IBM zSystems Tech Bytes

Presented by the Washington Systems Center

How to Measure Those New z16 Capabilities

John Burg
jpburg@us.ibm.com
Executive IT Specialist
October 6, 2022

Agenda

- z16 CPU MF Update
- New z16 SMF 30 support
  - Crypto
  - AIU
- z16 CPU MF AIU Reporting

CPU Measurement Facility

- Introduced in z10 and later processors
- Facility that provides hardware instrumentation data for production systems
- Two Major components
  - Counters
    - **Cache and memory hierarchy information**
    - SCPs supported include z/OS and z/VM
  - Sampling
- New z/OS HIS started task
  - Gathered on an LPAR basis
  - Writes SMF 113 records
  - z/OS implementation instructions http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TC000066

New z16 formulas

- New z/VM Monitor Records
  - Gathered on an LPAR basis – all guests are aggregated
  - Writes new Domain 5 (Processor) Record 13 (CPU MF Counters) records
  - z/VM implementation instructions http://www.vm.ibm.com/perf/tips/cpumfhow.html
- Minimal overhead

Value of CPU Measurement Facility (CPU MF)

- Recommended Methodology for successful z Systems Processor Capacity Planning
  - Need on “Before” processor to determine LSPR workload
  - **TDA process requires CPU MF Counters enabled**
- Validate achieved IBM Z processor performance
  - Needed on “Before” and “After” processors
- Provide insights for new features and functions
  - Continuously running on all LPARs
  - **Efficiency is even more important today**
  - **New insights and features including: SIIS and AIU**

New

Capturing CPU MF data is an Industry “Best Practice”

z/OS SMF 113 Record

- SMF113_2_CTRVN2
  - –“1” = z10
  - –“2” = z196 / z114
  - –“3” = zEC12 / zBC12
  - –“4” = z13 / z13s
  - –“5” = z14
  - –“6” = z15
  - –“7” = z16

L1MP and RNI-based LSPR Workload Decision Table

L1MP	RNI	LSPR Workload Match
< 3%	>= 0.75 < 0.75	AVERAGE LOW
3% to 6%	>1.0 0.6 to 1.0 < 0.6	HIGH AVERAGE LOW
> 6%	>= 0.75 < 0.75	HIGH AVERAGE

Current table applies to z10 EC, z10 BC, z196, z114, zEC12, zBC12, z13, z13s, z14, z15 and z16 CPU MF data

Reminder

- RNI is **not** a Performance metric
  - RNI and L1MP allows one to match their workload to an LSPR workload
    - Any other use of RNI is not valid

Looking for z16 Migration “Volunteers” SMF data

- Want to validate / refine Workload selection metrics

Looking for “Volunteers”

(3 days, 24 hours/day, SMF 70s, 71s, 72s, 99 subtype 14s,113s per LPAR)

“Before z14 / z15” and “After z16”

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

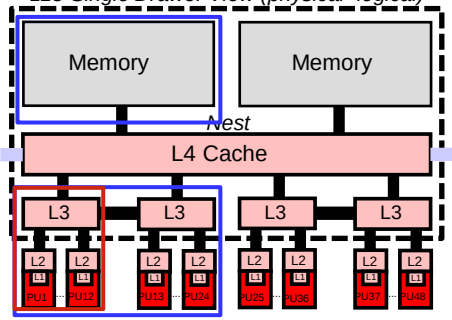
z16 Metrics

z16 vs z15 Hardware and Topology Comparison

z15 CPU 5.2 GHz
Caches L1 private 128k i, 128k d / core
L2 private 4 MB i, 4 MB d / core
L3 shared 256 MB / CP chip
L4 shared 960 MB / drawer

Topology
- 12 cores + 1 L3 / CP chip
- 2 CP chips / cluster
- 2 clusters + 1 L4 (48 engines) / drawer
- 5 drawers / CEC
- Book interconnect: numa star

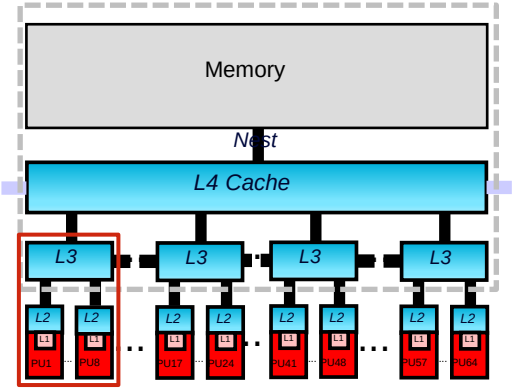
z15 Single Drawer View (physical+logical)



z16 CPU 5.2 GHz
Caches L1 private 128k i, 128k d / core
L2 *private* 32 MB unified / core
virtual victim L3 up to 7x32 = 224 MB / CP chip
virtual victim L4 up to 8x32x7 = 1.75 GB / drawer

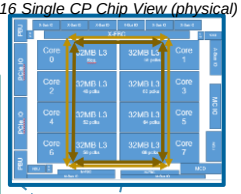
Topology
- 8 (core + L3)s / CP chip
- 2 CP chips / DCM
- 4 DCMs (64 engines) / drawer
- 4 drawers / CEC
- Book interconnect: numa star

z16 Single Drawer View (logical)

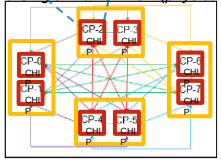


See Tech Bytes
z16 Technical Overview
Tue Oct 4th 3 PM

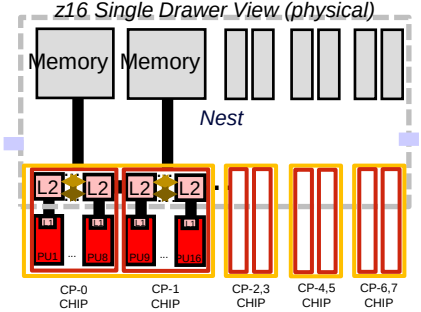
z16 Single CP Chip View (physical)



z16 Single Drawer View (physical)



z16 Single Drawer View (physical)



