The Many CPU Fields Of SMF

Cheryl Watson
Watson & Walker, Inc.
www.watsonwalker.com

November 5, 2013 Session VB

Agenda

- Sources of SMF CPU Usage
- What is a CPU Second?
- CPU Field Precision
- Normalization
- Address Space CPU Usage
- Service Class CPU Usage
- LPAR Usage
- CEC Usage
- CPU Variability
- z/OS 2.1
- References



Sources of CPU Information in SMF

- RMF CPU Records (Type 70)
 - CEC CPU usage, LPAR usage, zIIP usage, zAAP usage, IFL usage, CF usage
- RMF Workload Activity Records (Type 72)
 - CPU usage by service class period
- SMF Address Space Activity (Type 30)
 - CPU usage by address spaces, including crossaddress space, and cross-system usage

Additional Sources of CPU Info

- DB2 Records (Type 102)
- CICS Records (Type 110)
- MQ Records (Type 115)
- WAS Records (Type 120)
- WebSphere Message Broker (Type 117)
- HTTP Server (Type 103)
- Hardware (Type 113)
- RMF Monitor II (Type 79)
 - CPU usage by address spaces and enclaves
- ^⁴ **●** TSO/E (Type 32)

CPU Time Precision

- CPU fields
 - .01 most fields are in hundredths of seconds
 - .001 milliseconds
 - .000001 microseconds
 - .001024 1024-microseconds units (and 1.024-millisecond units)
 - .000128 128-microsecond units
 - .000001 TOD field, where bit 51 is one microsecond
 - .0000000625 one raw CPU or SRB service unit (a sixteenth of a microsecond) – not multiplied by service definition coefficient

What is a Second?

- A CPU second is defined as one clock second
- Theoretically, a job that takes one second of CPU time on a machine will take two seconds of CPU time on a machine that is half as fast, or one-half second on a machine that is twice as fast. Does this happen?
- For chargeback or capacity planning, how do you measure the speed of a machine?

Normalization

- Different Speed CECs
 - What is the normalization factor for chargeback or capacity planning?
 - Most sites use LSPR ratios, MIPS from CPU charts, or service units
 - Example:
 - z196 2817-501 1-way has an LSPR ratio of 1.05, is 588 MIPS, and has a published service unit/second (su/sec) rate of 30888.0309
 - z196 2817-701 1-way has an LSPR ratio of 2.15, is 1202
 MIPS, and has a published su/sec rate of 61776.0618
 - Notice ratios: 1202/588 = 2.04; 61776.0618/30888.0309= 2.0; 2.15 / 1.05 = 2.05

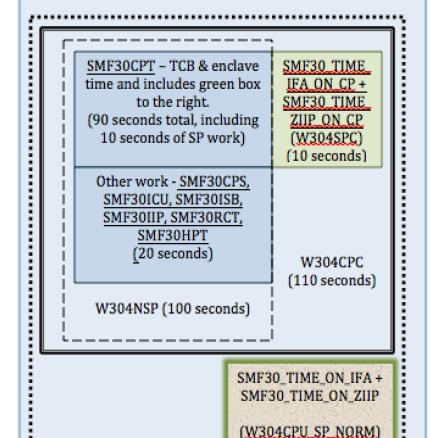
Normalization

- "Knee-capped" CPUs
 - Base CPs run at a slower (degraded) speed, while zIIPs and zAAPs run at base speed
 - For example, the zIIP and zAAP on a 2817-501 1way are the same speed as the 2817-701, which is twice as fast.
 - SMF records include normalization factor
 - The zEC12 has three series of machines that are knee-capped. The 4xx series is about 16% of a 7xx; the 5xx is about 42%; and the 6xx is about 63%

Address Space CPU Usage

- SMF Type 30 Records
 - 30.2 & 30.3 Written at end of interval
 - 30.4 Written at end of step
 - 30.5 Written at end of job
 - CPU times are in hundredths of seconds (.01 seconds)
 - Some of the CPU time can be from CPs, some from zIIPs, some from zAAPs, and some from other LPARs or CECs

