Cheryl's Hot Flashes #13

Cheryl Watson

Session 2509; SHARE 104 in Anaheim

March 4, 2005

Watson & Walker, Inc.

home of **Cheryl Watson's TUNING Letter, CPU Chart, BoxScore** and **GoalTender**

Thought for the Day

"Debugging is twice as hard as writing
the code in the first place. Therefore,
if you write the code as cleverly
as possible, you are, by definition,
not smart enough to debug it."

Brian W. Kernighan

Agenda

- Survey Questions
- User Experiences
  - z990/z890 update, PDSE CPU Time, CPENABLE, MSO SDC, VWLC, APARs
- This SHARE
- Outstanding Questions
- 6-Month Update
  - APARs, Chargeback, z990 Sizing, zAAP Processors

Survey Questions

- Who is running z/OS.e?
- How many sites are running z/OS 1.5? z/OS 1.6? Earlier than z/OS 1.4?
- Who is using Variable WLC pricing?
- Who has activated IRD CPU management?
- Who is considering On/Off Capacity on Demand in the next year?
- Who is using zAAP processors?
  - Who is considering them?

Survey Questions

- Who is running WebSphere Application Server on z/OS?
- Who uses SMF type 30 record CPU time for chargeback, data center reporting, pseudo-chargeback, capacity planning, etc.?
- Who is doing heavy cryptographic work?
- Who is using VSAM RLS for CICS?
  - Who is considering it?
- Who is using Transactional VSAM?
  - Who is considering it?

z990/z890 Update

- This SHARE, two sites told me about a strange occurrence
  - Both applied maintenance to their z990 within the last few weeks (Driver 55, MCL089 in EC Stream J13486) – this applies to z890 also
  - Both sites discovered that the service units per second for the CPs were divided by four (speed constant and consumed service units)
  - A site can install the maintenance but won't see the effect until a POR
  - MCL103 provides slip traps to allow a fix on the fly; POR still needed for final fix
  - Info APAR **II14006** (3Mar05) provides more information

z990/z890 Update

- What will this affect?
  - SMF and RMF records will show one-fourth the service units that are expected
  - Billing programs will need to be changed if using service units
  - Jobs will take longer to meet the duration, so will stay in first period longer
  - Resource groups won't apply for a long time
  - WLM sysplex balancing will probably be wrong
  - The 4-hour MSU value appears to be correct
  - Any reporting of service unit consumption based on Type 72 data will be wrong
  - Capacity planning using the data will be wrong

PDSE CPU Time

- IBM recommends converting PDS libraries to PDSE libraries
- Recent problem occurred with CPU overhead in PDSE and some sites have converted back
- Solution is to manage the hiperspaces, not back off PDSE

PDSE CPU Time

- History
  - z/OS 1.3, APAR **OW53245** (Jul02) created a new SMSPDSE address space to replace SMXC and SYSBMAS
  - Prior to this address space, most of the PDSE CPU time was recorded in user's address space
  - Problem 1
    - SMSPDSE is not restartable – IPL required if problems occurred (hang, dead-lock, SOS)
    - Solution – z/OS 1.6 can create two: SMSPDSE (for LNKLST concatenations) and restartable SMSPDSE1 (for others); new parms created

PDSE CPU Time

- History
  - Problem 2
    - PDSE program object libraries were not being cached correctly in hiperspace (they would fill up hiperspace, but never be deleted)
    - APAR **OA06884** (9Jun04) corrected the problem by finding LRU members and deleting them
    - Algorithm for looping through hiperspace is based on IGDSMXxx parameters LRUTIME and LRUCYCLES; size of hiperspace on HSP_SIZE
    - The default values for these caused increasingly higher CPU times in SMSPDSE (after IPL, CPU time dropped back down)

PDSE CPU Time

- Site 1
  - Opened PMR due to excessive CPU time in SMSPDSE address space (13% to 21% of the machine)
  - IBM opened APAR **OA08991** (04Jan05) – *High CPU Utilization Occurs with UA10647*
  - APAR said to see ++HOLD data; working as designed; user should tune parameters
  - If you don't want to cache, then back out PTF (very difficult) or turn off hiperspace usage with HSP_SIZE(0) – both require an IPL
  - Thanks to **Mike Mayne**, Huntsville Hospital System