Formulas – z16

Workload Characterization
L1 Sourcing from cache/memory hierarchy

Metric	Calculation – *note all fields are deltas. SMF113-1s are deltas. SMF 113-2s are cumulative.*
CPI	B0 / B1
PRBSTATE	(P33 / B1) * 100
L1MP	((B2+B4) / B1) * 100
L2P	((E145+E146+E169+E170) / (B2+B4)) * 100
L3P	((E147+E149+E150+E151+E171+E173+E174+E175) / (B2+B4)) * 100
L4LP	((E148+E152+E153+E154+E160+E161+E162+E163+E164+E165+E172 +E176+E177+E178) / (B2+B4)) * 100
L4RP	((E155+E166+E167+E168+E179) / (B2+B4)) * 100
MEMP	((E156+E157+E158+E159+E180+E181+E182+E183) / (B2+B4)) * 100
LPARCPU	(((1/CPSP/1,000,000) * B0) / Interval *in Seconds*) * 100

Updated May 31, 2022

Note these Formulas may change in the future

CPI – Cycles per Instruction
Prb State - % Problem State
L1MP – Level 1 Miss Per 100 instructions
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, zIPs) captured and uncaptured

B* - Basic Counter Set - Counter Number
P* - Problem-State Counter Set - Counter Number
See "The Load-Program-Parameter and CPU-Measurement Facilities" SA23-2260 for full description
E* - Extended Counters - Counter Number
See "IBM The CPU-Measurement Facility Extended Counters Definition for z10, z196/ z114, zEC12 /zBC12, z13/z13s, z14, z15 and z16" SA23-2261-07 for full description
CPSP - SMF113_2_CPSP "CPU Speed"

Formulas – z16 Additional

Metric	**Calculation**– *note all fields are deltas. SMF113-1s are deltas. SMF 113-2s are cumulative.*
Est Instr Cmplx CPI	CPI – Finite CPI
Finite CPI	E143 / B1
SCPL1M	E143 / (B2+B4)
Rel Nest Intensity	4.3*(0.45*L3P + 1.3*L4LP + 5.0*L4RP + 6.1*MEMP) / 100
Eff GHz	CPSP / 1000

Updated May 31, 2022

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 Load-Program-Parameter and CPU-Measurement Facilities" SA23-2260 for full description

E* - Extended Counters - Counter Number

See "IBM The CPU-Measurement Facility Extended Counters Definition for z10, z196/ z114, zEC12 /zBC12, z13/z13s, z14, z15 and z16" SA23-2261-07 for full description

CPSP - SMF113_2_CPSP "CPU Speed"

CPU MF SIIS Indicator can help Identify potential SIIS

- CPU MF can be used to help identify potential SIIS timeframes
  - Based on % of certain I Writes / D Writes sourced
  - LPAR view, identifies when it happens, not who is causing it
    - Identify the program(s) running in the time period, e.g. via zBNA Top Programs
    - Use a hot spot analyzer to find the issue
    - Remediate the source code to correct the issue

See *Identifying SIIS Inefficiency using CPU MF Counters* http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102806

Processor	SIIS Indicator %	Description	
zEC12 / zBC12	E130 / B4 * 100%	D Writes sourced with L2 intervention / D Writes	
z13 / z13s	E163 / B2 * 100%	I Writes sourced with L3 intervention / I Writes	
z14 / ZR1	E164 / B2 * 100%	I Writes sourced with L3 intervention / I Writes	
z15	E164 / B2 * 100%	I Writes sourced with L3 intervention / I Writes	
z16	E170 / B2 * 100%	I Writes sourced with L2 intervention / I Writes	***New***

Processor Topology

- Processor Topology
  - Where logical processors are assigned and run on the physical cores
  - Can be important in understanding changes in Performance
- Today – 3 Ways
  - HMC (one time – not dynamic nor in SMF)
    - LPAR Resource Assignment
  - SMF 99-14s (z/OS systems(s) view only)
  - LPAR Dumps (one time – need IBM to execute and process)
- New z16 Topology Support
  - All partitions (Processor view), not just z/OS
  - Recorded in SMF 70 record, so its dynamic
    - Includes a Topology change indicator
  - Supported in z/VM also

See Tech Bytes
Hot Topics
Thur Oct 13th 3 PM

New z16 Processor Topology Support

- Support
  - z/OS new Data Gatherer functionality is delivered with APAR OA62064
    - PTFs are available for z/OS V2.5, V2.4, and V2.3.
  - z/VM -support will be provided via apar for z/VM 7.1 and z/VM 7.2
    - Will be in the base of z/VM 7.3 (available in 3QT 2022)

- WSC Tools Support
  - zPCR and zCP3000

IBM zSystems Best Practice Performance Recommendations

- Set weights and logicals to meet needs (GCPs and zIIPs)
  1. Understand LPAR capacity requirements across time
  2. Manage weights to meet capacity requirements
  3. Assign logicals to meet weights (CPs by weight)
     - Only 1-2 more than needed to meet CPs by weight
     - Optimize for VHs
  4. Utilize zPCR to help assess 2 and 3

  See Best Practice for ***Number of Logical CPs Defined for an LPAR***
  www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD106388

- Utilize z16 (70s), SMF 99-14s and SMF 113s to understand topology and impact by polarity / logical processor

- Topology Change can occur for any change in
  – LPAR (de)activation, Weight, Logical Processor
    - Weight change includes IRD, WLM Capping (Defined Capacity and Group Capacity)

z16 Crypto and AIU SMF 30

Thanks to Eysha Powers and Scott Ballentine for intro slides

CPACF Usage Tracking

With IBM z16, IBM provides processor activity instrumentation to count cryptographic operations

Compliance

Which crypto was used?