Standard CPs (300 MIPS)



W304CPU (130 seconds) (20 seconds normalized) zIIP or zAAP (600 MIPS)

SMF30_TIME_ ON_IFA + SMF30_TIME_ ON_ZIIP (W3040NSP) (10 seconds)

SMF30_TIME_ IFA_ON_CP and SMF30_TIME_ ZIIP_ON_CP (5 seconds converted to SP time)

> W304CPU_SP (15 seconds)

- Work that ran on the standard CPs (1 of 2)
 - SMF30CPT TCB time, enclave time, preemptable SRB time, client SRB time and CPU time for work that was eligible for zIIPs & zAAPs, but that ran on the CP (last 2 fields on next page)
 - SMF30CPS SRB CPU time that ran on the CP
 - SMF30ICU TCB CPU time for initiator work; sum of SMF30ICU_STEP_INIT for this step and SMF30ICU_STEP_TERM from the previous step
 - SMF30ISB SRB CPU time for initiator work; sum of SMF30ISB_STEP_INIT for this step and SMF30ISB_STEP_TERM from the previous step

- Work that ran on the standard CPs (2 of 2)
 - SMF30IIP CPU time processing I/O interrupts (SLIH)
 - SMF30RCT Region control task CPU time (startup and swapping)
 - SMF30HPT CPU time spent moving Hiperspace data
 - SMF30_TIME_IFA_ON_CP Work that is eligible for a zAAP, but that ran on the CP
 - SMF30_TIME_ZIIP_ON_CP Work that is eligible for a zIIP, but that ran on the CP

- Work that ran on a zAAP or zIIP
 - SMF30_TIME_ON_IFA Work that ran on a zAAP
 - SMF30_TIME_ON_ZIIP Work that ran on a zIIP
- Potential work for zAAP
 - (SMF30_TIME_IFA_ON_CP * 256 / SMF30ZNF) + SMF30_TIME_ON_IFA
- Potential work for zIIP

- Total work that ran on a CP
 - SMF30CPT + SMF30CPS + SMF30ICU + SMF30ISB + SMF30ICU + SMF30IIP + SMF30RCT + SMF30HPT
- Potential work for the CP
 - Total of above + (SMF30_TIME_ON_IFA * SMF30ZNF / 256) + (SMF30_TIME_ON_ZIIP * SMF30SNF / 256)
- Note: SMF30ZNF and SMF30SNF = 256 if SPs are same speed as CPs

RMF Workload Activity Report

```
... INTERVAL 29.59.998
... INTERVAL 29.59.998
SERVICE POLICY PAGE
-SERVICE DEFINITION COEFFICIENTS-
IOC CPU SRB MSO
6.0 10.0 10.0 0.0000

SYSTEMS
---ID--- OPT SU/SEC CAP% --TIME-- INTERVAL
SYS1 00 35714.3 100 10.00.00 00.29.59
```

- Obtaining CPU time from service units:
 - SMF30SUS Copy of RmctAdjC number of sixteenths of one CPU microsecond per CPU service unit
 - SMF30CPC CPU service definition coefficient, scaled by 10
 - SMF30SRC SRB service definition coefficient, scaled by 10
 - SMF30CSU_L CPU service units; this is equivalent to SMF30CPT plus normalized SMF30_TIME_ON_IFA plus normalized SMF30_TIME_ON_ZIIP; new in z/OS 1.11
 - SMF30SRB_L SRB service units; this is equivalent to SMF30CPS; new in z/OS 1.11
 - SMF30ESU_L Independent enclave CPU service units; new in z/OS 1.11

