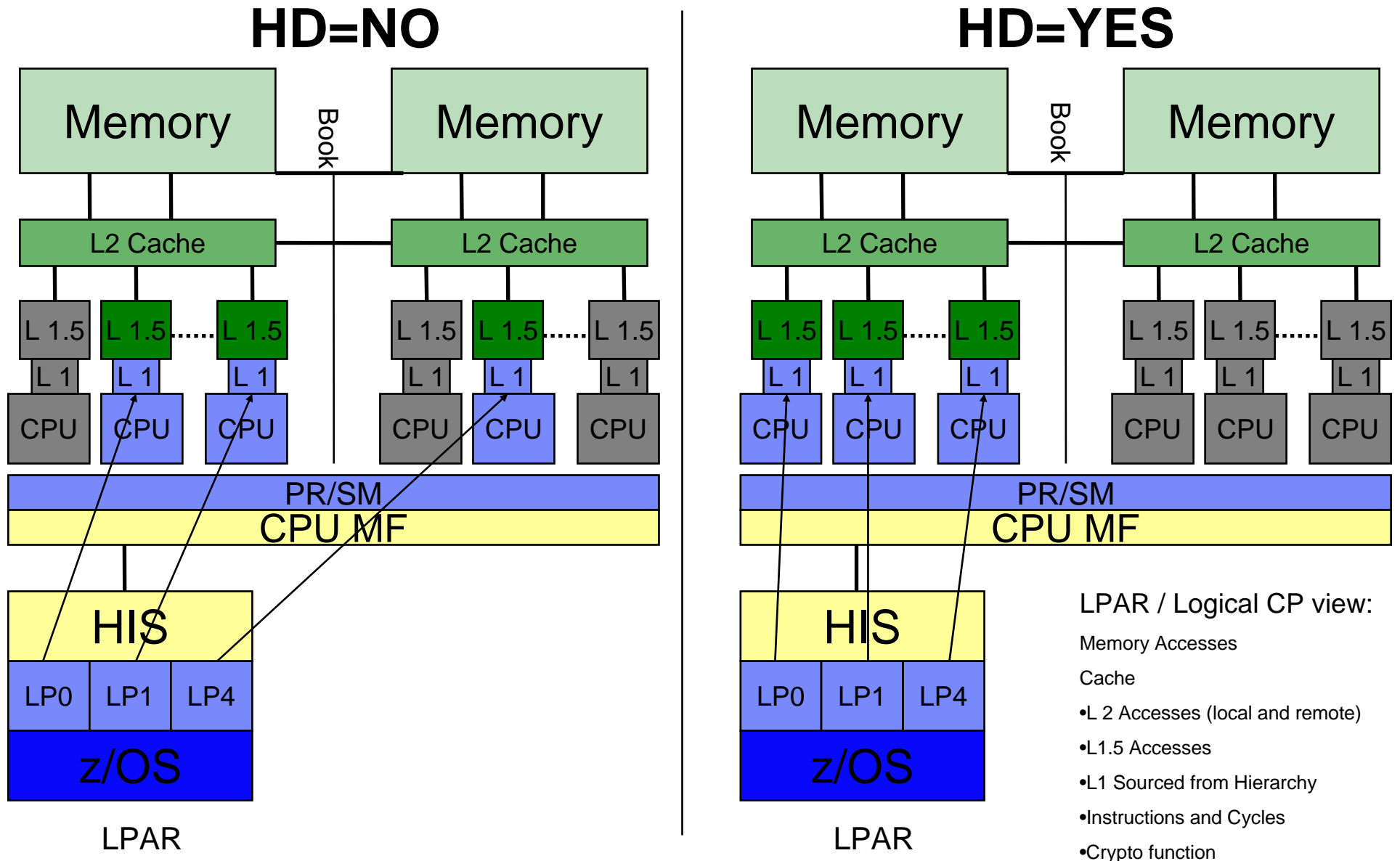
- **CPU MF Performance Metrics continues to help understand why performance changed**
  - LPAR Configuration Changes including
    - HD= Yes/No
  - 1 MB Vs 4k Pages
  - GHz measurement for State Changes
- **Customers continue to successfully run CPU MF COUNTERS collecting SMF 113s**
  - Over days/months without any reported performance impact, Turning on and leaving on
  - Volunteer Feedback: easy to enable, minimal time investment
- **SMF 113 Logical CPU IDs are equal to the SMF 70 Logical CPU IDs**
  - Directly identifies GCPs, zIIPs or zAAPs in SMF 113s with [APAR OA30486](#) for z10s and z196
- **LPAR Management Time is NOT included in LPARCPU time (SMF 113 Cycles)**
- **Utilize the Counter Version Number fields to map to technology**
  - SMF113\_2\_CTRVN2 – Crypto or Extended counter sets = “2” for z196 “1” for z10
- **z/VM CPU MF – native prototype in process**



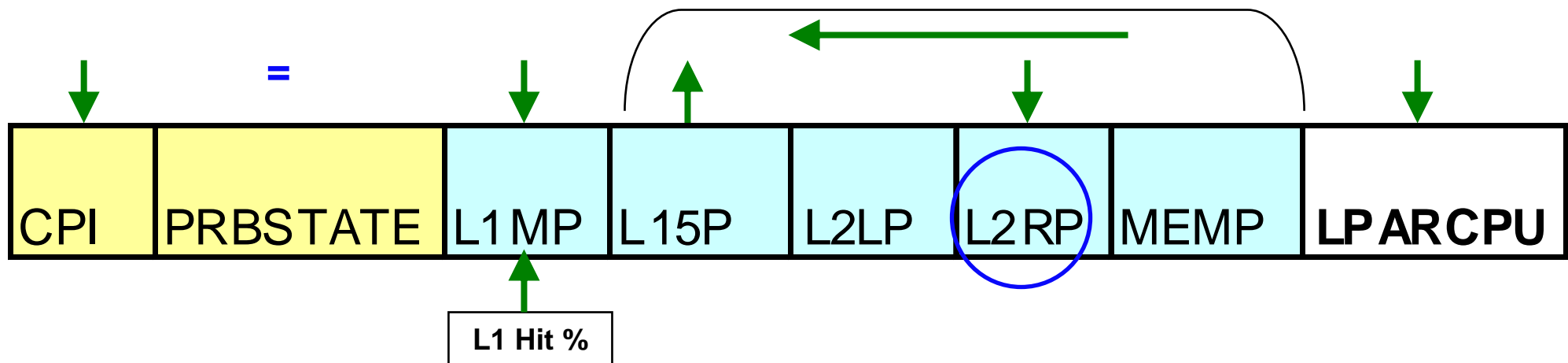
CPU MF can help provide cache/memory resource change insights