Performance

What was the frequency of crypto use?

Configuration

Did my crypto configuration change take effect?

https://www.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep_ca/1/897/ENUS122-001/index.html

Enhanced SMF Type 30 Records

Include new crypto counter sections that contain counters for CPACF cryptographic instructions utilized by a job in a given period. These sections are produced only for those instructions that are used.

These counters are correlated with z/OS jobs and users for the determination of the algorithms, bit lengths, and key security utilized by a given workload.

This data can aid in compliance, performance, and configuration.

IBM z16 ...
CPACF Usage Tracking

Deliver a new hardware managed counter set to track crypto algorithms, bit lengths and key security (e.g., AES 256 encrypted).

Provides evidence for:

- compliance (i.e., which crypto is used)
- performance (i.e., frequency of crypto use)
- configuration (i.e., proof of change)

For z/OS, the counters could be:

- Synchronized with the SMF recording interval
- Stored in new triplet section of existing SMF Type 30 records
- Configured on a system wide (i.e., IEASYSxx) basis

Available to all IBM Z operating systems

New and/or enhanced audit logs containing CPACF counter values

Correlated with the workload that invoked the CPACF function (e.g. PCI workload)

SMF Type 30 Record: Self-Defining Section

Offset (dec)	Offset (hex)	Name	Length	Format	Description
190	BE	SMF30USN	2	binary	Number of zEDC usage statistics sections.
192	C0	SMF0_End_V1	0	n/a	End of version 1.
192	**C0**	**SMF30CPO**	**4**	**Binary**	**Offset to the CrypCtrs section**
196	**C4**	**SMF30CPL**	**2**	**Binary**	**Length of the CrypCtrs section**
198	**C6**	**SMF30CPN**	**2**	**Binary**	**Number of CrypCtrs sections**
200	**C8**	**SMF30CPA**	**4**	**Binary**	**Number of CrypCtrs sections in subsequent (continuation) records**
204	CC	SMF30NPO	4	Binary	Offset to the NNPICtrs section
208	D0	SMF30NPL	2	Binary	Length of the NNPICtrs section
210	D2	SMF30NPN	2	Binary	Number of NNPICtrs sections
212	D4	SMF30NPA	4	Binary	Number of NNPICtrs sections in subsequent (continuation) records
216	**D8**	**SMF30_Cont_Recs_To_Follow**	**2**	**Binary**	**The number of continuation records to follow this record. When SMF30_RecCont_LastRec is on, this field will be zero.**
218	**DA**	**SMF30_RecCont_Flags**	**1**	**Binary**	**Record continuation flags Bit 0 - SMF30_RecCont_FirstRec - When on, this record is the first of a set of two or more continuation records Bit 1 - SMF30_RecCont_AdditionalRec - When on, this record is the second or subsequent but not last record in a set of two or more continuation records. Bit 2 - SMF30_RecCont_LastRec - When on, this record is the last of a set of two or more continuation records**
219	**DB**	**SMF0_End_V2**	**0**	**n/a**	**End of version 2.**

The value in SMF30SOF (not shown above) contains the offset to the first section of the record and is equal to the length of the SMF30SDS section. Constants SMF30SDS_Len_V1 and SMF30SDS_Len_V2 can be compared against SMF30SOF to determine which version of the record was created.

SMF Type 30 Record: Product / Subsystem Section

Offset (dec)	Offset (hex)	Name	Length	Format	Description
...					
2	2	SMF30RS1	1		Reserved.
3	3	SMF30PFlags	1	Binary	Product Flags Bit 0 - SMF30_zCBP **Bit 1 - SMF30_CrypCtrs_Active - Indicates that CrypCtrs is active** Bit 2 – SMF30_NNPICtrs_Active - Indicates that NNPICtrs is active

These new bits are in reserved space in the existing record, so no special version checking is needed to determine if these bits are available.

SMF Type 30 Record: Triplet Sections

This section can be located in the SMF Type 30 record using the following triplet fields, which are located in the 'header/self-defining' section:

- Offset: SMF30CPO
- Length: SMF30CPL
- Number: SMF30CPN - Reports the number of sections in the current record.
- Number: SMF30CPA - Reports the number of sections in subsequent SMF type 30 records.

Constants for each crypto counter entry ID are provided in macro IFASMFCN.

Offset (dec)	Offset (hex)	Name	Length	Format	Description
...					
0	0	SMF30_CrypCtrs_Entry_ID	2	Binary	CrypCtrs entry identifier
2	2	SMF30_CrypCtrs_Count	8	Binary	CrypCtrs count value

Note: Counters with non-zero values will be present, and there will be one section per non-zero counter. For example, if you have 5 non-zero counters, you will have 5 sections, not 1 section with 5 values.

Crypto / AIU SMF 30 Counters Support

- Support
  - Need APAR OA61511 and OA61395 to be installed to gather in SMF 30 records
    - PTFs are available for z/OS V2.5 and V2.4

SMF 30 Crypto and AIU NNPA Support

- Value resides in
  - Entry ID and Counts
  - AIU Neural Network Processor Assist (NNPA) instruction also uses this SMF 30 support
    - Utilized by WMLz and Db2
    - SMF 30 Counts are not available for zCX applications
  - The Counts represent a **Delta value** since last interval – WSC requirement
    - Can determine accesses / load in interval and correlate to other Performance metrics
    - Thus capturing Performance Value
      - What is CPU impact when utilizing Encryption algorithms?
      - Who is using (or not using) what Algorithm / Function?
      - What is the NNPA function “mix” and how does it relate to the “model”?

