Cheryl's Hot Flashes #5

Cheryl Watson

Session 2543, March 2, 2001

Watson & Walker, Inc.

publishers of 'Cheryl Watson's TUNING Letter'

http://www.watsonwalker.com

Technology • Connections • Results

Cheryl’s Hot Flashes

- Hardware Issues:
  - R10 LSPRs
  - MSUs Unveiled
  - z900 Notes
  - z900 Compression
  - Many CPs to fewer, faster CPs
- Software Issues:
  - DB2 Archive Logs on VTS
  - XISOLATE
  - COBOL
  - APARs
  - WLM Notes
- Good Reading

R10 LSPRs

- Review:
  - IBM runs benchmarks and publishes their results at: www.s390.ibm.com/lspr

  - Different workloads are measured: CICS/DB2, IMS, TSO, traditional batch (CB84), new batch (CBW2), and scientific (FPC1). FPC1 is usually indicative of SAS, C/C++, and Java work.
  - An average is derived from CICS, IMS, TSO, and CB84 called MIX.
  - MIX is the value used by most analysts to determine MIPS ratings.
  - Prior to the R10 benchmarks, the configuration for each of the workload runs was the same.

  - IBM (and I) have ***long recommended*** that you use workload ratings rather than the MIX rating.

R10 LSPRs

- R10 and z900 benchmarks:
  - TSO and CICS/DB2 were run in 64-bit mode, but the rest were run in 31-bit mode.
  - MIX was still calculated the same.
  - This new MIX is an even worse indicator of expected performance since YOU must run all workloads in either 64-bit or 31-bit mode.
  - This means that it's more important than ever to use the workload estimates or MIPS rather than a single number when using anything based off the new LSPRs.

R10 LSPRs

- There can be quite a difference between the workloads.
- From Cheryl Watson's January 2001 CPU Chart (MIPS per CP):

	MIX	TSO	CICS/DB2	IMS	CB84	CBW2	FPC1
9672-ZX7	151	151	152	158	143	189	161
2064-110	193	198	192	198	187	240	233
Change	28%	31%	26%	25%	31%	27%	45%

MSUs Unveiled

- MSUs stands for 'Millions of Service Units' and is calculated as:

MSU (rounded) = (su/sec * #_of_CPs*3600)/1000000

- As an example, the 2064-102 (a 2-way) has a su/sec value of 10891.7631 and an MSU rating of 78:

MSU (rounded) = (10891.7631 * 2 * 3600)/1000000 = 78

MSUs Unveiled

- But lately, the calculation doesn't match the ratings issued by the vendor. (It's used for marketing and pricing, after all!) As an example, the 2064-116 has a rating of 441 MSUs, but the calculation produces 468 MSUs - a difference of over 6%.
- Amdahl is even worse (since they calculate su/sec differently) and they have differences of up to 24%.
- So if you calculate MSUs from su/sec, understand that you won't be right on.

MSUs Unveiled

- When you start to use the IBM License Manager in z/OS on a z900, it's also important to understand this difference.