HiperDispatch attempts to align Logical CPs with PUs in the same Book



From CPU MF, HiperDispatch=YES May Decrease the L2 Remote %



CPI – Cycles per Instruction

PRBSTATE - % Problem State

L1MP – Level 1 Miss %

L15P – % sourced from L1.5 cache

L2LP – % sourced from Level 2 Local cache (on same book)

L2RP – % sourced from Level 2 Remote cache (on different book)

MEMP - % sourced from Memory

LPARCPU - APPL% (GCPs, zAAPs, zIIPs) captured and uncaptured

Potential Workload Characterization  
z10 L1 sourcing from cache/memory hierarchy

# HiperDispatch=Yes Customer Improvement on z10 721

		Day	Hour	CPI	Prb State	Est Instr Cmplx CPI	Est Finite CPI	Est SCPL1M	L1MP	L15P	L2LP	L2RP	MEMP	Rel Nest Intensity	LPARCPU	HD ?
		12	11.0	8.2	53.7	3.79	4.45	115	3.9	63.7	23.1	6.7	6.6	0.89	1745.1	No
		11	11.0	7.5	52.8	3.74	3.71	97	3.8	70.3	19.8	3.7	6.3	0.76	1632.3	Yes
HD=Yes % Improvement				1.10	1.02	1.01	1.20	1.19	1.01	0.91	1.17	1.80	1.05	1.17	1.07	

CPI – Cycles per Instruction

Prb State - % Problem State

Est Instr Cmplx CPI – Estimated Instruction Complexity CPI (infinite L1)

Est Finite CPI – Estimated CPI from Finite cache/memory

Est SCPL1M – Estimated Sourcing Cycles per Level 1 Miss

L1MP – Level 1 Miss %

L15P – % sourced from Level 2 cache

L2LP – % sourced from Level 2 Local cache (on same book)

L2RP – % sourced from Level 2 Remote cache (on different book)

MEMP - % sourced from Memory

Rel Nest Intensity – Reflects distribution and latency of sourcing from shared caches and memory

LPARCPU - APPL% (GCPs, zAAPs, zIIPs) captured and uncaptured

**HiperDispatch=YES resulted in a ~10% improvement as measured by CPU MF.**

Additional measurements over multiple days from traditional CPU/Transaction metrics should be used to validate HD=No Vs. Yes results

- **Partition has 21 logical processors**
- **2 additional partitions on the CEC**

\*\*\* **New - This is an evolving use of CPU MF** \*\*\*

## CPU MF can help measure the impact of 1 MB Pages in your environment

Test	CPI	PRBSTATE	Est Instr Cmplx	Est Finite CPI	Est SCPL1M	L1MP	L15P	L2LP	L2RP	MEMP	Rel Nest Intensity	LPARCPU	GHz	TLB1 Miss CPU% of Total CPU	TLB1 Cycles per Miss	PTE% of all TLB1 Misses
DB2 V10 4K PageFix=YES	4.46	1.29	2.63	1.83	26	7.13	94.72	4.64	0.01	0.63	0.09	28.2	4.4	16.0	83	19.2
DB2 V10 1MB PageFix=YES	4.26	1.13	2.58	1.68	23	7.25	96.56	3.03	0.01	0.41	0.06	33.9	4.4	15.6	65	13.7
	1.05					0.98	0.98	1.53						1.03	1.28	1.40

- **DB2 10 for z/OS Beta provides ability to specify 1 MB Pages for DB2 Buffer Pools**
- **1 MB Pages can help reduce TLB Page Table Entry Misses**
- **CPU MF can be used to help measure the 1 MB Page impact for your environment**
  - DB2 10 for z/OS Beta Customer ran DB2 Batch job that exercised 4k and 1MB pages (PageFix=Yes). LFArea=40M
    - The batch job executed 30M Selects, 20M Inserts, and 10M Fetchs
  - *CPU MF showed the following – but this is not necessarily representative of 1 MB Page results*
    - 40% reduction in Page Table Entry % (PTE) of all TLB1 Misses
    - 28% reduction TLB1 Cycles per Miss, 3% reduction TLB1 Miss CPU% of Total CPU
    - Lower CPI and Nest Intensity
    - DB2 Accounting report showed 1.4 % reduction in CPU time
- **As you implement 1 MB Page exploiters, use CPU MF to help measure the impact**
  - **Measure it in its intended Production LPAR**
- **See white paper “IBM System z10 Support for large pages”**
  - <http://www.research.ibm.com/journal/abstracts/rd/531/tzortzatos.html>

