

z/OS Performance **HOT** Topics

SHARE, Summer, 2012

Session: 11463

Kathy Walsh

IBM Corporation



Trademarks

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

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries. Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed.

Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Agenda

- Processor Information
 - Relative Nest Intensity
 - CPU Measurement Facility
 - zPCR Latest Status Information
- Performance and Capacity Planning Topics
 - WLM
 - HiperDispatch
 - DB2 and zIIPs
 - Other
- New z/OS Performance Support Overview
 - z/OS 1.13
 - z/OS 1.12
- ▶ Addendum
 - Older APARs or Performance Information

www.ibm.com/support/techdocs

Techdocs - the Technical Sales Library

This site provides access to the Technical Sales Support organization's technical information databases. It gives you access to the most current installation, planning and technical support information available from IBM pre-sales support, and is constantly updated. You can browse or search these databases by date, document number, product, platform, keywords, etc.

New to Techdocs? Take a look at our [detailed introduction](#), which describes the document categories available (those listed on the navigation area on the left side of this page).

Rather than browse these categories, as a convenience you may enter a search of the full **Techdocs** database, or of any category you wish, here:

Search: Allow word variants
 for:
 Hits: Order by: "Fuzzy" search
 Include docs updated: [Help for Search](#)

Also available: our [Advanced search](#), where you can select documents based on various assigned document attributes.

Components of Workload Capacity Performance

■ Instruction Complexity (Micro Processor Design)

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

■ Memory Hierarchy or “Nest”

- Many design alternatives
 - Cache (levels, size, private, shared, latency, MESI protocol), controller, data buses
- Workload effect
 - Quite variable
 - **Sensitive to many factors: locality of reference, dispatch rate, IO rate, competition with other applications and/or LPARs, and more**
 - Net effect of these factors represented in “Relative Nest Intensity”
- **Relative Nest Intensity (RNI)**
 - **Activity beyond private-on-chip cache(s) is the most sensitive area**
 - **Reflects distribution and latency of sourcing from shared caches and memory**
 - Level 1 cache miss per 100 instructions (L1MP) also important
 - Data for calculation available from CPU MF (SMF 113) starting with z10

CPU Measurement Facility

- Hardware Instrumentation Facility available on z10 GA2, z196, and the z114
 - n New z/OS component - Hardware Instrumentation Facility (HIS)
 - n Generates SMF 113.2 records
 - n z/VM support for CPU MF Counters via **APAR VM64961**
 - **z/VM 6.1 and z/VM 5.4 on z10s and z196s**

- Capacity Planning Changes
 - n This data needs to be collected and used to select appropriate workload when doing capacity planning
 - n When available zPCR will use the SMF 113 data to select workload

- RNI calculations for z196 and z114 changed slightly (6/2012)
 - n Refinement based on customer measurements and LSPR workloads
 - n No change to z10 RNI calculation

Session 11600: CPU MF - 2012 Update and WSC Experiences, Wed, 4:30 PM

New CPU Measurement Education – PRS4922

- WSC hosted two Part Webinar to provide CPUMF Education
 - Part 1
 - Introduction and Overview of CPU MF
 - Implementation Details
 - Solicited customer data to be used in Part 2
 - Each customer sent in SMF 113 records and received a customized report
 - Part 2
 - Approx. 40 customers responded with data
 - Detailed information on usage and meaning of the CPUMF metrics
 - Overview and profile information on the provided data

- Presentations, and replays are available
 - <http://www-03.ibm.com/support/techdocs/atstr.nsf/WebIndex/PRS4922>

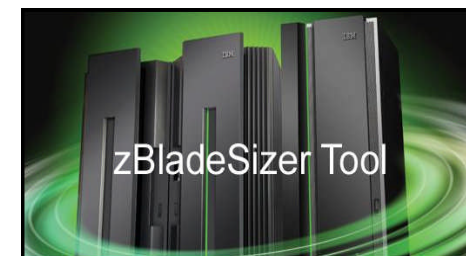
zPCR Latest Status

■ Version 7.9b (7/31/12) New Functions

- Algorithms concerning System z Multi-Book models have been enhanced
- Capacity results will likely differ from previous zPCR versions
 - zPCR algorithms for multi-book configurations are updated to recognize the order in which real CPs get allocated across books
 - GP, zAAP, and zIIP CPs are allocated upward, starting in the first book, IFL and ICF CPs are allocated downward, starting in the last book
 - An intersection of GP/zAAP/zIIP CPs with IFL/ICF CPs can occur in only one book
 - When such an intersection exists, zPCR applies the usual partitioning cost for CPs in that book only, while the cost for the CPs in the remaining books is less