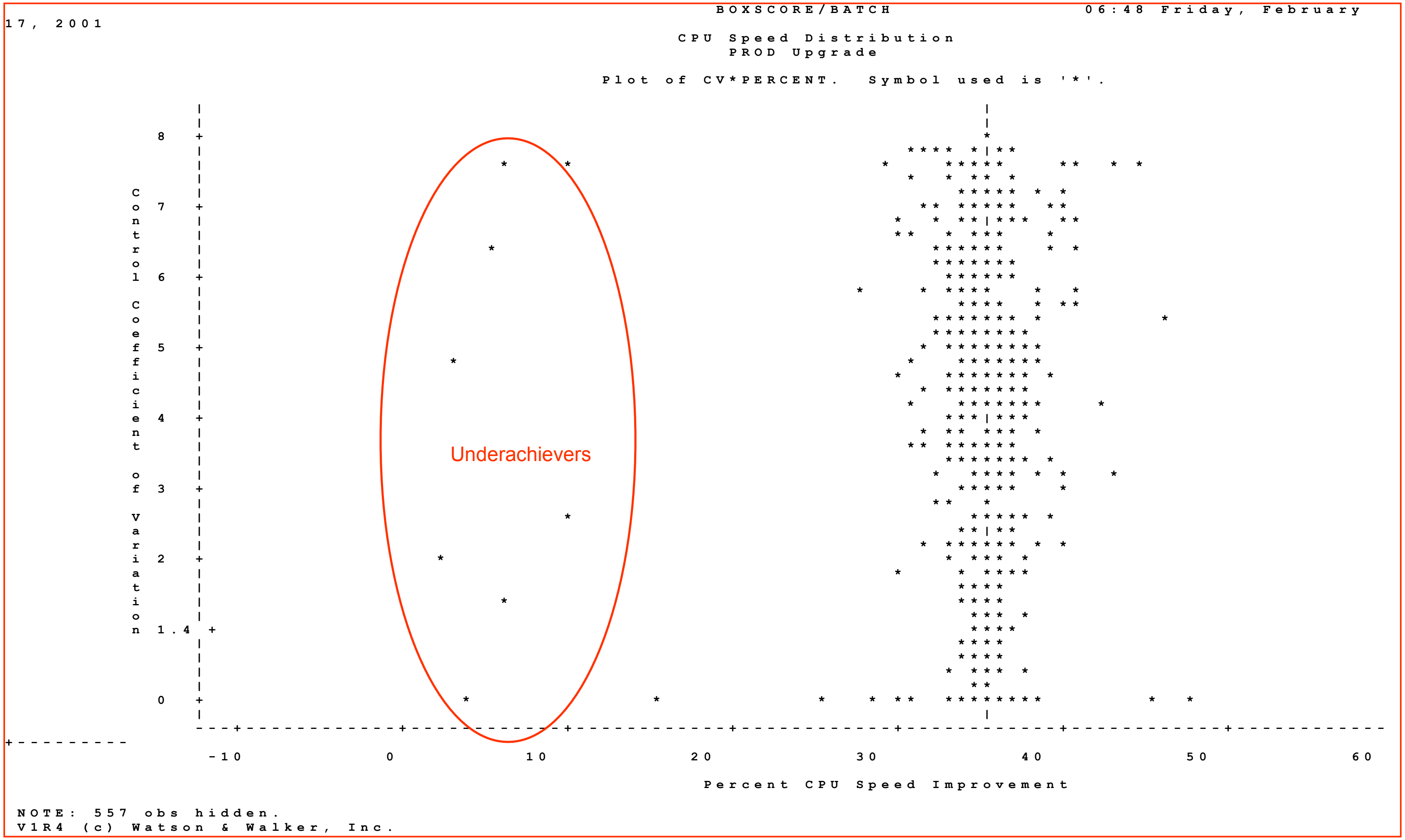
- On the z900s, you will specify a 'defined capacity' in MSUs and the software will compare that to the actual capacity used. They'll do the calculation based on the raw su/sec of the machine. They use the su/sec for the CEC, not what the LPAR might be using. Any relationship to marketing MSUs is purely coincidental.

z900 Notes

- About three weeks ago, I received a couple of notes from my customers indicating that their SAS runs on the z900 machines were not receiving the expected performance. Two had an under-capacity of 30%; one more on the MXG forum had a 15% underfall. See Cheryl's List # 50 at www.watsonwalker.com/archives.html.

- Early testing found some other programs also ran extremely poorly on the z900: Natural, the IMS log archiver DFSUARC0, and ADRDSSU (DSS). Up to 30% less than expected.

z900 Notes



```
                                        BOXSCORE/BATCH                    06:48 Friday, February
17, 2001
                                        CPU Speed Distribution
                                            PROD Upgrade

                                 Plot of CV*PERCENT.  Symbol used is '*'.

NOTE: 557 obs hidden.
V1R4 (c) Watson & Walker, Inc.
```

Session 2543 - Cheryl's Hot Flashes #5

z900 Notes

- In the z900, the data and instruction cache are separated. The cache width is 256 bytes. If you modify something in the area that has already been pulled into the instruction cache, it must be moved back over to the data cache. The invalidation takes some overhead.