PDSE CPU Time

- Site 2
  - Same problem. Changed LRUTIME from default of 15 to 60 (at another site, this changed CPU usage from an average of 16% to an average of 4%). IBM suggests using 60.
  - Site 1 (Mike) tried LRUTIME of 60 and CPU consumption in SMSPDSE was reduced by 75%
    - But CPU consumption still steadily increased each day since the previous IPL

PDSE CPU Time

- Site 3
  - High CPU in SMSPDSE after fix for APAR OA06884
  - Backing off the PTF (which is difficult) reduced the CPU consumption to previous levels
- Site 4
  - Set HSP_SIZE to 64 MB, LRUSIZE to 60 and LRUCYCLES to 120 and found good results

PDSE CPU Time

- Site 5
  - Info provided by **Norman Hollander** of CA (Thanks!)
  - Installation started missing service agreements; one of problems was tracked back to some PDSE libraries that held thousands of members
  - Conversion back to PDS libraries reduced test program from CPU time of 20 seconds to .25 seconds
  - Norman: *"Since January of 2004, there have been 165 APARs against PDSEs; since January of 2005, there have been 75. The latest was on 2/18/2005."*

PDSE CPU Time

- Related APARs from Norman:
  - APAR **OA10071**, *Loop in IGWIFELE leads to hang in PDSE Library ,* 10Jan05
  - APAR **OA08818**, *PDSE Index Manager Serviceability Enhancement*, 04Nov04, HIPER
  - APAR **OA07881**, *Loop Creating SRE's (Suspend Resume Elements),* 12Nov04, HIPER

PDSE CPU Time

- Ten other sites have asked me about the problem this week, so it's common
- See Redbook **SG24-6106-01 - Partitioned Data Set Extended Usage Guide** (draft with a major update, was last updated last week and discusses these APARs)
  - *"Program objects, being housed in PDSEs are eligible to be managed by LLA/VLF as well as the PDSE hiperspaces. In general, a program object that is in use by more than one address space concurrently may benefit from the use of the PDSE hiperspace.* ***However, if there is only one user then use of hiperspace may result in a waste of real storage and/or CPU cycles."***

PDSE CPU Time

- More from Redbook **SG24-6106-01**:
  - *"There will not be any benefit from using PDSE hiperspace if there is only one user of the data set at any one time. Only heavily used PDSEs, and those that stay open long enough to take advantage of it, should use PDSE hiperspace member caching.* ***Therefore, in general, it is better to avoid using PDSE hiperspace unless there is a demonstrated improvement from using it."***
  - Redbook recommendation: LRUCYCLES=120, LRUTIME=60; HSP_SIZE=128 as starting place if CPU time is a problem

PDSE CPU Time

- Other notes found while researching this topic
  - APAR **OA10426**, *"LLA Refresh High CPU Utilization During Member Disconnect Processing,"* HIPER, R708, 15Feb05 refers to high CPU when frequent LLA refreshes are done on a PDSE library
  - APAR **OA09955**, *"IEBCOPY PDSE to PDSE Does Not Notify VLF of New / Changed Member,"* OPEN, 22Nov04
  - APAR **OA09162**, *"PDSE Program Object Libraries Not Being Cached in Hiperspace,"* 14Jan05
  - Search IBM-Main archives for 'SMSPDSE' **(bama.ua.edu/archives/ibm-main.html)**

PDSE CPU Time

- Recommendations
  - Don't back off PDSE program object libraries
  - Control use of hiperspaces if CPU time or storage is of concern
  - Options to reduce impact:
    - Turn off hiperspace usage totally (HSP_SIZE=0) – requires an IPL
    - Reduce amount of hiperspace with HSP_SIZE **– requires an IPL**
    - Reduce amount of LRU analysis (raise LRUTIME to 60, reduce LRUCYCLES to 120)
    - Use SMS storage class to identify only those PDSE libraries to be cached **(MSR of 3 or less will cause hiperspace caching)**
    - Last ditch – identify just the members you want cached and put them in separate library (we can't recommend this)
    - If using hiperspace caching and CPU creep occurs, submit a PMR to IBM – there is no outstanding APAR for the CPU creep problem

CPENABLE

- WSC **Flash10337**
  - *"z/OS CPENABLE Settings IBM 9672 / zSeries Processor"*, 25Feb05
  - New recommendation for z990 and z890 is (10,30) instead of (0,0) because IBM benchmarks showed an improved ITR of 0 to 5% when using (10,30). The best improvement is for the higher n-way processors.
  - This was mentioned in **Kathy Walsh's** session 2500 and **Riaz Ahmad's** session 2800
  - But several discussions during the week made me realize that people are confused….so….