Session 11599: zPCR Capacity Sizing Lab - Part 1 Introduction and Overview, Tue, 9:30 AM
11497: zPCR Capacity Sizing Lab - Part 2 Hands-on Lab, Thur, 11:00 AM

zBX Capacity Planning Sizing Tool



- IBM zBladeEXTR
 - Identify servers and time periods to consider in the analysis
 - Uses nmon data from AIX covering the time periods
 - Run by the installation to reduce amount of data needed to be sent to IBM
- IBM zBladeSizer
 - Provides optimal solution into a zBX Mod 002 including CPU, memory, network I/O, and disk I/O based on configuration constraints
 - Allows definition of availability and performance criteria for the environment
 - Supports
 - **POWER7™** blades
 - **System x®** blades - Linux servers only (Target availability: 4Q12)
- Supports new and existing zBX environments
- Contact your IBM Account team to arrange a sizing

Session 11601: zEnterprise eXposed! Part 1: Intro to zEnterprise Perf Mgmt, Thur, 9:30 AM

Session 11603: zEnterprise eXposed! Part 2: Experiences with the zEnterprise Unified Resource Manager, Thur, 11:00 AM

WLM – OA32298 – New Function

- New I/O Priority Manager feature in the IBM System Storage DS8700 and DS8800 provides favored processing of selected I/O requests by throttling other I/O requests which are less important
- Controlled by new IEAOPTxx parameter
 - STORAGESERVERMGT = YES | NO
 - Requires IO Priority Management in WLM Policy to be set to YES
 - Specifying YES removes control unit delay samples from service classes with velocity goals
 - May see change in achieved velocity when enabled if significant control unit queue delay (part of PEND time)
- Activated by WLM passing an I/O management field with the I/O request to the storage server
 - WLM sets this field for each address space and enclave and its value will be propagated by IOS when it sends the I/O request to the storage servers
 - WLM provides following information by service class period:
 - Response Time Goal: Goal achievement (derived from PI) and specified importance
 - Velocity Goal: Specified velocity goal and importance
 - System Goals: Not managed by Server I/O Priority Manager
 - Discretionary Goal: Always eligible to be throttled by Server I/O Priority Manager
- **White Paper: IBM System z DS8000 I/O Priority Manager**
<http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102074>

CICS Response Time Management Enhancement

- OA35428 (WLM) and OA35617 (SMF) - New Function
 - WLM manages CICS regions solely based on the transaction service class mix being executed
 - Depending on mix, WLM is generally not able to distinguish CICS TORs from AORs
 - CICS AORs typically run more CPU intensive work which do not require the same fast access as TORs
 - At higher utilization levels, >85% busy noticeable queue (QRMod) delays within the TORs can be recognized
 - Increases end-to-end response times of the CICS transactions, and reduces the throughput of CICS work
 - Of most concern in exclusively CICS environments with little postponeable work
 - More visible in a HIPERDISPATCH=YES environment
 - Number of logical processors in use is minimized in order to increase cache hits and thereby throughput of the work in the system

CICS Response Time Management Enhancement

- Possible circumvention:
 - Exempting CICS TORs from being managed using the goals of the transaction service classes
 - CICS TORs then managed towards the velocity goals
 - Drawback is transaction statistics not available for managing CICS work

- New Function:
 - New WLM Classification option *BOTH* for managing CICS TORs
 - TOR managed to velocity goals of the region but transaction completions are still tracked so WLM can manage CICS service classes with response time goals
 - Option *BOTH* should only be used for CICS TORs with noticeable queue delays at higher utilization levels
 - CICS TORs should be defined with a higher importance than the response time service classes for the CICS transactions
 - CICS AORs should remain at the default TRANSACTION