**Warning: These numbers come from a synthetic Benchmark and do not represent a production workload**

# DB2 10 for z/OS Beta Customer – RMF for 1 MB Page

P A G I N G   A C T I V I T Y									
z/OS V1R10		SYSTEM ID SYSA		START 06/30/2010-12.45.00		INTERVAL 000.15.00		PAGE 2	
CONVERTED TO z/OS V1R11 RMF		END 06/30/2010-13.00.01		CYCLE 5.000 SECONDS					
DPT = IEAOPTXX		MODE = ESAME		CENTRAL STORAGE MOVEMENT RATES - IN PAGES PER SECOND					
-----									
HIGH UIC (AVG) = 65535 (MAX) = 65535 (MIN) = 65535									
WRITTEN TO		READ FROM		*--- CENTRAL STORAGE FRAME COUNTS ---*					
CENTRAL STOR		CENTRAL STOR		MIN		MAX		AVG	
HIPERSPACE		0.00		1		1		1	
PAGES									
VIO		0.00		0		0		0	
PAGES									
-----									
FRAME AND SLOT COUNTS									
-----									
CENTRAL STORAGE				LOCAL PAGE DATA SET SLOT COUNTS					
		MIN	MAX	AVG			MIN	MAX	AVG
(15 SAMPLES)									
AVAILABLE		158,574	161,884	159,692	AVAILABLE SLOTS		2,854,758	2,854,758	2,854,758
SQA		10,497	10,595	10,529	VIO SLOTS		0	0	0
LPA		5,734	5,735	5,734	NON-VIO SLOTS		757	757	757
CSA		39,739	39,921	39,850	BAD SLOTS		0	0	0
LSQA		15,186	15,198	15,193	TOTAL SLOTS		2,855,515	2,855,515	2,855,515
REGIONS+SWA		539,686	542,913	541,822	SHARED FRAMES AND SLOTS				
TOTAL FRAMES		786,432	786,432	786,432					
FIXED FRAMES									
NUCLEUS		2,608	2,608	2,608	CENTRAL STORAGE		6,428	6,557	6,489
SQA		9,636	9,733	9,668	FIXED TOTAL		98	98	98
LPA		94	94	94	FIXED BELOW 16 M		0	0	0
CSA		1,550	1,550	1,550	AUXILIARY SLOTS		0	0	0
LSQA		14,324	14,334	14,331	TOTAL		8,389	8,518	8,450
REGIONS+SWA		49,347	49,392	49,359	MEMORY OBJECTS AND FRAMES				
BELOW 16 MEG		77	77	77					
BETWEEN 16M-2G		13,456	13,498	13,467	OBJECTS COMMON				
TOTAL FRAMES		77,568	77,680	77,611	SHARED				
STORAGE REQUEST RATES									
GETMAIN REQ		0		LARGE					
FRAMES BACKED		0		FRAMES COMMON					
FIX REQ < 2 GB		0		COMMON FIXED					
FRAMES < 2 GB		0		SHARED					
REF FAULTS 1ST		0		1 MB					
NON-1ST		0							

# Formulas – Additional TLB

<b>Metric – z10</b>	<b>Calculation – <i>note all fields are <b>deltas</b> between intervals</i></b>
TLB1 CPU Miss % of Total CPU	$( (E145+E146) / B0 ) * 100$
TLB1 Cycles per TLB Miss	$(E145+E146) / (E138+E139)$
PTE % of all TLB1 Misses	$(E140 / (E138+E139) ) * 100$

<b>Metric – z196</b>	<b>Calculation – <i>note all fields are <b>deltas</b> between intervals</i></b>
TLB1 CPU Miss % of Total CPU	$( (E130+E131) / B0 ) * 100$
TLB1 Cycles per TLB Miss	$(E130+E131) / (E144+E145)$
PTE % of all TLB1 Misses	$(E146 / (E144+E145) ) * 100$

**Note these Formulas may change in the future**

TLB1 CPU Miss % of Total CPU - TLB CPU % of Total CPU

TLB1 Cycles per TLB Miss – Cycles per TLB Miss

PTE % of all TLB1 Misses – Page Table Entry % misses

B\* - Basic Counter Set - Counter Number

See “The Set-Program-Parameter and CPU-Measurement Facilities” SA23-2260-0 for full description

E\* - Extended Counters - Counter Number

See “IBM The CPU-Measurement Facility Extended Counters Definition for z10” SA23-2261-0 for full description or “The CPU-Measurement Facility Extended Counters Definition for z10 and z196” SA23-2261-01 for full description

# z10 and z196 CPU MF COUNTERS Summary

- **Traditional metrics continue to provide the best view of Performance**
  - CPU MF can help explain why a change occurred
- **First Step completed in Workload Characterization for Capacity Sizing**
  - Relative Nest Intensity calculation today gives a hint to zPCR
- **Volunteers are still needed for our Workload Characterization study for refinement**
  - Feedback from Volunteers is this is very easy to enable, with a minimal time investment
- **CPU MF has a very low overhead to run and is easy to implement**
  - Less than 1/100 of a second for HIS address space in 15 minute interval
  - Customers are successfully running CPU MF in Production Today
- **Recommend enabling CPU MF COUNTERS on z10s and z196s today!**
  - To supplement current performance metrics (e.g. from SMF, RMF, DB2, CICS), turn on and leave on
  - APAR OA30486 required for z196s and recommended for z10s
- **CPU MF Overview and WSC Experiences Techdoc TC000041**
  - <http://www.ibm.com/support/techdocs/atmastr.nsf/WebIndex/TC000041>
  - CPU MF presentation and a detailed write up for enabling CPU MF





# Acknowledgements

- **Many people contributed to this presentation including:**

Riaz Ahmad

Greg Boyd

Jane Bartik

Harv Emery

Gary King

Frank Kyne

Steve Olenik

Bob Rogers

Bill Schray

Brian Smith

Bob St John

Elpida Tzortzatos

Kathy Walsh

# Disclaimer



Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Thank You  
for attending!