- Obtaining CPU time from service units:
 - To convert service units to CPU time in microseconds (.000001 seconds):

```
TCB time = (SMF30CSU_L * (SMF30SUS / 16)) / (SMF30CPC / 10)

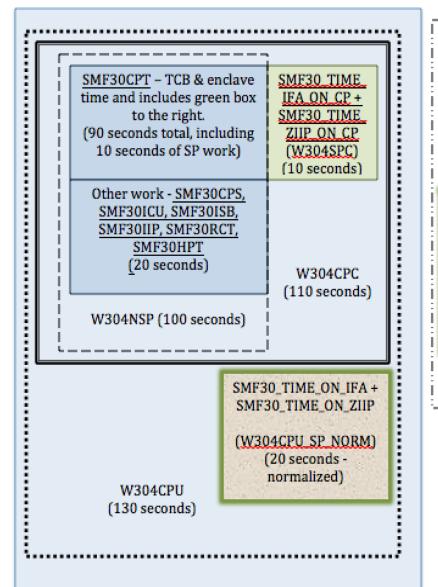
SRB time = (SMF30SRB_L * (SMF30SUS / 16)) / (SMF30SRC / 10)

Independent enclave time = (SMF30ESU_L * (SMF30SUS / 16))

/ (SMF30CPC / 10)
```

- Why?
 - Use when precision of .01 is not sufficient
 - For TCB time from service units, remember to back out the zIIP and zAAP normalized times

Standard CPs (300 MIPS)



zIIP or zAAP (600 MIPS)

SMF30_TIME_ ON_IFA + SMF30_TIME_ ON_ZIIP (W3040NSP) (10 seconds)

SMF30_TIME_ IFA_ON_CP and SMF30_TIME_ ZIIP_ON_CP (5 seconds converted to SP time)

W304CPU_SP (15 seconds)

Consistency?

- Most consistent is to normalize everything back to a CP and charge on that time (from diagram, that would be W304CPU of 130 seconds of 300 MIPS processor)
- Also consistent is to have two values and charge different rates, so one is non-specialty work that can only run on a CP and the other is specialty work that would prefer to run on an SP (from diagram, that would be W304NSP of 100 seconds and W304CPU_SP of 15 seconds of 600 MIPS processor)
- Actual time spent on each isn't consistent because it depends on parameter settings and the current load on the CPs and SPs

- Other CPU times for the obsessives (like me):
 - SMF30ASR CPU time used by preemptable SRBs and client SRBs; this is included in SMF30CPT
 - SMF30ENC CPU time used by independent enclaves when in a WLM enclave; this is included in SMF30CPT
 - SMF30DET Similar field for dependent enclaves
 - SMF30_ENCLAVE_TIME_ON_IFA Independent enclave time spent on zAAP; this is included in SMF30_TIME_ON_IFA
 - SMF30_DEP_ENCLAVE_TIME_ON_IFA Similar field for dependent enclaves

- Other CPU times for the obsessives:
 - SMF30_ENCLAVE_TIME_IFA_ON_CP CPU time used by independent enclaves on a CP that are eligible for zAAPs; this is included in SMF30_TIME_IFA_ON_CP
 - SMF30_DEP_ENCLAVE_TIME_IFA_ON_CP Similar field for dependent enclaves
 - SMF30_ENCLAVE_TIME_ON_ZIIP Independent enclave time spent on zIIP; this is included in SMF30_TIME_ON_ZIIP
 - SMF30_DEPENC_TIME_ON_ZIIP Similar field for dependent enclaves

- Other CPU times for the obsessives:
 - SMF30_ENCLAVE_TIME_ZIIP_ON_CP CPU time used by independent enclaves on a CP that are eligible for zIIPs; this is included in SMF30_TIME_ZIIP_ON_CP
 - SMF30_DEPENC_TIME_ZIIP_ON_CP Similar field for dependent enclaves
 - SMF30_ENCLAVE_TIME_ZIIP_QUAL Normalized independent enclave time qualified to be on a zIIP; the eligible time achieved is in SMF30_TIME_ON_ZIIP and SMF30_TIME_ZIIP_ON_CP
 - SMF30_DEPENC_TIME_ZIIP_QUAL Similar field for dependent enclaves