 - SMF 30 record updated with new bit, SMF30CRM, to indicate address space matched a classification rule of '*BOTH*'

CICS Response Time Management Enhancement

- New WLM BOTH classification support causes the function introduced in APAR OA34801 to be obsolete
 - OA34801 introduced support for IEAOPTxx parameter REPORTCOMPLETIONS
 - Specifies whether to allow transaction response time data to be reported for transaction service classes even if the subsystem work manager region is exempted from being managed towards the transaction goal
 - Default value is NO
 - Function is removed in z/OS V1R13 and higher
 - Function in z/OS V1R11 and z/OS V1R12 is still supported but not recommended and new WLM Classification option BOTH is the recommended solution

Session 10891: Workload Manager Update for z/OS 1.13 and 1.12, Mon, 3:00 PM

z/OS LDAP Tivoli Directory Services and WLM

- OA36644 – Provides ability to define performance goals for work within the LDAP server
 - Support is always active and a default service class needs to be defined in the LDAP classification subsystem
 - Failure to do this will result in LDAP work running in SYSOTHER
 - Creates a never ending independent enclave
 - New White Paper WP102151: Managing LDAP Workloads via Tivoli Directory Services and z/OS
<http://www.ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP102151>
 - Migration Action: LDAP subsystem name was used previously in WLM so verify the rules are as you intended

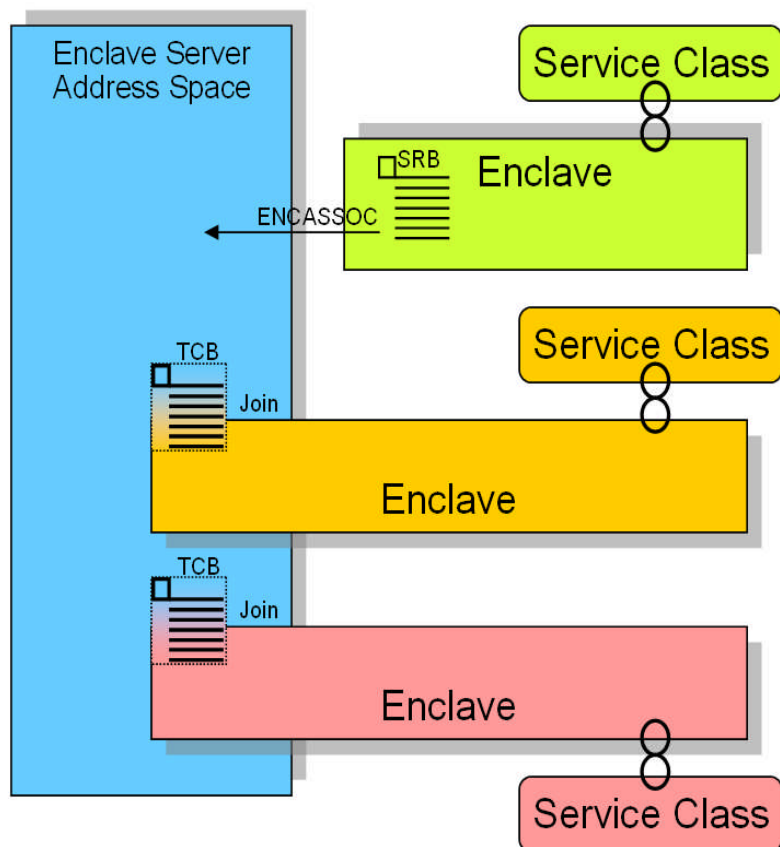
New Enclave Based Workloads

- Exploiters are continually making additional use of both independent and dependent enclaves
 - Watch release level migrations to ensure new ones are properly classified
 - Default classification is SYSOTHER
- Use a monitor like SDSF's enclave panel to check
 - LDAP
 - System Rexx
 - TCPIP IPSEC
 - GRS Monitor