# Appendix

# CPU MF – Lessons Learned since August 2009

- **CPU MF Performance Metrics can be used to help understand why performance changed**
- **Customers are successfully running CPU MF COUNTERS collecting SMF 113s**
  - Over days and months without any reported performance impact
  - Feedback from Volunteers is this is very easy to enable, with a minimal time investment
- **SMF 113 Logical CPU IDs are equal to the SMF 70 Logical CPU IDs**
  - Can match up SMF 113s & SMF 70s to identify GCPs, zIIPs or zAAPs
  - Can see the unique Vertical Polarity Logical CPs cache/memory characteristics
    - E.G. Vertical Mediums may have higher L2 Remote activity
- **In multi-book z10 ECs there can be L2 Remote Activity even if  $\leq 12$  GCPs**
  - Because of I/O activity from SAPs as the data is initially stored in the Remote L2
- **Utilize the Counter Version Number fields to map to technology**
  - Number is increased for a change in meaning or number of counters
    - SMF113\_2\_CTRVN1 – Basic or Problem-State counter sets
    - SMF113\_2\_CTRVN2 – Crypto or Extended counter sets

# CPU MF Update – Lessons Learned since March 2009

- L1 Miss % can be determined from CPU MF COUNTERS
- z10 EC must be at bundle #20 or higher for CPU MF COUNTERS
- IRD considerations
  - If CPU goes offline, only activity within internal is recorded in an Intermediate record, then
    - If no activity in follow on 15 minute interval(s), Intermediate record is not cut for the CPUID
      - No Final record when HIS is ended
    - When activity resumes, Intermediate record is written for CPUID
- New **APAR OA27623** to add “CPU Speed” to SMF 113 and to HIS COUNTERS output
  - *Processor speed for which the hardware event counters are recorded. Speed is in cycles / microsecond* - “4404” for z10 EC
  - SMF 113 new field: SMF113\_2\_CPSP - 4 byte binary
  - Simplifies conversion of Cycles into “Time”
- Customers are successfully running CPU MF COUNTERS (and collecting SMF 113s) over 24 hours
- Analyze the “major” LPARs on a z10 at the same time

# Documentation

- *MVS Commands* SA22-7627-19
  - Setting up hardware event data collection 1-39
- *The Set-Program-Parameter and CPU-Measurement Facilities* SA23-2260-0
  - Full description of Basic, Problem-State and Crypto Counter Sets
- *IBM The CPU-Measurement Facility Extended Counters Definition for z10* SA23-2261-0
- *IBM The CPU-Measurement Facility Extended Counters Definition for z10 and z196* SA23-2261-01
- WSC Short Stories and Tall Tales
  - SHARE Summer 2009 Denver - Session 2136 – John Burg
- *CPU MF Overview and WSC Experiences* Techdoc TC000041 available March 26 2010
  - <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TC000041>
  - SHARE Winter 2010 presentation and detailed write up for enabling CPU MF – John Burg
- ITSO Red Book reference Planned for 4QT 2010
  - *Exploiting System z LPAR Capacity Controls* - SG24-7846. 2 Part Book:
    - Part 1 - CPU MF
    - Part 2 - HiperDispatch, Group Capacity Controls, hard/soft capping
    - Draft available ~April 1
  - <http://www.redbooks.ibm.com/redbooks.nsf/home?ReadForm&page=drafts>