CPENABLE

- Back to the basics
  - When data transfer is complete on an I/O device, the channel subsystem will attempt to find an idle CP or a busy CP that's enabled for interrupts
  - The subsystem will interrupt the CP, which brings in the I/O FLIH (first level interrupt handler) to handle the I/O
  - This may result in another SSCH being issued to that same device
  - The I/O interrupt handler then issues a TPI (test pending interrupt) to see if more I/O interrupts are waiting

CPENABLE

- Back to the basics
  - CPENABLE(10,30) was the original default value and meant the following:
    - Enable only the highest number CP for interrupts
    - If the number of I/O interrupts handled by TPI is over 30%, then enable another CP
    - If the number of I/O interrupts handled by TPI is less than 10%, then disable one of the CPs
    - From RMF:

CPU NUMBER	I/O TOTAL INTERRUPT RATE	% I/O INTERRUPTS HANDLED VIA TPI
1	61.58	1.97
2	25.57	4.66
3	42.64	2.70
4	813.2	10.11
TOTAL/AVERAGE	943.0	9.10

CPENABLE

- Recommendations
  - Realize that for LPAR and large I/O systems, the %TPI will tend to be higher, resulting in I/O elongation
  - In 1993, IBM recommended that all shared LPARs should use CPENABLE=(0,0)
  - My rule of thumb was that no more than 15-20% of I/Os should be handled via TPI, otherwise change CPENABLE
  - I found many sites with 95% production LPAR and 5% test, so many sites could still use (10,30)
  - In 1996, IBM found that CMOS had too much overhead with (0,0), so they changed the recommendation for CMOS (only) back to (10,30)
  - In this new flash, IBM is recommending (10,30) for z890s and z990s

CPENABLE

- RMF report with CPENABLE (0,0)

```
CPU          ...   I/O TOTAL        % I/O INTERRUPTS
NUMBER       ...   INTERRUPT RATE   HANDLED VIA TPI
0                  447.4             3.87
1                  447.0             3.86
TOTAL/AVERAGE      894.4             3.86
```

- Findings
  - (0,0) usually produces best ITR, but uses more CPU time
    - I/O interrupts are handled as soon as possible
    - Cache hits cause higher CPU time
  - (10,30) consumes least amount of CPU time
    - I/Os are delayed when LPAR isn't dispatched
    - Takes less CPU time because other work is not interrupted
    - On z990, ITR is improved, especially for high n-ways

CPENABLE

- See **Kathy Walsh's** session 2500
  - She discussed WSC's findings
  - Also strongly recommended that IRD sites apply APAR **OA05798** before changing to (10,30) or IRD may remove the only CP that was enabled for interrupts

MSO SDC

- History
  - Original IEAIPSxx service definitions for resource usage were commonly set at:
    - MSO (main storage occupancy) = 3.0
    - IOC (I/O coefficient) = 5.0
    - CPU (TCB time) = 10.0
    - SRB (SRB time) = 10.0
  - But when storage grew, MSO became too large and skewed results
  - So code was changed to allow setting MSO=.0001

MSO SDC

- **History**
  - I recommended setting the MSO SDC to 0 because I could never, ever, get the calculated storage values from MSO to match page-seconds, RMF type 72 storage measurements, or RMF MII storage frames (and it was still huge!)
  - IBM later recommended setting it to 0
  - But some consultants and users still believed in it

MSO SDC

- Recent Events
  - Customer asked **Barry Merrill** on the MXG listserver why the MSO units seemed to be wrong (he was using MSO-0.0001)
  - Barry researched it and found out from IBM that the code was wrong internally and truncation in the calculation resulted in all MSO values less than .0122 being set to 1
  - APAR **OA10641**, "*SMF30MSC MSO Coefficient Requires Scaling When Being Used in Calculations*," 1Mar05
  - Barry sent note to MXG listserver on 22Feb05 describing the situation