| <u>SDSF ENCLAVE DISPLAY</u> | | | | | | | SYSD | ALL | LINE 1-12 |
|-----------------------------|------------|---------------|----------|-----------------|-----|-----|-----------------|-----|-----------|
| COMMAND INPUT ==> | | | | | | | | | SC |
| NP | NAME | <u>SSType</u> | Status | <u>SrvClass</u> | Per | PGN | <u>RptClass</u> | | |
| | 3C000F54BE | LDAP | INACTIVE | OPS_HIGH | 1 | | RLDAPPRB | | |
| | 5C000F54C0 | LDAP | INACTIVE | OPS_HIGH | 1 | | RLDAPDEF | | |
| | 60000F54BD | LDAP | ACTIVE | OPS_HIGH | 1 | | RLDAPGEN | | |
| | 2000000001 | STC | INACTIVE | SYSTEM | 1 | | | | |
| | 3800007944 | STC | INACTIVE | SYSSTC | 1 | | | | |
| | 2800000003 | STC | INACTIVE | SYSSTC | 1 | | RTCPIP | | |
| | 2C00000004 | TCP | INACTIVE | SYSOTHER | 1 | | | | |

Overview: WLM Enclave Server Address Spaces

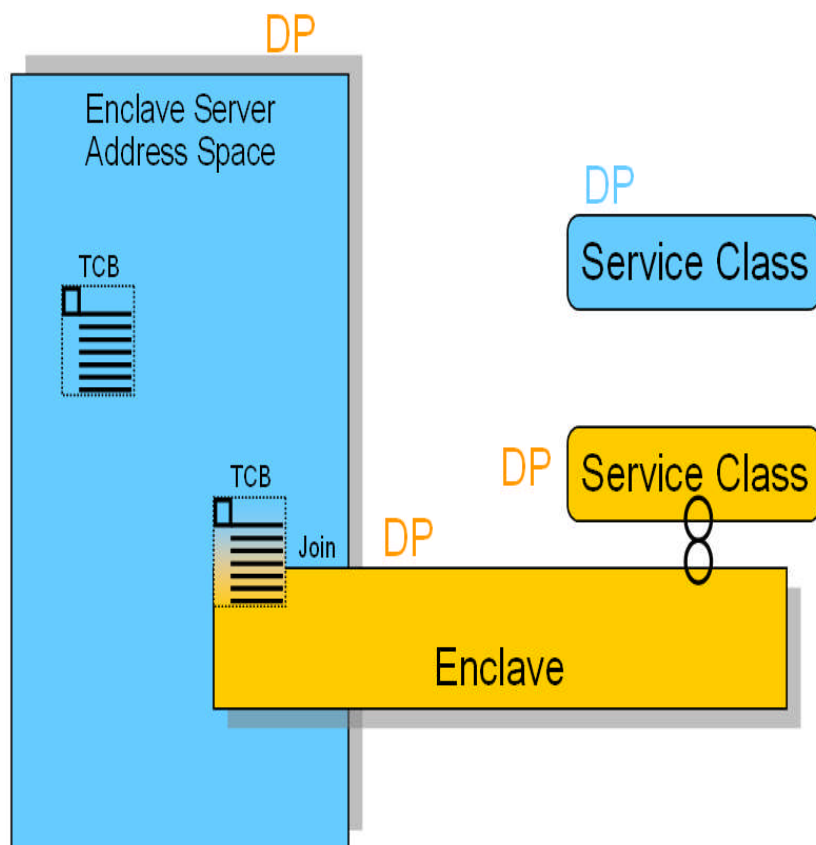
A Short Retrospective



- An address space becomes an enclave server when
 - An enclave SRB issues SYSEVENT ENCASSOC
 - A TCB of the address space joins an enclave
 - And specifies ENCLAVESERVER=YES (the default)
- Assumption (Programming Model)
 - All work being executed within the address space is related to enclaves
 - No significant amount of work (TCBs) executing in the address space not related to enclaves
- Enclave Server Management
 - CPU and I/O DP is derived from service class of most important enclaves
 - Meaning: No CPU and I/O management exists for server address spaces' service class
 - Storage management is done directly