# APAR OA27623 “CPU Speed” – HIS COUNTERS Output

```

HIS019I EVENT COUNTERS INFORMATION VERSION 1
FILE NAME: SYSHIS20090615.112833.CNT
COMMAND: MODIFY HIS,B,TT='UA46797',CTRONLY,CTR=ALL
COUNTER VERSION NUMBER 1: 1 COUNTER VERSION NUMBER 2: 1

COUNTER SET= BASIC
COUNTER IDENTIFIERS:
  0: CYCLE COUNT
  1: INSTRUCTION COUNT
  2: L1 I-CACHE DIRECTORY-WRITE COUNT
  3: L1 I-CACHE PENALTY CYCLE COUNT
  4: L1 D-CACHE DIRECTORY-WRITE COUNT
  5: L1 D-CACHE PENALTY CYCLE COUNT

START TIME: 2009/06/15 11:28:33 START TOD: C4574FEC19DF7217
END TIME: 2009/06/15 12:19:09 END TOD: C4575B3B6919C911
COUNTER VALUES (HEXADECEMAL) FOR CPU 00 (CPU SPEED = 4404 CYCLES/MIC):
  0- 3 00000017978F0641 000000044358C932 000000000C3DB63E 00000001A038D222
  4- 7 00000000223375DD 00000004F5D256E8 -----

START TIME: 2009/06/15 11:28:33 START TOD: C4574FEC19E10D97
END TIME: 2009/06/15 12:19:09 END TOD: C4575B3B691AE091
COUNTER VALUES (HEXADECEMAL) FOR CPU 05 (CPU SPEED = 4404 CYCLES/MIC):
  0- 3 00000016D275AAA9 00000004395C24A6 000000000C2E714E 000000019E57E8E0
  4- 7 00000000219A39DC 00000004E4C3881F -----

START TIME: 2009/06/15 11:28:33 START TOD: C4574FEC19E29817
END TIME: 2009/06/15 12:19:09 END TOD: C4575B3B69188C11
COUNTER VALUES (HEXADECEMAL) FOR CPU 0A (CPU SPEED = 4404 CYCLES/MIC):
  0- 3 0000000002803BE2 00000000001B9237 000000000000D93D 0000000000A5B310
  4- 7 0000000000021461 0000000001D9D453 -----

START TIME: 2009/06/15 11:28:33 START TOD: C4574FEC19E43D97
END TIME: 2009/06/15 12:19:09 END TOD: C4575B3B691C7411
COUNTER VALUES (HEXADECEMAL) FOR CPU 0B (CPU SPEED = 4404 CYCLES/MIC):
  0- 3 0000000002513682 00000000001B92C2 000000000000F3FE 000000000095A685
  4- 7 000000000002092A 0000000001D32119 -----

START TIME: 2009/06/15 11:28:33 START TOD: C4574FEC19E58997
END TIME: 2009/06/15 12:19:09 END TOD: C4575B3B691D5311
COUNTER VALUES (HEXADECEMAL) FOR CPU 0C (CPU SPEED = 4404 CYCLES/MIC):
  0- 3 0000000002460D98 00000000001B8F79 000000000000FF29 00000000007A4435
  4- 7 0000000000020BEF 0000000001AFF518 -----

START TIME: 2009/06/15 11:28:33 START TOD: C4574FEC19E73D97
END TIME: 2009/06/15 12:19:09 END TOD: C4575B3B691E2E91
COUNTER VALUES (HEXADECEMAL) FOR CPU 0D (CPU SPEED = 4404 CYCLES/MIC):
  0- 3 00000000021ADEE1 00000000001B9152 000000000000E98F 0000000000644954

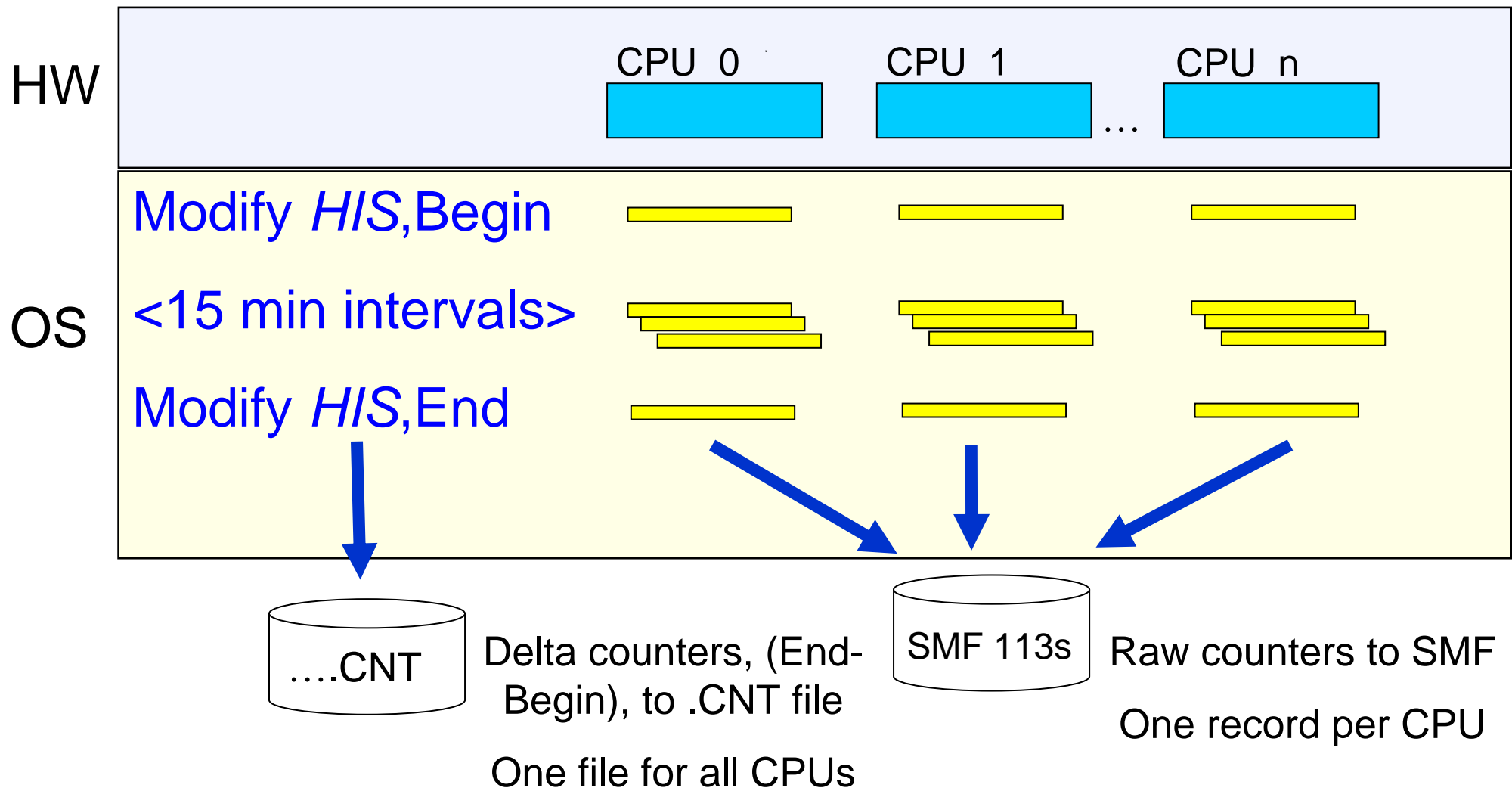
```

These numbers come from a synthetic Benchmark  
and do not represent a production workload



# How it works

## Hardware Instrumentation Counters



# What data is in the CPU MF – per Logical CP

- **Basic Counters (and Problem) per CPU – (1)**

- Cycles
- Instructions
- L1 Cache Sourcing **basic** information

- **Crypto Counters per CPU – (1)**

- Counts and Cycles by Crypto function

- **Extended Counters – per CPU (Model Dependent) – (2)**

- Cache Hierarchy Information and more
  - z10 L1 Sourcing **detailed** information

z10 L1 Cache  
Hierarchy  
Sourcing

1 - See “The Set-Program-Parameter and CPU-Measurement Facilities” SA23-2260-0 for full description

2 - See “IBM The CPU-Measurement Facility Extended Counters Definition for z10” SA23-2261-0 for full description

See Appendix for Basic, Problem and Crypto Counters

# What we did

- **Set up CPU MF on WSC z10 and z/OS 1.10**
- **Started/Modified HIS and collected SMF 113s and \*.CNT Data**
  - Ran “COUNTERS” mode, COUNTERS=ALL (Basic, Problem, Crypto, Extended) via:
  - **“F HIS,B,TT='EncrypCounters2',PATH='/his/',CTRONLY,CTR=ALL”**
- **Ran DASD dumps**
  - DASD dumps sequentially over 20 minute duration
  - With option: ENCRYPT(CLRTPDES) -
- **Built sample reports with a REXX exec**
  - Used \*.CNT output to as input
  - Validated with SMF 113s
  - Reports
    - Basic Counters
    - Basic / Extended Counters - z10 L1 Cache Hierarchy Sourcing Report
    - Crypto Counters