- Other CPU times for the obsessives:
 - SMF30ICU_STEP_TERM Initiator TCB time for step termination of the previous step; included in the SMF30ICU field of that step; new in z/OS 1.12
 - SMF30ISB_STEP_TERM Similar field for SRB time; included in the SMF30ISB field of that step; new in z/OS 1.12
 - SMF30ICU_STEP_INIT Initiator TCB time for step initiation of this step; is included in the SMF30ICU; new in z/OS 1.12
 - SMF30ISB_STEP_INIT Similar field for SRB time; included in the SMF30ISB field; new in z/OS 1.12
 - SMF30OST z/OS UNIX services requested by APPC/MVS work; included in SMF30CPT or SMF30CPS

- And even more:
 - SMF30UCT TCB time for registered product; included in other fields; also recorded in Type 89 record
 - SMF30UCS SRB time for registered product; included in other fields; also recorded in Type 89 record
 - SMF30_Highest_Task_CPU_Percent Largest percent of TCB time used by any task in this address space; new with APAR OA39629 (13Jul2012) for z/OS 1.12/1.13
 - SMF30_HIGHEST_Task_CPU_Program Program name associated with previous field; new with APAR OA39629
 - IBM has new tool to use these newest fields called zEnterprise Batch Workload Analyzer (zBNA)
 - For an article on zBNA, send an email to <u>marketing@watsonwalker.com</u> with name, company & address for a free copy of our last Tuning Letter

- Work that executes on another system:
 - Enclaves may run on other systems (other LPARs, and even other CECs)
 - SMF type 30 record can have multiple segments to show that work (each system is identified by field SMF30MRS)
 - SMF30MRA CPU rate adjustment factor (the number of sixteenths of one microsecond per CPU service unit)
 - SMF30MRD CPU time used by dependent enclaves on another system

- SMF Type 97
 - Contains CPU time for work run on this system, but sent by another system

Service Class Period CPU Usage

- RMF Type 72.3 Records
 - 72 Written at end of RMF interval
 - CPU times are in service units and microseconds

CPU Usage

- R723CCPU TCB service units including zAAP & zIIP time on CP, client SRBs, and enclaves
- R723CSRB SRB service units
- R723RCT RCT in microseconds
- R723HST Hiperspace time in microseconds
- R723IFAT zAAP time in microseconds
- R723IFCT zAAP time spent on CPs in microseconds
- R723CSUC zIIP service units spent on CPs; included in R723CCPU
- R723CIFA zAAP service units
- R723CIFC zAAP service units spent on CPs; included in R723CCPU

- Fields used for normalization:
 - R723MCPU CPU (TCB) service definition coefficient * 10,000
 - R723MSRB SRB service definition coefficient * 10,000
 - R723MADJ Adjustment factor for CPU rate
 - R723NFFI Normalization factor for zAAP; calculate normalized time on CP by multiplying with this value and dividing by 256
 - R723NFFS Normalization factor for zIIP; use same calculation
 - R723NADJ Nominal adjustment factor for CPU rate (see note)
 - R723CECA CEC adjustment factor (see note)
 - Note: z196 capacity change supported with APAR OA30968 in z/OS 1.12/1.13

- Obtaining CPU time from service units:
 - To convert service units to CPU time in microseconds (.000001 seconds):

```
TCB_time = (R723CCPU * (R723MADJ / 16))
/ (R723MCPU / 10000)
SRB time = (R723CSRB * (R723MADJ / 16))
/ (R723MSRB / 10000)
```

- Total CPU time on CPs =
 TCB_time + SRB_time + R723RCT + R723IIT +
 R723HST
- Total zIIP and zAAP time = R723IFAT + R723CSUP

- Relating RMF type 72.3 CPU total usage with SMF type 30 data:
 - RMF does not contain initiator time
 - RMF does not contain eligible zIIP/zAAP time
 - SMF precision of .01 is not very accurate
 - It's sometimes difficult to get good times for comparison (SMF and RMF would need to have similar intervals, with the same SYNC, and SMF would need to be creating interval records)