- Natural - turned out to be a technique in the way a program was coded. Changing the program resolved the issue.

- No word yet on ADRDSSU.

z900 Notes

- IMS log archiver, DFSUARC0, is being worked on. Some of the commonly used routines were less than 256 bytes long, so data and instructions were in the same cache.

z900 Notes

- SAS Results
  - SAS had some prolog code that was less than 256 bytes where data was modified within 256 bytes of the instruction. This occurred for every function call in SAS and caused quite a bit of overhead on the z900.
  - SAS provided two zaps to change this logic, and (in most cases) solves the problem:

    For V8: **www.sas.com/service/techsup/unotes/SN/004/004291.html**

    For V6: **www.sas.com/service/techsup/unotes/V6/G/G952.html**
  - The good news is that these also improve SAS run times on the G5 and G6 processors by about 5%.

z900 Notes

- What this means to you:
  - If you run SAS on a G5, G6 or z900, be sure to install the latest SAS zaps.
  - If you're a vendor, consider whether your software has repetitive code that modifies itself or storage within 256 bytes of the instructions.
  - When upgrading to a new processor, be sure to check each of the vendor packages on the new processor. (As an example, in one of the sites I mentioned in my list, a customer was having problems with a Natural application. A change to the application code produced significant improvement on the z900.) You might have similar examples.

z900 Notes

- What this means to you:
  - For other considerations on coding for D-cache and I-cache, see the IBM-Main archives at bama.ua.edu/archives/ibm-main.html and search for a string of entries with 'z/Architecture I-cache' in the heading.
  - Use reentrant code whenever possible. Watch out for very high use, very tiny, routines (less than 256 bytes).
  - Evaluate your workloads after any processor change. See the technique described by **Al Coley** of WSC in session 2531 for determining stable and repeatable workloads (CPU per I/O and CPU per Transaction).

z900 Compression

- Hardware compression took a major performance hit starting with G4.
- From z900 Ivory Letter:
  - "Hardware data compression on the IBM z900 (Freeway) has been enhanced to provide significant performance gains over IBM 9672 G6 (turbo) processors. Depending upon the length of the records: a 3 to 4 times improvement has been demonstrated for compression and a 2 to 3 times improvement has been demonstrated for expansion. For both expansion and compression, the longer the record length, the higher the performance gain. With expansion, records with lower compression ratios have higher performance gains".
- I'd be interested in hearing about any benchmarks on this improved compression.

Many CPs to Fewer, Faster, CPs

- Long standing problem: moving work from many CPs to fewer, but faster, CPs. Problem comes up once again with z900s, but the phenomenon exists for any similar situation.
- Be very careful. Too few CPs can cause queuing for less important work and service levels may not be met.
- Phenomenon especially noticed in large CICS shops.

Many CPs to Fewer, Faster, CPs

- Problem is aggravated when people want to place too many LPARs on a single CP. Horrible things occurred to the site that moved 8 LPARs from a 6-way machine to a 2-way machine (even with 20% more MIPS).
- Old rule of no more LPs than twice the number of CPs still holds.
- See my article at our Web site (select 'Tuning Letter'; then Download Sample Issue; then read Focus article).

DB2 Archive Logs on VTS

- In my TUNING Letter (2000, No. 5) article on VTS, I made the comment that VTS isn't good for DB2 archive logs.

- Several people questioned that recommendation, since many sites use VTS for these logs.

- Here's the feedback from several users:
  - If logs are large there could be recall issues in recover situations. If a physical tape must be mounted, time to first byte will be much faster with real devices than virtual ones.

  - If you run out of tapes, you can't log, which stops DB2 users in their tracks.

DB2 Archive Logs on VTS

- We use it and have no major issues. We do have to account for the lack of it during D/R testing.