Overview: WLM Enclave Server Management

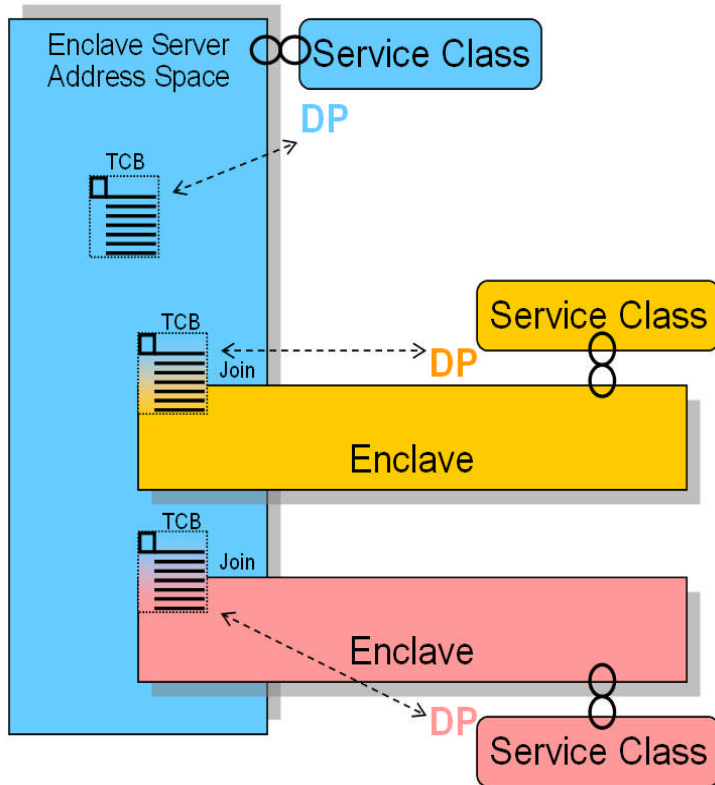
Is There a Possible Problem?



- What if the programming model isn't true?
 - There is significant work running in TCBs not associated with enclaves
 - Example: Garbage collection for a JVM (WAS) or common routines which provide service for the enclave TCBs
 - Is it sufficient to manage this work in the same way as the enclaves?

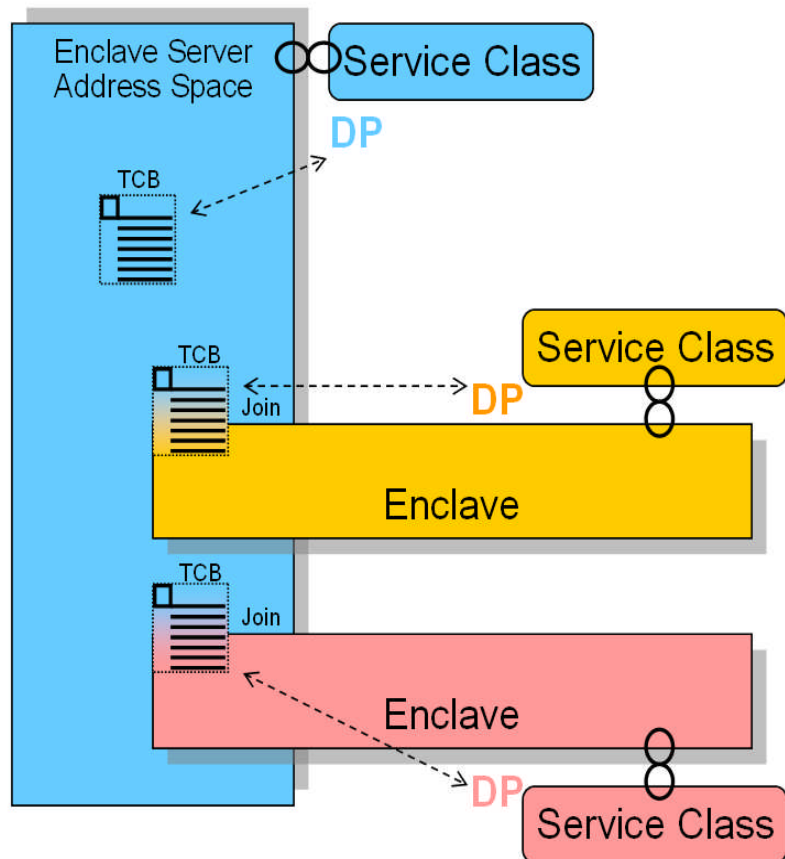
- What happens if no enclaves are running (note: this applies to queue servers only)
 - And the address space is swapped out?
 - A mechanism exists to swap in the address space but this mechanism assumes the swap in is only for a queue server task which wants to select a unit of work and then joins the enclave
 - And even if the address space stays swapped in?
 - The TCBs running within the address space just stay with the DP and IOP from the last enclave being associated with the address space
 - No CPU or I/O adjustment is performed