RMF Workload Activity Report

```
. . . INTERVAL 29.59.998

REPORT BY: POLICY=DAYTIME

DAYTIME WLM SERVICE POLICY

---SERVICE--- SERVICE TIME ---APPL %---

IOC 156748K CPU 18505.31 CP 1079.1

CPU 6609M SRB 3388.175 AAPCP 1.48

MSO 0 RCT 6.049 IIPCP 3.19

SRB 1210M IIT 171.501

HST 13.059 AAP 60.34

AAP 1086.112 IIP 87.50

IIP 1575.015
```

Service time is in seconds; APPL % is in percent of a single CP

RMF Workload Activity Report

```
... INTERVAL 29.59.998
... INTERVAL 29.59.998
SERVICE POLICY PAGE

-SERVICE DEFINITION COEFFICIENTS-

IOC CPU SRB MSO
6.0 10.0 10.0 0.0000

SYSTEMS
---ID--- OPT SU/SEC CAP% --TIME-- INTERVAL
SYS1 00 35714.3 100 10.00.00 00.29.59
```

Sample calculations:

- CPU SUs (6609M) + SRB SUs (1210M) = 7819M
- CPU time = (7,819,000,000 / 10) / 35714.3 = 21893.20 seconds
- From RMF report, CPU time = 18505.31 + 3388.175 = 21893.5 (COOL – it matches!)
- Total CPU time is 21893.20 + 6.049 +171.501 +13.059 = 22083.809
- ZAAP CPU time on ZAAP = 1086.112 seconds; and from AAP % .6034 * 1800 1086.12 (COOL!)
- zIIP CPU time on zIIP = 1575.015 seconds; and from IIP% .8750 * 1800 = 1575.0 (COOL!)
- CP % = 1079.1%, and from (22083.809 1086.112 1575.015) / 1800 = 10.79% (This just gives me goosebumps!)

LPAR CPU Usage

Source is RMF Type 70 CPU Record

CPU			TIME %			LOG PROC	I/O	I/O INTERRUPTS		
NUM	TYPE	ONLINE	LPAR BUSY	MVS BUSY	PARKED	SHARE %	RATE	% VIA TPI		
0	CP	100.00	68.61	68.51	0.00	100.0 HIG	H 331.3	33.30		
1	CP	100.00	70.04	69.97	0.00	100.0 HIG	H 228.2	33.38		
2	CP	100.00	64.05	63.99	0.00	100.0 HIG	H 177.9	33.86		
3	CP	100.00	69.16	69.09	0.00	100.0 HIG	H 405.8	31.88		
4	CP	100.00	68.57	68.49	0.00	100.0 HIG	н 280.0	31.94		
5	CP	100.00	62.20	62.14	0.00	100.0 HIG	H 203.5	32.82		
6	CP	100.00	68.69	68.58	0.00	100.0 HIG	н 376.8	32.03		
7	CP	100.00	68.25	68.18	0.00	100.0 HIG	H 243.9	31.87		
8	CP	100.00	62.86	62.81	0.00	100.0 HIG	H 182.4	32.92		
9	CP	100.00	68.40	68.31	0.00	100.0 HIG	н 329.1	31.85		
A	CP	100.00	81.74	81.39	0.00	100.0 HIG	н 1319	28.34		
В	CP	100.00	60.66	60.60	0.00	100.0 HIG	H 208.3	33.54		
С	CP	100.00	72.18	74.15	0.00	100.0 HIG	H 296.2	33.13		
D	CP	100.00	78.85	84.13	0.00	100.0 HIG	н 1196	30.66		
E	CP	100.00	66.20	66.16	0.00	100.0 HIG	н 12018	14.35		
F	CP	100.00	66.64	66.59	0.00	100.0 HIG	H 151.8	34.25		
10	CP	100.00	65.29	65.22	0.00	95.0 MED	182.1	34.65		
11	CP	100.00	0.00		100.00	0.0 LOW	0.00	0.00		

LPAR CPU Usage

More of RMF Type 70 Record:

CPU		TIME] %		LOG PRO	OC .	I/O INT	TERRUPTS	
NUM TYPE	YPE ONLINE LPAR BUS		MVS BUSY	PARKED	SHARE	%	RATE % VIA TPI		
							1		
3B CP	100.00	0.00		100.00	0.0	LOW	0.00	0.00	
TOTAL/AVER		19.37	68.73	100.00	1695	пом	18130	20.10	
1011111/1111111	1102	13.37	00.70		1000		10100	20.10	
My calculat	ion:		1168.31 (sar	ne as 17 *	68.73)			11	
40 AAP	100.00	43.69	43.58	0.00	100.0	HIGH			
41 AAP	100.00	20.78	20.75	0.00	50.0	MED			
TOTAL/AVER	AGE	16.12	32.17		150.0			1	
								1	
My calcula	tion:		64.33					1	
								1/\	
3C IIP	100.00	56.56	56.34	0.00	100.0	HIGH			
3D IIP	100.00	36.94	36.91	0.00	50.0	MED		/	
TOTAL/AVER	AGE	23.37	46.62		150.0			1	
								/ /	
My calcul	ation:		93.25						

LPAR CPU Usage

- Capture Ratios
 - CPs from LPAR view 1168.31%
 - CPs from Workload view 1079.1%
 - CP capture ratio = (100 * 1079.1) / 1168.31 = 92.4%
 - zAAPs from LPAR view 64.33%
 - zAAPs from workload view 60.34%
 - zAAP capture ratio = (100 * 60.34) / 64.33 = 93.8%
 - zIIPs from LPAR view 93.25%
 - zIIPs from workload view 87.50%
 - zIIP capture ratio = (100 * 87.5) / 93.25 = 93.8%

CEC CPU Usage

RMF Type 70 Record

MVS PARTITION NAME IMAGE CAPACITY NUMBER OF CONFIGURED PARTITIONS WAIT COMPLETION DISPATCH INTERVAL					Ŋ	SYS1 5001 16 NO YNAMIC	NUMBER OF	PHYSICAL CP AAP IFL ICF IIP	PROCESSORS		
	· PAR	TITION	DATA				AVERAG	E PROCESS	OR UTILIZATIO	N PERCENT.	AGES
NAME	S	WGT	MS	SU ACT	PROCE NUM	ESSOR- TYPE	 LOGICAL P EFFECTIVE			L PROCESS EFFECTIVE	
SYS1	А	339	0	969 6	60.0	CP	19.15	19.37	0.22	19.15	19.37
*PHYSICAL	*								0.85		1.21
TOTAL									1.52	51.19	53.07

- LPAR usage here is 19.37% of 60 CPs, which is 1162% compared to LPAR view of 1168%
- 60 CPs of CEC are 53.07% busy or 3184.2% (only 32 CPs needed)

CEC CPU Usage

RMF Type 70 Record

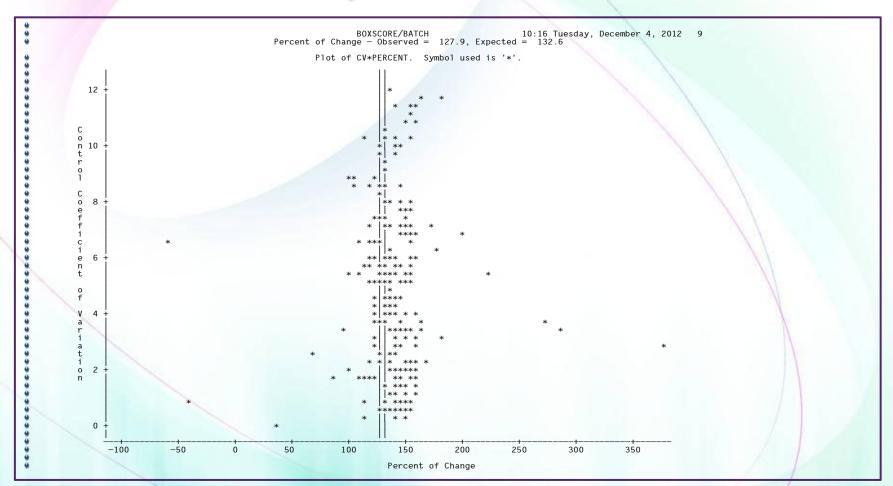
PARTITION DATA								AVERAGE	PROCESSOF	UTILIZATIO	N PERCENTA	GES
NAME	S	WGT	MS DEF	U ACT	PROCE NUM	ESSOR- TYPE		. LOGICAL PRO	OCESSORS TOTAL	PHYSICA LPAR MGMT	AL PROCESSO EFFECTIVE	
SYS1	А	375	0		4	AAP		15.95	16.12	0.17	15.95	16.12
PHYSICAL	f									1.05		1.05
TOTAL										1.51	98.01	99.52
	А	375	0		4	IIP		22.93	23.37	0.44	22.93	23.37
PHYSICAL	•									2.50		2.50
TOTAL										3.46	45.64	49.10