MSO SDC

- What to do now?
  - Set MSO to zero or simply refuse to use the MSO measurement for anything of value
  - Changing the MSO SDC changes the value used to determine duration
  - You can calculate the effect of changing the value of any SDC
  - This is documented in a paper I wrote in 1994 called *"Positioning for Goal Mode,"* and you can find it under "Articles" on our Web site
  - If you don't change duration and set MSO to 0, then work will stay in first period much longer

VWLC

- Variable Workload License Charges (VWLC)
  - Can save significant amounts of money from IBM
  - Here's one example for just the z/OS and CICS TS software budget:
    - Move was from two z900-101 machines - 476 MIPS or 82 MSUs (software cost was $1.21M/year)
    - Upgraded one of the machines to a 102 machine - 688 MIPS or 119 MSUs (45% growth)
      - PSLC software cost would be $1.62M/year
      - Full workload charges would be $1.27M/year
      - VWLC usage so far is 78 MSUs for $850K to $950K/year

VWLC

- That's great news, but there's also bad news:
  - Some installations are telling us that a few ISVs listed on IBM's Web site as supporting VWLC don't seem to have their heart in it
  - If you hear of some good or bad experiences, please let us (or IBM) know
  - In the example given, the IBM software reduction only offset part of the ISV increase
  - CA has received the most positive comments regarding support for VWLC – congratulations, CA!
  - BMC uses "Reference Pricing"

APARs

- SMF Type 30
  - OA09118 (SMF), OA09081 (SRM)
    - z/OS 1.6 – these APARs add service definition coefficients (e.g. CPU=10.0, SRB=10.0…) to SMF Type 30 records; closed Oct04
    - You can now convert back and forth between service units and CPU time without having to match to the type 72 records
    - Yippee!

APARs

- GRS/MIM
  - APAR **OA06611,** *GRS Exit Enhancement*, 29Sep2004
  - Improves GRS/MIM performance
  - Reduces pathlength by eliminating some of the prebatch exit calls
  - CA-MIM 4.6 supports this facility (need PTF for high CPU time)
  - Wells Fargo implemented parts of new APAR, plus CA-MIM updates and application changes to remove enqueue/dequeues to a data space – results are that CA-MIM/GRS performs better in z/OS 1.4 than in OS/390 R10

APARs

- WLM
  - APAR **OA08903,** *IWMRCOLL RCAECUSE Field May Contain Incorrect Values*, 6Oct04
  - If two jobs end at the same time, the CPU using counts could be wrong

- Uni-processor/VSAM RLS - Open Problem
  - SMSVSAM (used by VSAM RLS) may have a problem on a uni-processor
  - RLS has a danger of getting caught in a hang in a single processor environment due to the fact that it is a heavy SRB user. Apparently, a processor can be in a situation where it needs replies to requests but the replies themselves need SRB resources and the single processor has already used up whatever SRB resource there was. Therefore a second processor must be in the plex in order to have access to another SRB resource. Doc APAR may come out of this.
  - Both of these thanks to **Jim Peterson**, Washington Mutual

This SHARE – Kudos

- Kudos for integrity goes to **Janet Sun**
  - Janet gave an excellent presentation on *DFSMS Basics: ICF Catalog Management* (session 3018)
  - It was all about managing catalogs manually
  - She never even hinted that her company, Mainstar Software Corporation, has a product (Catalog RecoveryPlus) that does all that painful work more easily, more reliably, and washes the dishes as well! See **www.mainstar.com** and session 3062 for more info.
  - Excellent recommendation: every site should be taking catalog backups daily (or more often), running IDCAMS Examine and Diagnose daily, and creating disaster recovery backups daily

This SHARE – Session Handouts

- Great session handout references
  - 8221, *What's New in z/OS Language Environment*, **Thomas Petrolino**, IBM
  - 2892, *IBM eServer zSeries 990*, **Harv Emery**, IBM, great terminology collection (over 70 terms)
  - 1320, *The Good, the Bad, and the Really Ugly: DSNZPARM*, **William Favero**, IBM
  - 3042, *Transactional VSAM (DFSMStvs),* **Ruth Ferziger**, IBM (includes excellent paper)