SMF 30 Crypto Entry IDs

- Over 150 Crypto Entry Ids
  - What Algorighm(s), key(s), and how many occurrences in an address space?
  - See SYS1.MACLIB(IFASMFCN)

Counter number	Processor activity counted
1	KM-DEA function ending with CC=0
2	KM-TDEA-128 function ending with CC=0
3	KM-TDEA-192 function ending with CC=0
4	KM-Encrypted-DEA function ending with CC=0
5	KM-Encrypted-TDEA-128 function ending with CC=0
6	KM-Encrypted-TDEA-192 function ending with CC=0
7	KM-AES-128 function ending with CC=0
8	KM-AES-192 function ending with CC=0
9	KM-AES-256 function ending with CC=0
10	KM-Encrypted-AES-128 function ending with CC=0
11	KM-Encrypted-AES-192 function ending with CC=0
12	KM-Encrypted-AES-256 function ending with CC=0
13	KM-XTS-AES-128 function ending with CC=0
14	KM-XTS-AES-256 function ending with CC=0
15	KM-XTS-Encrypted-AES-128 function ending with CC=0
16	KM-XTS-Encrypted-AES-256 function ending with CC=0
118	PCC-Compute-XTS-Parameter-Using-Encrypted-AES-256 function ending with CC=0

SMF 30 Crypto Example

- z16 simple IFASMFDP step
  - Encrypting output data set
  - Results in CPACF Counts for algorithms

```
//STEP8    EXEC PGM=IFASMFDP
//*
//* CREATES EF ENCRYPTED DATA SET PHYSICAL SEQ DATA SET
-  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  1 Line(s
//*
//INDD1    DD  DSN=JPBURG.WSCSYSD.SMF.SYSD.SEP1417.T03,DI
-  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  4 Line(s
//OUTDD1   DD  DSN=JPBURG.WSCSYSD.SMF.SYSD.T03.EFEN2,DISP
-  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -_ 1 Line(s
//           DATACLAS=DSENCS,
//           STORCLAS=ATSCTS,
//           DSKEYLBL='SAMPLE.SECRET.AES256.KEY001',
```

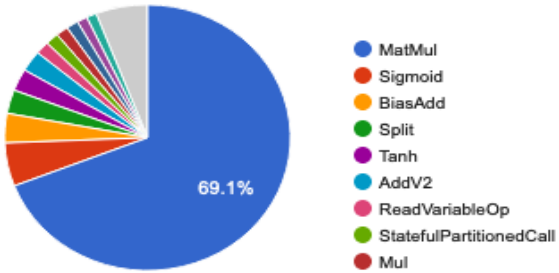
```
Subtype Record 4
SMF ID SYSD Work Type JES2 Ho Min Sec 091237  Year Mon Day 20220415
Usage Offset 0 Length 76 Number 0
zEDC Offset 1375 Length 64 Number 0
z16 Crypto Offset 1525 Length 10 Number 2  Num Sections in Addl Records 0  Yes Crypto Counters
Crypto Counters - Entry Id:  16 Count Value: 73395 Description: KM-XTS-Encrypted-AES-256 function ending w cc=0
Crypto Counters - Entry Id:  118 Count Value: 24477 Description: PCC-Compute-XTS-Parameter-Using-Encrypted-,
             AES-256 function ending with CC=0
z16 NNPI (AI) Offset 1519 Length 10 Number 0  Num Sections in Addl Records 0   No AIU Counters
TYP   PFlags   RV  4    96   05
Job Step Program JPBURGE8 STEP8    IFASMFDP
EXCPs 49001 SSCHs 17205
CPU Time and SRB Time in Seconds 0.18   0.13
Highest task % and Program Name 0 IEFIIC
Tot, TCB, SRB and I/O Ser Units 34333 17177 12252 4904
Total Instructions 689240763  Instruction Flags (0 is normal) = 0
TCB and SRB Time in Microseconds 183579.1875   130943.25  CPI  2.372
```

AI Accelerator

- More than just matrix multiplication!

- AI Functions/Macros abstracted via NNPA instruction
  - Elementwise, Activation
  - Normalization, Pooling
  - Matrix-multiplication
  - Convolution
  - Conv+Scale+Activate
  - MatMul+Compare/Activate
  - RNN activation

- Functions can be added per firmware update



RNN model

Function group	#	Function support in GA1
Elementwise ops	0x10	NNPA_EL_ADD
	0x11	NNPA_EL_SUB
	0x12	NNPA_EL_MUL
	0x13	NNPA_EL_DIV
	0x14	NNPA_EL_MIN
	0x15	NNPA_EL_MAX
Activation ops	0x20	NNPA_LOG
	0x21	NNPA_EXP
	0x31	NNPA_RELU
	0x32	NNPA_TANH
	0x33	NNPA_SIGMOID
Norm op.	0x34	NNPA_SOFTMAX
	0x40	NNPA_BATCHNORM
Pooling	0x50	NNPA_AVGPOOL2D
	0x51	NNPA_MAXPOOL2D
Systolic ops	0x70	NNPA_CONVOLUTION
	0x71	NNPA_MATMUL_OP
	0x72	NNPA_MATMUL_OP_BCAST23
RNN	0x60	NNPA_LSTMACT
	0x61	NNPA_GRUACT
	0x00	NNPA_QAF

SMF 30 AIU NNPA Entry IDs

- 27 NNPA Entry Ids
  - What NNPA functions and how many occurrences in an address space?
  - Tensor attributes
    - 6 Entry IDs, 22-27
  - See SYS1.MACLIB(IFASMFCN)