- ZAAP usage is 16.12% of 4 zAAPs or 64.48% compared to LPAR view of 64.33%
- zIIP usage is 23.37% of 4 zIIPs or 93.48% compared to LPAR view of 93.25%

CPU Variability

- Now that you are comfortable with the CPU fields and their precision, consider the variability of a CPU second.
- In my last Hot Flashes presentation, I included the following slide. It shows how jobs behaved after an upgrade. The average improvement was 127%, but some steps saw no improvement and others saw 300% improvement.
- Conclusion there is NO golden normalization factor!

CPU Variability



New in z/OS 2.1

- SMF Type 30, Counter Section
 - Activated when SMFCOUNT is specified in SMFPRMxx (or set with SETSMF command) and Hardware Instrumentation Services (HIS) is enabled for the Basic Counter Set
 - Records number of instructions executed on:
 - CP as TCB (non-enclave)
 - CP as SRB (non-enclave)
 - CP as preemptable or client SRB (non-enclave)
 - zIIP/zAAP (non-enclave)
 - CP but eligible for zIIP/zAAP (non-enclave)
 - CP as independent enclave
 - zIIP/zAAP as independent enclave
 - CP but eligible for zIIP/zAAP as independent enclave
 - CP as dependent enclave
 - zIIP/zAAP as dependent enclave
 - CP but eligible for zIIP/zAAP as dependent enclave

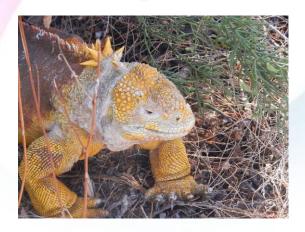
New z/OS 2.1 - **WOW!**

- RMF XP (RMF Distributed Platform Performance)
 - RMF XP can collect performance data from:
 - AIX on System p (12 subtypes)
 - Linux on System x (9 subtypes)
 - Linux on System z (11 subtypes)
 - Windows on System x (5 subtypes)
 - In z/OS 2.1, RMF XP records type 104 records to SMF
 - Obtain CPU usage, memory, I/O, configuration (e.g. number of CPUs)
 - NOW you can report on your entire complex in a single report for management

References

- IBM MVS System Management Facilities (SMF) SA22-7630
- SHARE in Anaheim #11264 SMF 101 Everything You Should Know About SMF and More, Thu, 3 pm, Cheryl Watson
- SHARE in Anaheim #11609 z/OS WLM Update for z/OS 1.13 & 1.12, Horst Sinram
- RMF Performance Management Guide SC33-7992
- RMF Report Analysis SC33-7990
- Redbook Effective zSeries Performance Monitoring using Resource Measurement Facility (RMF) – SG24-6645
- Cheryl Watson's Tuning Letter 2004 No. 3 & 2012 No. 4 SMF CPU fields
- 2013 SHARE in Boston #13707 Introducing the IBM zBC12 and zEC12 GA2 Hardware, Harv Emery

Thank you!



Cheryl Watson Walker with partner, husband, and best friend Tom Walker in the Galapagos (www.tomandcheryltravels.me)



- Email: technical@watsonwalker.com
- **Website:** <u>www.watsonwalker.com</u>