This SHARE – z990 Service Times

- Understanding Differences Between z900 and z990 Service Time Measurements
  - Session 2574 by **Dr. H. Pat Artis**, Performance Associates, Inc. is a knockout!
  - Abstract: “With the introduction of the z990, IBM increased the precision of the timing measurements provided by the I/O subsystem from 128 to 0.5 microseconds. This session will explore the benefits and impacts of this change and answer the question *did my service time really increase?*”
  - Storage vendors and several large data centers that used his product, ***PAI/O Driver for z/OS***, were finding that service times on the z990 were .2 to .3 milliseconds higher than on the z900 –-- Argh! (they said)

This SHARE – z990 Service Times

- z900 and z990 Service Times
  - Pat determined that there was not a problem with the z990 channel subsystem, nor did the I/Os take longer on the z990. But instead, the new precision captured more of the service time (i.e. dropped less of the time due to truncation).
  - For an understanding of exactly why this is true, see Pat's session on the SHARE Web site and his full paper on his own Web site at www.perfassoc.com.

Outstanding Questions

- IMS/DB2 CPU Time
  - We reprinted an article of **Barry Merrill's** which said that for DB2-under-IMS transaction accounting, DB2-caused CPU time is included in the CPU time in the IMS 07 log record. The implication is that ALL DB2-caused CPU time is included in the 07 record.
  - A reader reported "Our experience is that main task CPU time is included, while parallel subtask CPU time is not included."
  - Does anybody know more about this?

Outstanding Questions

- z990 Crypto
  - Scary story – User had a job that ran for two hours on a z900 crypto facility, but ran for more than 24 hours on a z990
  - Has anybody had similar problems?
  - This was confirmed by another site yesterday; saw similar results (10 times longer); was told that it was due to lack of multiple paths; be very careful with sizing!
  - **SG24-6870-00 - zSeries Crypto Guide Update**, 11Apr03
    - New word? "If you move PCICC cards, those cards will be zeroized."

6-Month Update - APARs

- GRSQ Performance Enhancement
  - OA07975
    - 29Sep2004, HIPER
    - CICS dumps were taking 10-30 minutes and using over 10 million CPU service units
    - Usermod available to bypass collection of GRSQ information dropped elapsed time to 1-2 minutes and reduced CPU service unit consumption by 90%

Thanks to **Curt Thompson** of the Principal Financial Group

6-Month Update - Chargeback

- Chargeback Issues
  - With variable z990 CPU times and zAAP processors, repeatable (or even fairly consistent) chargeback costs are becoming more difficult to obtain
  - A change in the amount of storage can also dramatically change the CPU time usage of a job
  - Consider tier/range charging:
    - CPU time between 0-10 seconds = $1.50
    - CPU time between 10-60 seconds = $5.00
    - CPU time between 60-300 seconds = $10.00
    - Etc., etc.
  - Variability then won't affect the customers as often

6-Month Update – z990 Sizing

- zSeries LSPRs
  - New LSPR workloads for zSeries (primitives):
    - CB-L – Commercial Batch (similar to CBW2)
    - CB-S – Commercial Batch Short (similar to CB84)
    - OLTP-W – Online (similar to CICS/DB2)
    - OLTP-T – Online (similar to IMS)
    - WASDB – WebSphere
  - Important to use for sizing z990s

- ***TUNING* Letter** (2004 No. 2)
  - Whole issue on this topic will be made available on our Web site next week (see Sample Issues)

6-Month Update – z990 Sizing

- IBM has free tools for sizing
  - Marketing reps can use free internal tools for processor sizing studies (zPCR and CP2000)
  - Three new custom workloads:
    - Online: (OLTP-W + OLTP-T + WASDB) / 3
    - Other: (CB-L + CB-S) / 2
    - Low I/O: (.60 * CB-L) + (.20 * OLTP-W) + (.20 * WASDB)
    - Low I/O is defined as images that have fewer than 30 DASD SSCHs/Second per used MSU
  - Most installations are in a low I/O environment

6-Month Update – z990 Sizing

- This SHARE – My session 2537 – z990 Performance & Capacity Considerations
  - If you don't use zPCR or CP2000, you may be disappointed
  - IBM rep is required to run CP2000 prior to completing a contract for a z990
  - Our experience – the people at the site who confirm the capacity of a new machine have never heard of these tools
  - It's imperative that you get the tools and understand the results before confirming your hardware order