Solution: WLM Non Shell Server Management



- New OPT Parameter
 - ManageNonEnclaveWork = {**No**|Yes}
 - Default: No (no change to previous releases)
 - Causes everything in the address space which is not associated to an enclave to be managed towards the goals of the external Service Class of the servant address space
- Advantages
 - Enclave (Queue) server address spaces in which no enclave is running will be managed as usual address spaces
 - The importance and goal of the external service class for the address space now has a meaning
- Requires Policy Evaluation
 - Verify goal settings for server address spaces
 - This is a deviation from the past when the service class for servers was only important for startup, shutdown and recycles

Solution: WLM Non Shell Server Management



Disadvantages

- Non-enclave work is directly related to the enclave work – part of the cost of the application
- The OPT setting is LPAR wide which means the same operational method must be used by every environment
 - No separation of production from test
- Servant address space lower goal than enclave work
 - Housekeeping function runs lower than the workload which depends upon it
- Servant address space higher goal than enclave work
 - On a recycle the CPU intensive restart process can impact currently running work
 - May need additional service class for test regions
- Recommendation: Take the default of NO unless you have specific problems
 - Watch the servant swap rate

WLM APARs

- OA38280
 - Reduce sampling overhead by removing IVSK instruction
 - CPU savings seen in WLM address space
 - Related to number of PBDE control blocks in the LPAR

- OA36459
 - Not calculating the capacity used by vertical mediums and vertical low processors correctly

- OA38367
 - After a change in WAS classification rules all new trans in the changed service class timed out
 - Problem with not unbinding the old service class from the servants and not assigning the new service class queue to the servants

SRM APARs

- OA38529
 - During a pageable or Aux storage shortage a non-swappable address space may be set non-dispatchable
 - VERBX SRMDATA shows address space swapped due to a capping delay
 - Delay samples were incremented incorrectly and indicated capping
 - Change:
 - Correctly indicate via VERBX SRMDATA to indicate non-dispatchability due to pageable or Aux storage shortage
 - Capping delay sample will not include time marked non-dispatchable due to pageable or AUX storage shortage

WP101229 - HiperDispatch White Paper V2

- Updated for the z196 and other common questions
- Discussion of meaning of MVS Busy with HD=YES
- Lists factors which influence potential HiperDispatch improvement
 - Processor cache technology
 - Number of physical processors
 - Size of the z/OS partition
 - Logical : Physical processor ratio
 - Memory reference pattern
 - Exploitation of IRD Vary CPU Management
- Lists “Rule of Thumb” Expectations for z10 and z196
- Discusses importance of accurately set dispatch priorities for workloads

HiperDispatch

■ OA36054

- Beginning with z/OS 1.13 when running on an IBM zEnterprise z196 or z114 the default for Hiperdispatch will be YES

| Share of the partition - assumes 1.5 logical to physical ratio | Number of Physical CPs + zIPs + zAAPs | | | |
|--|---------------------------------------|-------|--------|--------|
| | <=16 | 17-32 | 33-64 | 65-80 |
| 0 <= share in processors < 1.5 | 0% | 0% | 0% | 0% |
| 1.5 <= share in processors < 3 | 2-5% | 3-6% | 3-6% | 3-6% |
| 3 <= share in processors < 6 | 4-8% | 5-9% | 6-10% | 6-10% |
| 6 <= share in processors < 12 | 5-11% | 7-13% | 8-14% | 8-16% |
| 12 <= share in processors < 24 | - | 8-16% | 10-18% | 11-21% |
| 24 <= share in processors < 48 | - | - | 11-21% | 12-24% |
| 48 <= share in processors <= 80 | - | - | - | 14-26% |