## SYSHISyyyymmdd,hhmmss.CNT Output unformatted

```

- HIS019I EVENT COUNTERS INFORMATION
- FILE NAME: SYSHIS20090207.161102.CNT

- COMMAND: MODIFY HIS,B,TT='EncrypCounters2',PATH='/his/',CTRONLY,CTR=ALL
- COUNTER VERSION NUMBER 1: 1 COUNTER VERSION NUMBER 2: 1

- COUNTER SET= BASIC
- COUNTER IDENTIFIERS:
-   0: CYCLE COUNT
-   1: INSTRUCTION COUNT
-   2: L1 I-CACHE DIRECTORY-WRITE COUNT
-   3: L1 I-CACHE PENALTY CYCLE COUNT
-   4: L1 D-CACHE DIRECTORY-WRITE COUNT
-   5: L1 D-CACHE PENALTY CYCLE COUNT

- START TIME: 2009/02/07 16:11:02 START TOD: C3B6ADBE7AD83D26
- END TIME: 2009/02/07 16:31:19 END TOD: C3B6B24700FC45A5
- COUNTER VALUES (HEXADECIMAL) FOR CPU 00:
-   0- 3 0000004689BEBF20 0000000433831366 0000000014CF0790 000000021B57E0D8
-   4- 7 000000002A620C97 0000000B25C43DBC -----

- START TIME: 2009/02/07 16:11:02 START TOD: C3B6ADBE7AD95826
- END TIME: 2009/02/07 16:31:19 END TOD: C3B6B24700FD3625
- COUNTER VALUES (HEXADECIMAL) FOR CPU 01:
-   0- 3 00000048CFB22F1D 000000048D23D49A 00000000154D89E5 0000000229B662EA
-   4- 7 000000002C1F067B 0000000B8087F6A7 -----

- START TIME: 2009/02/07 16:11:02 START TOD: C3B6ADBE7ADABCA6
- END TIME: 2009/02/07 16:31:19 END TOD: C3B6B24700FE1525
- COUNTER VALUES (HEXADECIMAL) FOR CPU 04:
-   0- 3 00000021DE76A328 0000000A8F16E5E9 0000000000022392 00000000008AC8F2
-   4- 7 000000001B92F07B 000000035E926CFD -----

- COUNTER SET= PROBLEM-STATE
- COUNTER IDENTIFIERS:
-   32: PROBLEM-STATE CYCLE COUNT
-   33: PROBLEM-STATE INSTRUCTION COUNT
-   34: PROBLEM-STATE L1 I-CACHE DIRECTORY-WRITE COUNT
-   35: PROBLEM-STATE L1 I-CACHE PENALTY CYCLE COUNT
-   36: PROBLEM-STATE L1 D-CACHE DIRECTORY-WRITE COUNT
-   37: PROBLEM-STATE L1 D-CACHE PENALTY CYCLE COUNT

```

**Description**

**Start / End time**

**Counters per CPU - 00**

**Counters per CPU - 01**

**Counters per CPU - 04**

# Sample Report – Basic Counters

```
*** z10 Summary - BASIC Counters Information ***
*** TOTAL for all CPUs ***
```

```
Cycle Count          625429033.94/Sec
Instruction Count     68153013.72/Sec
L1 I-Cache Directory-write Count  580653.65/Sec
L1 I-Cache Penalty Cycle Count  15076029.05/Sec
L1 D-Cache Directory-write Count  1572649.35/Sec
L1 D-Cache Penalty Cycle Count  91824855.27/Sec
```

Normalized Basic Counters to per Second

L1 Index and Directory Write Counts used In Cache  
Hierarchy Sourcing

```
Total z10 Busy : 4.79% - for the 3 CPUs
```

## L1 Miss % can be derived from CPU MF information

• Instruction Count is the base. If instructions are not in z10 L1 Cache, then they must be “Sourced” from the z10 hierarchy. The Total “Sourced” is the Total Write Count, the “Misses”

• **L1 Miss % = Directory Write Counts (I+D) / Instruction Counts**

• **3.2% = (580,653.65 + 1,572,649.35) / 68,15,013.72**

These numbers come from a synthetic Benchmark  
and do not represent a production workload