Counter number	Processor activity counted
1	NNPA-ADD function ending with CC=0
2	NNPA-SUB function ending with CC=0
3	NNPA-MUL function ending with CC=0
4	NNPA-DIV function ending with CC=0
5	NNPA-MIN function ending with CC=0
6	NNPA-MAX function ending with CC=0
7	NNPA-LOG function ending with CC=0
8	NNPA-EXP function ending with CC=0
9	Reserved for IBM use
10	NNPA-RELU function ending with CC=0
11	NNPA-TANH function ending with CC=0
12	NNPA-SIGMOID function ending with CC=0
13	NNPA-SOFTMAX function ending with CC=0
14	NNPA-BATCHNORM function ending with CC=0
15	NNPA-MAXPOOL2D function ending with CC=0
16	NNPA-AVGPOOL2D function ending with CC=0
17	NNPA-LSTMACT function ending with CC=0
18	NNPA-GRUACT function ending with CC=0
19	NNPA-CONVOLUTION function ending with CC=0
20	NNPA-MATMUL-OP function ending with CC=0
21	NNPA-MATMUL-OP-BCAST23 function ending with CC=0
22	NNPA function with conditions as described in “Common Oper
23	NNPA function with conditions as described in “Common Oper
24	NNPA function with conditions as described in “Common Oper
25	NNPA function with conditions as described in “Common Oper
26	NNPA function with conditions as described in “Common Oper
27	NNPA function with conditions as described in “Common Oper

SMF 30 AIU NNPA Example

- z16 AIU NNPA Example 1
  - Including Tensor

```
Subtype Record 4
SMF ID S01  Work Type OMVS Ho Min Sec 143016  Year Mon Day 20220302
Usage Offset 0 Length 76 Number 0
zEDC Offset 1279 Length 64 Number 0
z16 Crypto Offset 1519 Length 10 Number 0  Num Sections in Addl Records 0  No Crypto Counters
z16 NNPI (AI) Offset 1525 Length 10 Number 5  Num Sections in Addl Records 0   Yes AIU Counters
AIU Counters - Entry Id:  16 Count Value: 14 Description: NNPA - AvgPool2D
AIU Counters - Entry Id:  22 Count Value: 10 Description: NNPA - Tensor Small bat
AIU Counters - Entry Id:  25 Count Value: 3 Description: NNPA - Tensor 1 Mb frame suitable
AIU Counters - Entry Id:  26 Count Value: 11 Description: NNPA - Tensor 2 Gb frame suitable
AIU Counters - Entry Id:  27 Count Value: 220 Description: NNPA - Tensor AIU data area access exception
TYP   PFlags   RV  4    96   05
Job Step Program TESTFLR1 *OMVSEX  BPXPRECP
EXCPs 6827 SSCHs 112
CPU Time and SRB Time in Seconds 12.92   0
```

SMF 30 AIU NNPA Example

- z16 AIU NNPA Example 2

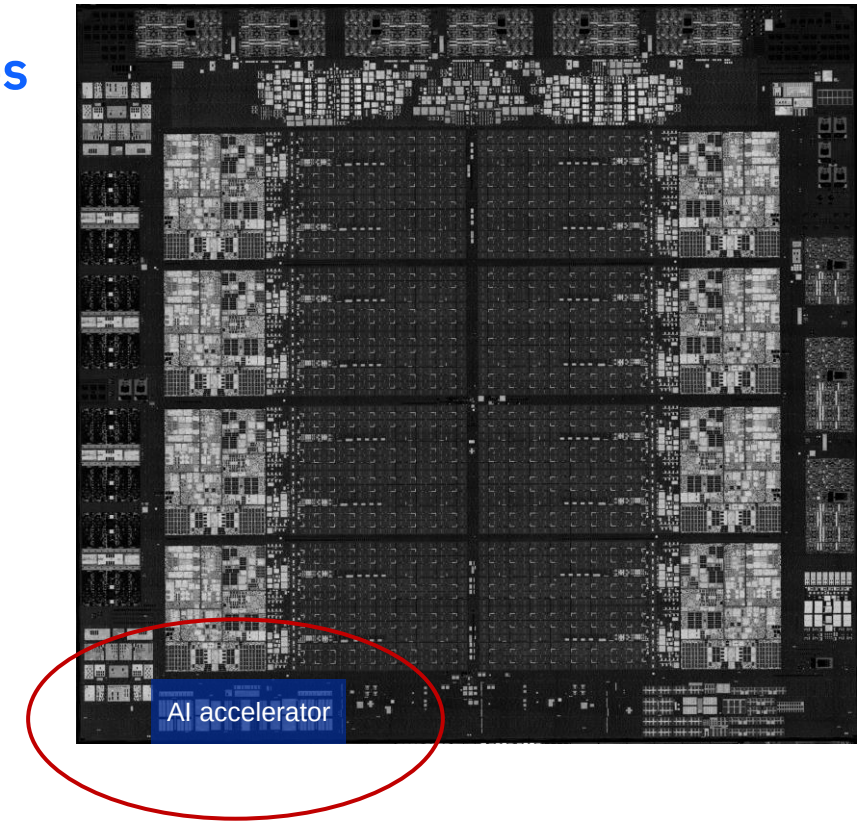
```
Subtype Record 4
SMF ID S01  Work Type OMVS Ho Min Sec 151301  Year Mon Day 20220302
Usage Offset 0 Length 76 Number 0
zEDC Offset 1279 Length 64 Number 0
z16 Crypto Offset 1519 Length 10 Number 0  Num Sections in Addl Records 0  No Crypto Counters
z16 NNPI (AI) Offset 1525 Length 10 Number 8  Num Sections in Addl Records 0   Yes AIU Counters
AIU Counters - Entry Id:  1 Count Value: 1 Description: NNPA - Add
AIU Counters - Entry Id:  2 Count Value: 1 Description: NNPA - Sub
AIU Counters - Entry Id:  3 Count Value: 1 Description: NNPA - Mul
AIU Counters - Entry Id:  4 Count Value: 1 Description: NNPA - Div
AIU Counters - Entry Id:  5 Count Value: 2 Description: NNPA - Min
AIU Counters - Entry Id:  22 Count Value: 6 Description: NNPA - Tensor Small bat
AIU Counters - Entry Id:  24 Count Value: 6 Description: NNPA - Tensor Very small size tensor
AIU Counters - Entry Id:  27 Count Value: 2 Description: NNPA - Tensor AIU data area access exception
TYP   PFlags   RV  4    96   05
Job Step Program TESTFLR5 *OMVSEX  BPXPRECP
EXCPs 1666 SSCHs 107
CPU Time and SRB Time in Seconds 0.03   0
```