- Logs are typically placed into a facility that is DR'able. The VTS is not DR'able.

- We are archiving DB2 logs to virtual tape. We then have automation to copy the virtual tape to a real tape to be sent off site for disaster recovery. I think that disaster recovery would main be the reason to avoid virtual tape. I have noticed two big benefits 1) The log archive is quick - about 10 seconds for a 200MB log file. 2) When we had a tape outage, DB2 continued to archive normally; but we had to stop IMS before it ran out of log datasets.

DB2 Archive Logs on VTS

- We are using VTS for DB2 logs without any problems, and have been doing so for about five months. We cut one log on the local VTS and the other log is cut on a remote VTS for offsite backup. SMS rules are needed to ensure the allocations via dynalloc go to the correct VTS. The only problem we have today is if the remote link is down and we cannot cut the offsite logs.

- There is no disaster recovery for data in a VTS except for another VTS at a remote location or backup to tape which may cause heavy thrashing in the VTS depending on when the backup is taken and if the data has been de-staged down to physical tape (and what kind of VTS you have).

XISOLATE

- Situation: Most key data sets have duplicate copies for availability purposes (e.g. sysplex couple data sets, JES checkpoint). These should be kept on separate hardware for best availability: different channels, different directors, different DASD, and different power.
- **Don Chesarek** (IBM) developed a tool called XISOLATE to automate the checking process to ensure that critical data sets are on different DASD.
- This is useful for availability, performance, and business reasons.

XISOLATE

- You provide list of devices and data sets to batch job and run it at IPL and once a day. It produces a flat file with isolation information and SYSLOG message.

- XISOLATE is a non-warranted program, but code and documentation can be obtained at **ftp://ftp.software.ibm.com/s390/mvs/tools**.

- XISOLATE has identified availability exposures in every installation that has used it during its development.
- The documentation also provides good recommendations on which data sets you should have isolated in Appendix A.

COBOL

- Discontinuance of service:
  - Effective March 31, 2001, IBM is discontinuing: 5668-958 VS COBOL II (Compiler, Library, Debug Facility) MVS & VM. Replaced by -- 5688-197 COBOL for MVS & VM Release 2. (See announcement 997-070)
  - Effective December 31, 2001, IBM is withdrawing service from: 5688-197 IBM COBOL for MVS and VM. Replaced by -- 5648-A25, IBM COBOL for OS/390 and VM Version 2. (See announcement 900-220)

APARs

- From last SHARE:
  - APAR **OW44517** (Hiper) - After OW39132, IEAVEOR (IOC003) may not decrement ASCBTCBS (07/17/01) - Performance degradation and/or high LPAR management time (33 MIPS to 69 MIPS in uncaptured time for one site)
  - **Kathy Walsh** in session 2500 (WSC Performance Hot Topics) pointed out that APAR **OW46338** is PE of OW44517. So if you applied OW44517, then also apply OW46338 (R5+, 1/2/01).

APARs

- From **Michael Friske** of Fidelity Investments:
  - He recommends looking at catalog APAR **OW46456** - ECS Sysplex in wait after a Rebuild. (HIPER, 2/8/01).
  - Also see catalog/VSAM APARs for serious CAS SP252 fragmentation: **OW37739** (SMS 1.4, 1.5, 2/18/99), **OW45828** (SMS 1.3, 8/31/00), and **OW46165** (SMS 1.3-R10, 9/22/00). Occurs when using Jobcats, Stepcats, or when CATMAX is exceeded.

More APARs

- From **Norman Hollander**:
  - High CPU Utilization in TCPIP Started Task due to SNMP agent. Problem caused CPU utilization to push to 100% due to design problem with the SNMP agent. Two fixes have been created back to 2.7. APARs **PQ41211** (R5+, 10/4/00) and **PQ44493** (R5+, 2/2/01) have fixes available. APAR **PQ45543** is closed (1/20/01) but PTFs have not been issued yet.
- See WSC Flash10045
  - APAR **OW40167** provides new feature for DFSMSdss. Allows new keyword, CONSOLIDATE, to consolidate multiple data extents.