# Sample Report – Basic / Extended Counters

## z10 L1 Cache Hierarchy Sourcing

```

*** z10 Summary - EXTENDED Counters Information ***
*** TOTAL for all CPUS ***
*** L1 Cache Hierarchy Report from L1.5, L2 & Memory ***

Source for L1                                     %          RATE Unit
-----
Dir Write L1 Inst Cache from L1.5                26.40%       568501.41 /Sec
Dir Write L1 Data Cache from L1.5                52.40%      1128297.67 /Sec
Dir Write L1 Inst Cache from L2 on same Book       0.54%        11690.52 /Sec
Dir Write L1 Data Cache from L2 on same Book      20.16%      434192.22 /Sec
Dir Write L1 Inst Cache from L2 NOT on same Book   0.00%         2.19 /Sec
Dir Write L1 Data Cache from L2 NOT on same Book   0.22%        4703.85 /Sec
Dir Write L1 Inst Cache from Memory on same Book   0.00%         4.05 /Sec
Dir Write L1 Data Cache from Memory on same Book   0.00%        63.44 /Sec
Dir Write L1 Inst Cache from Memory NOT on same Book 0.02%        455.48 /Sec
Dir Write L1 Data Cache from Memory NOT on same Book 0.25%        5392.17 /Sec
Total                                             100.00%      2153303.00 /Sec

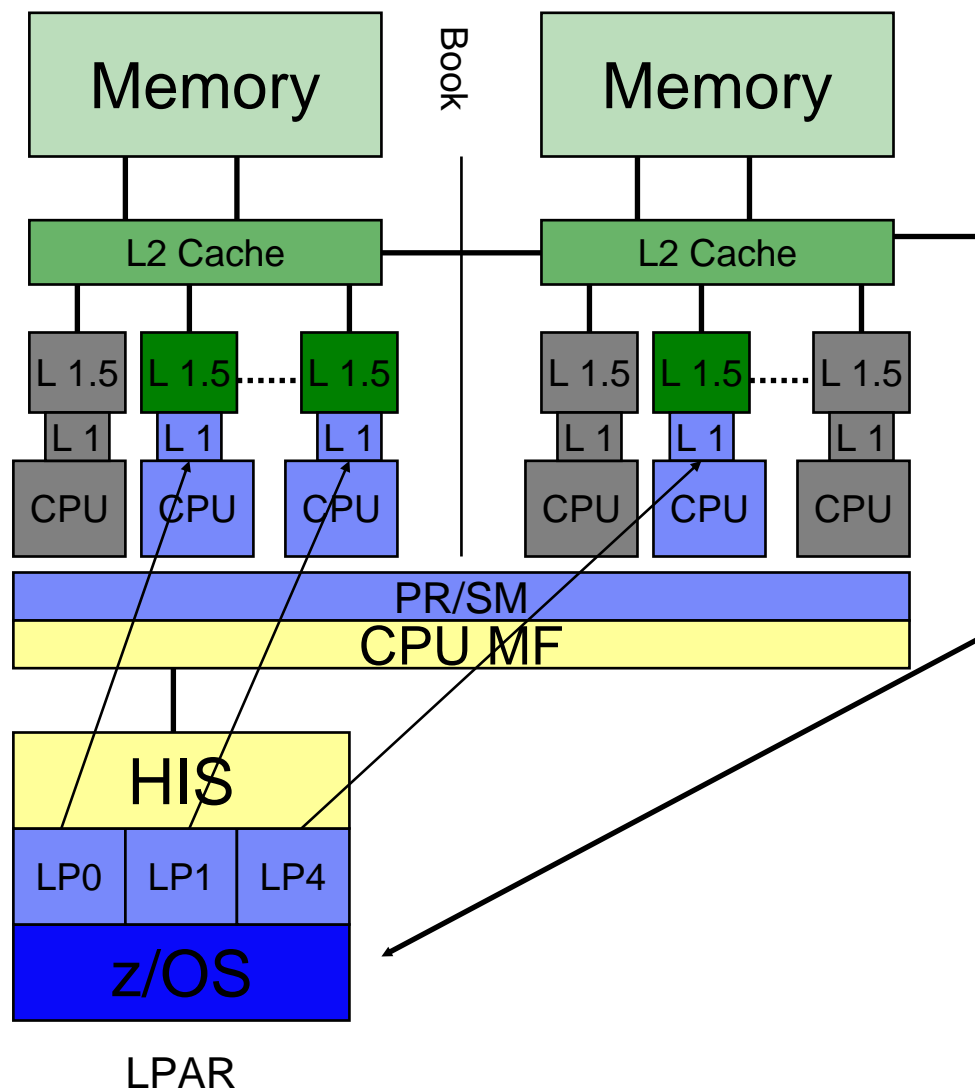
```

Various Sources from Extended Counters

← Total L1 Sourcing from Basic Counters

These numbers come from a synthetic Benchmark  
and do not represent a production workload

# CPU MF and HIS provide a z/OS logical view of z10 Resource Usage and Cache Hierarchy Sourcing



## LPAR / Logical CP view:

Memory .27%

### Cache

L 2 (local and remote) 20.92%

L1.5 78.8%

### L1 Sourced from above Hierarchy

2.15 Million / Sec

3.2% L1 Miss %

96.8% L1 Hit %

These numbers come from a synthetic Benchmark and do not represent a production workload

# Sample Report – Crypto Counters

```

*** z10 Summary - CRYPTO Counters Information ***
*** TOTAL for all CPUs ***

PRNG Function Count          0/Sec
PRNG Cycle Count            0/Sec
PRNG Blocked Function Count  0/Sec
PRNG Blocked Cycle Count    0/Sec
SHA Function Count          0.73/Sec
SHA Cycle Count            592.47/Sec
SHA Blocked Function Count   0/Sec
SHA Blocked Cycle Count     0/Sec
DEA Function Count          6277.39/Sec
DEA Cycle Count            332273396.24/Sec
DEA Blocked Function Count   0/Sec
DEA Blocked Cycle Count     0/Sec
AES Function Count          0/Sec
AES Cycle Count            0/Sec
AES Blocked Function Count   0/Sec
AES Blocked Cycle Count     0/Sec

*** CRYPTO BUSY SUMMARY ***

PRNG  Crypto Busy:  0.00% - for the 3 CPUs
SHA   Crypto Busy:  0.00% - for the 3 CPUs
DEA   Crypto Busy:  2.55% - for the 3 CPUs
AES   Crypto Busy:  0.00% - for the 3 CPUs
-----
Total Crypto Busy:  2.55% - for the 3 CPUs

```

This information may be useful in determining:

- When and What encryption function is occurring (Count)?
- How many cycles are being used?

The encryption facility executed both SHA functions and TDES functions for this specific test.