Measuring AIU Performance with CPU MF

IBM z16 Integrated Accelerator for AI

Centralized On-chip accelerator shared by all cores

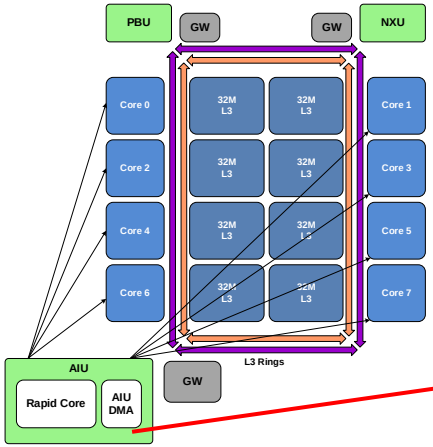
- Very low and consistent inference latency
- Compute capacity for utilization at scale
- Variety of AI models ranging from traditional ML to RNNs and CNNs
- Security – provide enterprise-grade memory virtualization and protection
- Extensibility with future firmware and hardware updates

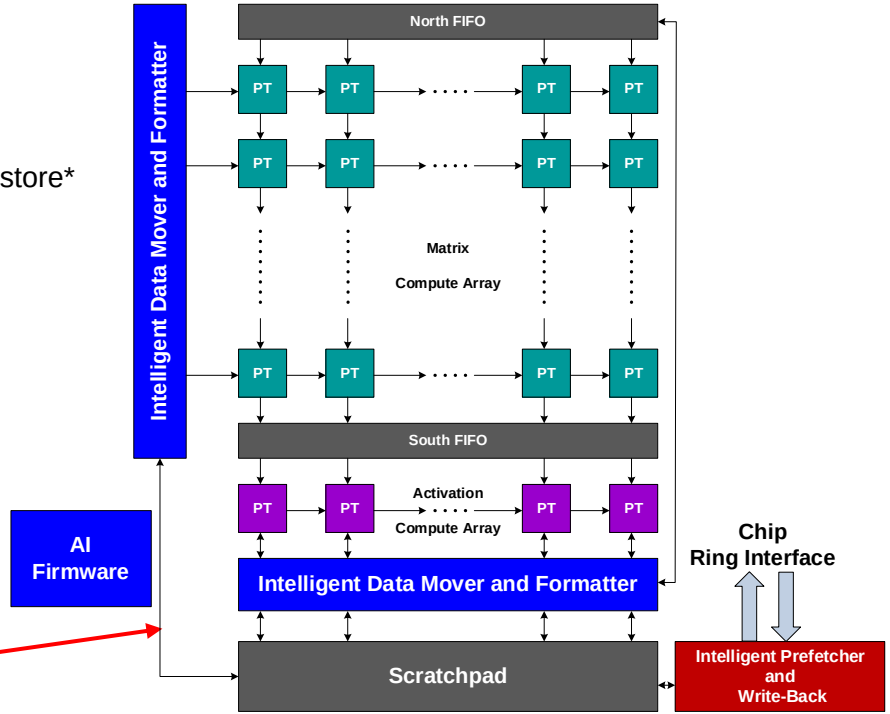


See Tech Bytes
Delivering AI on z16
Thur Oct 13th 11 AM

AI Accelerator

- Central on-chip AI accelerator, shared among all cores;
  - Similar concept to compression accelerator in IBM z15
- Aggregate of >6 TFLOPS / chip
- Over 200 TFLOPS on 32-chip system
- Neural Network Processing Assist (NNPA) instruction
  - Memory-to-memory CISC instruction
  - Operates directly on tensor data in user space
- Connected to chip ring core: 120+ GB/s read, 80+ GB/s store*
- Matrix array for matrix multiplication and convolution
- Special engines for complex functions





* for properly conditioned AI models

z16 CPU MF and AIU

- z16 CPU MF Formulas published at GA (May 31) **Updated**
  - CPU Formulas
- AIU Metrics **New**
  - In addition to address space SMF 30 AIU counts, CPU MF can be utilized
  - CPU MF has thread level if running in SMT2
  - LPARCPU equivalents, is absolute load, where 100 = 1 “Engine”
    - **AIUCPU** – Total AIU CPU
      - **WAIUCPU** – Waiting for access to AI
        - Maximum is logical processors * CF * 100
          - If zIIPs running SMT2 scale by Capacity Factor (CF), 1 for GCPs
      - **CAIUCPU** - Executing on AIU
        - Maximum is 100 for a Chip (up to 8 cores driving 1 AIU accelerator)
  - **AIUCPI**
    - AIU Executing Cycles per completed instruction

SA23-2261-07

Counter 267 – Increments by one for every NUERAL NETWORK PROCESSING ASSIST instruction executed.

Counter 268 – Increments by one for every NUERAL NETWORK PROCESSING ASSIST instruction executed that ended in Condition Codes 0, 1 or 2.

Counter 269 – Cycles CPU spent obtaining access to IBM Z Integrated Accelerator for AI.