WLM Notes

- Compat mode disappears in 1Q02 - yeah!
  - The EWCP sessions on WLM migration reminded me that there are still sites that need to prepare for conversion to goal mode by getting their IPS/ICS structures organized. An article in my newsletter in 1995 addressed those issues. I'll post it on my Web site for everyone's use. Wait for a week, then see the article on positioning for goal mode at: **www.watsonwalker.com/articles.html**

- When proceedings come out, look at these sessions on migration:
  - 2514 - WLM Migration by Jim McCoy has good list of resources
  - 2513 - WLM Migration by Norman Hollander has good list of APARs
  - Other EWCP sessions on WLM migration: 2510, 2511, 2512, 2515, 2516, 2539

WLM Notes

- Largest amount of time is changing home-grown reports for management - take your time and do it on your test system

- I/O Priority Management
  - When I/O priority management is turned on, the velocity calculation is calculated as:

$$\frac{(\text{CPU Using} + \text{I/O Using})}{(\text{CPU Delay} + \text{CPU Using} + \text{Storage Delay} + \text{I/O Using} + \text{I/O Delay})}$$

  - Where:
    - I/O Using = Connect time + disconnect time
    - I/O Delay = Pend time + IOSQ

WLM Notes

- I/O Priority Management
  - This made no sense to performance analysts who have always believed that disconnect time (seek time and RPS misses) is a delay.
  - Because of the calculation, there were several problems: velocities were quite variable; the poorer the DASD response times, the more that jobs looked like they were running better than expected and they'd get knocked down by WLM (getting the reverse effect of what's needed); the amount of disconnect time was disproportional to other delays and dominated the velocity calculation.
  - I always recommended that people NOT turn on I/O priority management. So did WSC. See my TUNING Letter, 2000 No. 5, pg 55.

WLM Notes

- I/O Priority Management
  - Now with dynamic CHPID management, you MUST use I/O priority management.
  - So new APAR, **OW47667**, will be available to change the calculation to omit disconnect time from the I/O USING. It's currently OPEN. Congratulations, WSC!
  - Apply the APAR, recalculate your velocities, and THEN you can turn on I/O Priority Management.
  - Related APAR:
    - RMF APAR **OW43954** - "High Disconnect Times May Occur For Devices That Present Large Device Active Times"; this causes Ex Vel to show as 100%. (RMF R6-R9, 8/8/00).

WLM Notes

- From **John Arwe**:
  - In R10, the new classification by system name does not apply to batch, but R10 does add classification by "jesplex" name (JES2 MAS, JES3 &whatever) via the "subsystem collection name" attribute. This solves the problem of customers with multiple jesplexes in one sysplex not being able to classify by job class (because the same job class could be defined differently in each jesplex).

- Compat Mode APAR:
  - After adding RSU008 to an R8 system, one user had trouble meeting their SLAs. (90% of TSO users completing in first period instead of 99%). APAR **OW43091** (Enclave CPU time is missing; R3+, 6/1/00) is in error. The new APAR is **OW47558** (2//2/01, OPEN).

WLM Notes

- Problem: operators move work to different service class than intended and forget to move it back. This results in many jobs sitting in the wrong service class. How to move them all back into the correct service class?
  - From John Arwe: You have two choices...
    - 1. Issue a RESET RESUME on each individual job.
    - 2. Rest all with: MODIFY WLM, MODE=COMPAT and then MODIFY WLM, MODE=GOAL.

Good Reading

- Online manuals:
  - ESA Reference Summary, SA22-7209, in PDF form: **www.s390.ibm.com/ftp/books/os390/pdf/sa227209.pdf**

- Neat white papers at **www.ibm.com/support/techdocs**:

WP100185 - WSC WLM Migration Guide and Checklist - V2 expected by end of March

Questions

- Email: cheryl@watsonwalker.com
- Web site: http://www.watsonwalker.com