Since CPU MF is new, this information is not available from RMF today

Need to analyze more Customer data

These numbers come from a synthetic Benchmark and do not represent a production workload



# Image Profile Security Customization for HIS

TSYSHMC: Customize/Delete Activation Profiles - Mozilla Firefox

https://9.82.36.91/hmc/wcl/T2867#W2860\_treeSel

## Customize Image Profiles: TSY:TOSP2 : TOSP2 : Security

- TSYS:TOSP2
  - TOSP2
    - General
    - Processor
    - Security**
    - Storage
    - Options
    - Load
    - Crypto

**Partition Security Options**

- ☒ Global performance data control
- ☐ Input/output (I/O) configuration control
- ☒ Cross partition authority
- ☐ Logical partition isolation

**Counter Facility Security Options**

- ☒ Basic counter set authorization control
- ☒ Problem state counter set authorization control
- ☒ Crypto activity counter set authorization control
- ☒ Extended counter set authorization control
- ☒ Coprocessor group counter sets authorization control

**Sampling Facility Security Options**

- ☒ Basic sampling authorization control

Save Copy Profile Paste Profile Assign Profile Cancel Help

# Counter Data

- **Basic Counter Set**

- Cycle count
- Instruction count
- Level-1 I-cache directory write count
- Level-1 I-cache penalty cycle count
- Level-1 D-cache directory write count
- Level-1 D-cache penalty cycle count

- **Problem State Counter Set**

- Problem state cycle count
- Problem state instruction count
- Problem state level-1 I-cache directory write count
- Problem state level-1 I-cache penalty cycle count
- Problem state level-1 D-cache directory write count
- Problem state level-1 D-cache penalty cycle count

- **Extended Counter Set**

- Number and meaning of counters are model dependant

# Counter Data

- **Crypto Activity Counter Set (CPACF activity)**
  - PRNG function count
  - PRNG cycle count
  - PRNG blocked function count
  - PRNG blocked cycle count
  - SHA function count
  - SHA cycle count
  - SHA blocked function count
  - SHA blocked cycle count
  - DES function count
  - DES cycle count
  - DES blocked function count
  - DES blocked cycle count
  - AES function count
  - AES cycle count
  - AES blocked function count
  - AES blocked cycle count

## SMF Record type 113, subtype 2

Layout: (SMF manual, HISYSMFR macro)

- Standard SMF record header ('1C'x bytes)
- SMF record control information
  - TOD when SMF record is written, etc.
  - Offset, length, and number of data sections
- Data section
  - TOD when counter data was captured
  - CPU number
  - Offset, length, and number of Counter Set Sections
  - Offset, length, and number of Counter Sections
  - Counter Set Sections
    - Counter Set type (1=BASIC, 2=PROB, 3=CRYPTO, 4=EXT)
    - Bit mask identifying the counters being recorded in array
      - ✓ e.g. 'FC00000000000000'x => counters 0-5 are valid
  - Counter Sections – 8-byte counter values (contiguous)

## z10 versus z9 hardware comparison

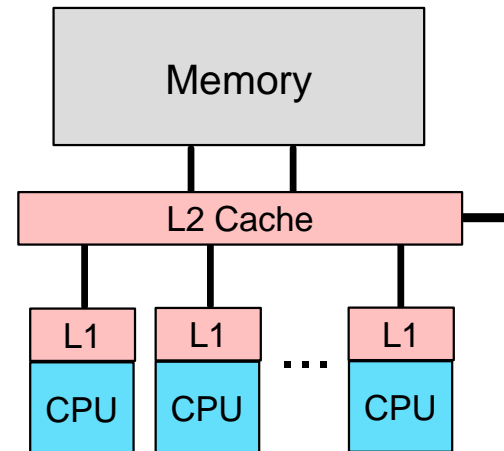
### ■ z9 EC

#### ▶ CPU

- 1.7 Ghz
- superscalar

#### ▶ Caches

- L1 private 256k i, 256k d
- L2 shared 40 mbs / book
- book interconnect: ring



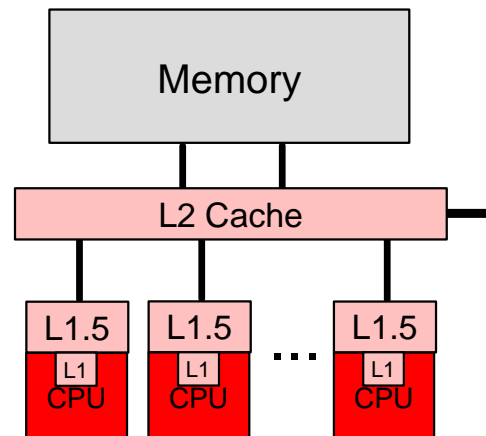
### ■ z10 EC

#### ▶ CPU

- 4.4 Ghz
- redesigned pipeline
- superscalar

#### ▶ Caches

- L1 private 64k i, 128k d
- L1.5 private 3 mbs
- L2 shared 48 mbs / book
- book interconnect: star



# Usage & Invocation - Additions to the .CNT file

- **The .CNT file adds a new line to describe the state (new version identifier)**

- When a state change was detected and STATECHANGE=STOP

```
HIS019I  EVENT COUNTERS INFORMATION VERSION 2
FILE NAME:  SYSHISyyyyymmdd.hhmmss.000.CNT
COMMAND:  MODIFY HIS,xxxx
STATE CHANGE:  YES,STOP
COUNTER VERSION NUMBER 1: xxxx  COUNTER VERSION NUMBER 2: xxxx
```

- When a state change was detected and STATECHANGE=IGNORE

```
HIS019I  EVENT COUNTERS INFORMATION VERSION 2
FILE NAME:  SYSHISyyyyymmdd.hhmmss.000.CNT
COMMAND:  MODIFY HIS,xxxx
STATE CHANGE:  YES,IGNORE
COUNTER VERSION NUMBER 1: xxxx  COUNTER VERSION NUMBER 2: xxxx
```

- When a state change was detected and STATECHANGE=SAVE

```
HIS019I  EVENT COUNTERS INFORMATION VERSION 2
FILE NAME:  SYSHISyyyyymmdd.hhmmss.000.CNT
COMMAND:  MODIFY HIS,xxxx
STATE CHANGE:  YES,SAVE
COUNTER VERSION NUMBER 1: xxxx  COUNTER VERSION NUMBER 2: xxxx
```

- When no state change was detected

```
HIS019I  EVENT COUNTERS INFORMATION VERSION 2
FILE NAME:  SYSHISyyyyymmdd.hhmmss.000.CNT
COMMAND:  MODIFY HIS,xxxx
STATE CHANGE:  NO
COUNTER VERSION NUMBER 1: xxxx  COUNTER VERSION NUMBER 2: xxxx
```