Counter 270 – Cycles CPU is using IBM Z Integrated Accelerator for AI

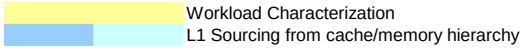
Sample WSC z16 CPU MF Metrics

CPI | Work Load Mix | CPI Micro Proc | CPI L2 & 'Nest' | How Often? 100% of L1 Misses | How Far? | How Much? | LSPR Workload Match

Hour	CPI	Prb State	Instr Cmplx CPI	Finite CPI	SCPL1M	L1MP	L2P	L3P	L4LP	L4RP	MEMP	Rel Nest Intensity	LPARCPU	Eff GHz	Machine Type	LSPR Wkld
18.50	2.09	33.46	1.19	0.90	22	4.1	96.6	0.9	2.1	0.0	0.4	0.23	6.3	5.2	Z16	LOW

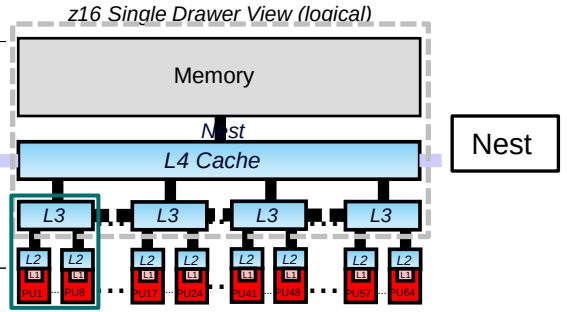
Finite CPI
Important Metric for LPAR Controls

CPI – Cycles per Instruction – A rate of delivery metric
- **EICPI – Estimated Instruction Complexity CPI – Indicates portion of CPI related to the microprocessor**
- **EFCPI - Estimated Finite CPI – Indicates portion of CPI related to the L2 private and shared caches (Nest)**



L1MP Sourced from Cache Hierarchy

Cycles / Instructions



Measuring the AIU Accelerator with CPU MF – Example 2

- Fraud Detection test
  - –P53 System
  - –1 Interval
- Utilized RMF CPU Activity
- Utilized 99-14s Topology
  - –SMF 70 Topology APAR OA62064 was not installed
- Created CPU MF Analysis to identify AIU Accelerator Busy

P53 CPU MF – Fraud Detection AIU Test SMF 99-14 Topology

- SMF 99-14 Topology
  - Still valid for z16
- RMF and 99-14s identify activity on
  - Drawer 1, DCM 4, Chip 2
  - Across 6 zIIPs in SMT2 mode

18	IIP	100.00	95.12	88.53	0.00	96.58	91.87
				86.14	0.00		
19	IIP	100.00	92.92	85.71	0.00	96.39	89.56
				85.54	0.00		
1A	IIP	100.00	63.10	56.78	0.00	100.00	63.10
				51.22	0.00		
1B	IIP	100.00	50.19	44.55	0.00	98.05	49.21
				41.13	0.00		
1C	IIP	100.00	40.11	35.27	0.00	96.00	38.51

C P U A C T I

z/OS V2R4 SYSTEM ID P53 DATE
RPT VERSION V2R4 RMF TIME

--- PU ---			TIME %			MT %	
NUM	TYPE	ONLINE	LPAR BUSY	MVS BUSY	PARKED	PROD	UTIL
				32.61	0.00		
1D	IIP	100.00	32.32	28.16	0.00	96.29	31.12
				26.04	0.00		

MULTI-THREADING ANALYSIS

CPU TYPE	MODE	MAX CF	CF	AVG TD
CP	1	1.000	1.000	1.000
IIP	2	1.369	1.249	1.677

P53 CPU MF – Fraud Detection AIU Test

- CPU MF by pool / logical
  - SMT2 thread level for each logical
- **AIUCPI**
  - AIU Exec Cycles per completed instruction
- **AIUCPU** – Total AIU CPU
- **WAIUCPU** – Waiting for access to AIU
  - Maximum is logical processors * 100
  - zIIPs SMT scale by capacity factor
- **CAIUCPU** - Executing on AIU
  - Maximum is 100 for a Chip

6 zIIPs 2 Threads each

SYSID	DAY	HOUR	CPU NUM RMF	Thread	Z16 CPI	POOL	CPID	LPARCPU	Z16 EICPI	Z16 EFCPI	L1MP	Z16 SCPL1M	AIUCPI	AIUCPU	WAIUCPU	CAIUCPU
P53	7	21.33	18	0	5.57	6	48	91.90	4.18	1.39	3.96	35	47,026.98	56.07	41.92	14.14
P53	7	21.33	18	1	5.65	6	49	89.40	4.26	1.39	4.00	35	47,784.31	55.11	41.50	13.61
P53	7	21.33	19	0	5.62	6	50	88.90	4.24	1.38	3.98	35	47,553.72	54.93	41.43	13.50
P53	7	21.33	19	1	5.63	6	51	88.80	4.25	1.38	3.99	35	47,565.40	54.82	41.34	13.48
P53	7	21.33	1A	0	8.50	6	52	58.90	7.22	1.29	3.60	36	41,213.35	45.70	38.49	7.21
P53	7	21.33	1A	1	9.91	6	53	53.20	8.50	1.41	3.83	37	42,081.49	41.85	35.72	6.13
P53	7	21.33	1B	0	10.96	6	54	46.20	9.53	1.43	3.73	38	41,377.42	37.60	32.64	4.96
P53	7	21.33	1B	1	12.03	6	55	42.70	10.52	1.51	3.83	39	41,567.97	34.91	30.50	4.40
P53	7	21.33	1C	0	12.93	6	56	36.60	11.39	1.54	3.73	41	41,106.93	30.46	26.91	3.54
P53	7	21.33	1C	1	13.80	6	57	33.90	12.20	1.61	3.80	42	41,680.38	28.14	24.91	3.22
P53	7	21.33	1D	0	13.33	6	58	29.20	11.79	1.54	3.52	44	41,204.98	24.35	21.70	2.66
P53	7	21.33	1D	1	14.40	6	59	27.00	12.77	1.64	3.67	45	41,324.69	22.46	20.03	2.43
															Total ==>	89.28