6-Month Update – z990 Sizing

- Kudos to IBM
  - LSPR Web site was updated Oct04 to include two new workloads: LOIO and TIMIX
  - **www-1.ibm.com/servers/eserver/zseries/lspr/**
  - TIMIX is “transaction intensive” mix
  - Special thanks to:
    - **Walt Caprice**
    - **Kathy Walsh**
    - **Gary King**

6-Month Update – z990 Sizing

- Software Pricing for z990s
  - Work with ISVs who don't understand what is happening
    - IBM provided about a 10% reduction in software pricing for z990s, and reduced the published MSUs by 10%
    - Note that this was done BEFORE IBM identified the low I/O condition
    - But many ISVs are still using average MIPS for pricing – wrong!
  - Determine costs BEFORE ordering an upgrade!

6-Month Update – z990 Sizing

- z990 Sizing Considerations
  - From our analysis, low I/O sites going from a z900 to a z990 are typically seeing about 10% underperformance if using the MIX MIPS workload. If you're using the low I/O customized workload, then it's very close.
  - From our experiences, the z990s produce the greatest variation of any processor we've ever seen
  - Based on IBM's zPCR sizing, you should be sure you can handle a capacity that is 5% less than you expect, even when using low I/O (can be hundreds of MIPS)
  - **Are z990s Underperforming? – The answer is NO, as long as you size correctly, tune your system, adjust your LPAR configurations and consider storage changes.**

6-Month Update – z990 Sizing

- Gartner publishes single number MIPS value
  - Six months ago they dropped their published MIPS by about 10%, which is probably to account for the low I/O effect (10% apparent underperformance if using MIX or average MIPS when moving from a z900 to z990)
  - We disagree with a single published MIPS value; especially now that customized workloads are more applicable than MIX MIPS
  - We think, instead, that ISVs should be using the published software MSUs for charging instead of MIPS (these happen to be 10% below hardware MSUs)
  - Because of low I/O, you may not see a drop in software costs on the z990s

6-Month Update – zAAP Processors

- Great new offering from IBM
  - Can significantly reduce costs for Java workloads (WebSphere, DB2, IMS and CICS)
  - A zAAP CP is $125K (about 25% the cost of a standard CP), and there are no software costs associated with it
  - You need to be very current: z890 or z990, z/OS 1.6, current releases of subsystems
  - Capacity planning and chargeback will need to be changed

6-Month Update – zAAP Processors

- Chargeback Considerations
  - E.g. – Current SMF type 30 CPU time doesn't include IFA time; instead, a new field is added for IFA CPU time
  - A job could run on a zAAP engine one day and on a standard CP the next day – do you charge less for the zAAP because it's costing you less, or do you charge the customer the same in order to be consistent?
  - If you charge the same, then the customers won't see any benefit from the new hardware

6-Month Update – zAAP Processors

- Capacity Planning Considerations
  - **Kathy Walsh's** recommendation: do separate reporting for each type of processor
  - If you are getting a zAAP to reduce software costs, then you need to run with IFACROSSOVER=NO and force ALL Java work to the zAAP (may not provide the best performance)
  - If you are getting a zAAP to simply increase current capacity at the lowest cost, then use defaults of YES for IFACROSSOVER and IFAHONORPRIORITY

6-Month Update – zAAP Processors

- z890 Considerations
  - A job could run on a zAAP engine at 365 MIPS for part of a run and a standard CP at 27-365 MIPS for part of a run – how do you charge?
- Important APAR **OA09113**
  - Increased CICS CPU times and response times when running Java on zAAP processors
  - HIPER, 25Jan05
- See **Kathy Walsh's** session 2825 this SHARE

6-Month Update – zAAP Processors

- Caution for reporting:
  - SMF type 30 CPU time does not include normalized IFA time (yet), but RMF Workload Activity Report's TCB time does include normalized IFA time
  - See APAR **OA10901** (OPEN) which adds zAAP normalization factor to SMF type 30s – Thanks to **Kathy Walsh** (session 2500) and **Bernie Pierce**

Questions

- **Email:**
  - cheryl@watsonwalker.com
- **Web site: www.watsonwalker.com**