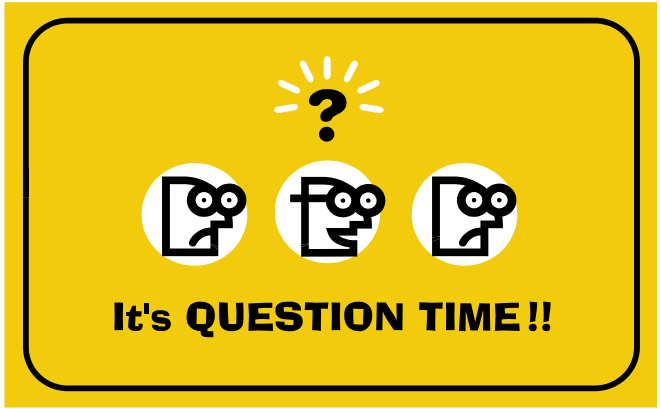
Q: Given 6 zIIPs resided on a Chip, how busy Is the AIU Accelerator?

A: Sum of CAIUCPU / 100 * 100% = 89.28%

Summary

- CPU MF continues to evolve and add value
  - Its an IBM Best Practice to collect CPU MF data on all LPARs

- There are some really cool new z16 performance metrics
  - New SMF 30 support
    - Crypto
    - AIU
  - CPU MF AIU Reporting



Any questions, follow-up, or requests for enhancements
please contact the authors or send an email to:
cpstools@us.ibm.com

- Session Evaluation link is provided in the Chat for this session.
- Please fill out a session evaluation as it does help us greatly!
- 3 PM EDT – Jacob Emery *Automated Openshift on IBM zSystems Deployment*

Notices and disclaimers

Notices and disclaimers

— Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of 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. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

— The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

— IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml

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.

Formulas – z16 Additional TLB

Metric	**Calculation** – *note all fields are deltas. SMF113-1s are deltas. SMF 113-2s are cumulative.*
Est. TLB1 CPU Miss % of Total CPU	((E130+E135) / B0) * (E143 / (B3+B5)) *100
Estimated TLB1 Cycles per TLB Miss	(E130+E135) / (E129+E134) * (E143 / (B3+B5))
PTE % of all TLB1 Misses	N/A with processor design change
TLB Miss Rate	(E129 + E134) / interval

Updated May 31, 2022

Note these Formulas may change in the future

Est. TLB1 CPU Miss % of Total CPU - Estimated TLB CPU % of Total CPU

Estimated TLB1 Cycles per TLB Miss – Estimated Cycles per TLB Miss

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

TLB Miss Rate – TLB Misses per interval (interval is defined by user for length of measurement and units)

B* - Basic Counter Set - Counter Number

P* - Problem-State Counter Set - Counter Number

See "The Load-Program-Parameter and CPU-Measurement Facilities" SA23-2260 for full description

E* - Extended Counters - Counter Number

See "IBM The CPU-Measurement Facility Extended Counters Definition for z10, z196/ z114, zEC12 /zBC12, z13/z13s, z14, z15 and z16" SA23-2261-07 for full description

CPSP - SMF113_2_CPSP "CPU Speed"

Understanding CPU MF Metrics - 1

- CPI – Cycles per Instruction
  - EICPI – Estimated Instruction Complexity CPI – Indicates portion of CPI related to the microprocessor
  - EFCPI - Estimated Finite CPI – Indicates portion of CPI related to the L2 private and shared caches (Nest)
- PRB – The % of Problem State instructions. This is an indicator of the workload mix, so a changing of PRB%, may indicate different workload mixes running.
- ESCPL1M – Estimated sourcing cycles per L1 Miss
- L1MP – Level 1 Miss Percentage – The average Level 1 miss percentage per 100 instructions. It is an indicator of **"How Often"** the instructions and data are not found in the L1 cache, and must be sourced further out in the cache hierarchy. It is a component in matching to the LSPR workload. If L1MP is ~>6%, it may be an indicator of CICS Threadsafe opportunity.
- L2P - Level 2 Cache Miss Percentage – The percent of misses sourced from the private Level 2 cache
- L3P – Level 3 Cache Miss Percentage - The percent of misses sourced from the shared Level 3 cache
- L4LP – Level 4 Local Cache Miss Percentage - The percent of misses sourced from the shared Level 4 Local cache
- L4RP – Level 4 Remote Cache Miss Percentage - The percent of misses sourced from the shared Level 4 Remote cache

Understanding CPU MF Metrics - 2

- MEMP – The Memory Cache Miss Percentage - The percent of misses sourced from the shared memory
- RNI – The Relative Nest Intensity – **"How Far"** out in the Nest are Instructions and Data sourced. It is a component in matching to the LSPR workload.
  - z13 RNI: 2.3*(0.4*L3P + 1.6*L4LP + 3.5*L4RP + 7.5*MEMP) / 100
  - z14 RNI: 2.4*(0.4*L3P + 1.5*L4LP + 3.2*L4RP + 7.0*MEMP) / 100
  - z15 RNI: 2.9*(0.45*L3P + 1.5*L4LP + 3.2*L4RP + 6.5*MEMP) / 100
- LPARCPU – This is a measurement of **"How Much"** load is running. 100% equals 1 Engine
- LSPR WKLD – The LSPR Workload this system matches to based on its L1MP and Relative Nest Intensity (RNI).
- **TLB Metrics**
- ETLBCPUP –The estimated CPU % related to TLB misses. Some portion of this amount may be able to be reduced with Large Pages.
- PTEP - The Page Table Entry % of TLB misses. If PTEP is >40%, it may be an indicator of applicability of Large Pages to reduce CPU.
- ETLBCYPM – The estimated TLB sourcing cycles per TLB Miss