■ OA30476

- LPARs with >64 logicals must run with Hiperdispatch=YES

HiperDispatch APARs

- OA37736 – New Function
 - WLM uses free processor capacity of the CEC to determine if an LPAR should unpark one or more vertical lows
 - Apportionment didn't take into account relative weight of LPARs on CEC
 - Small LPARs could unpark and take free processor capacity leaving a larger LPAR unable to unpark vertical lows
 - LPARs may then not have proportional access, based on weight, to whitespace
 - Changes:
 - Now unpark an additional vertical low if CPU consumption of the partition is below its current weight **plus** newly introduced unused capacity share
 - Unparking only happens when no free processor capacity and the LPAR has CPU demand
 - SMF 99.12 and 99.14 record mappings will be published
 - 99.12 – Hiperdispatch Interval data
 - 99.14 – Hiperdispatch Topology data

Workload Promotion

■ OA30068

- PDSE hang can occur on various PDSE latches due to address space getting swapped out by WLM while holding PDSE resources
- PDSE contention couldn't be resolved by blocked workload support since latch holder was swapped out
 - Would require the address space to be made non-swappable
- PDSE latch processing is changed to add SYSEVENT ENQHOLD function to allow SRM to boost the service of the latch holder
 - Improves swapin recommendation value

WORKLOAD ACTIVITY

--PROMOTED--

| | |
|-----|-------|
| BLK | 0.000 |
| ENQ | 0.000 |
| CRM | 0.000 |
| LCK | 0.275 |
| SUP | 0.000 |

Check and understand why there are CPU times in any service classes

Local Lock Promotion - LCK

- In HiperDispatch while a WEB is suspended while holding a local/CML lock, z/OS will promote the WEB to dispatch priority x'FF'
- Done so work will run at a high priority until it releases its Local/CML lock
- Monitors will not display the WLM designated Dispatch Priority
- May introduce CPU delay to high importance workloads

z/OS 1.10 RMF Enhancements

- RMF Monitor III provides reports about spin and suspend locks
- Suspend Lock report will display the address spaces which hold locks and which are suspended
 - Spin Lock report will display how often global locks are held and who is spinning
- Specify LOCK | NOLOCK in RMF Monitor III
- Reporting of lock statistics is intended to help analyze lock contention in the system

z/OS 1.13 RMF GRS & Supervisor Delay Monitoring

- Collect and display system-wide contention information and contention information at an address space level in
 - Requires LOCK to be specified in RMF Monitor III
 - New SMF 72 subtype 5 record
 - New RMF XML Postprocessor Serialization Delay Report (SDELAY)

- New information
 - System Suspend lock types:
 - CMS
 - CMSEQDQ
 - CMSLatch
 - CMSSMF
 - Local
 - CML Lock Owner and
 - CML Lock Requestor
 - GRS lock types:
 - GRS Latch locks
 - GRS Enqueue Step
 - GRS Enqueue System and
 - GRS Enqueue Systems locks

Enhanced SMF 30 Reporting

- New SMF 30 function to provide enhanced reporting in the CPU Accounting section
 - z/OS 1.12 and above
- Highest percent of CPU time used by a single task in the address space in the interval, step, or job
 - SMF30_Highest_Task_CPU_Percent
- Program name associated with the task with the highest percentage of CPU time in the address space
 - SMF30_Highest_Task_CPU_Program

Session 11309: The Many CPU Fields of SMF, Tue, 1:30 PM

New DB2 / WLM Goal Setting Documentation

- Updated WLM classification information for DB2 started tasks
 - New section published June, 2012

- Recommends:
 - IRLMPROC in SYSSTC
 - ***ssnm*MSTR, *ssnm*DBM1, *ssnm*DIST and WLM-managed stored procedure address spaces** should be in a service class with a high importance and an aggressive velocity goal
 - May need to be marked CPU Critical
 - Recommends LPARs which run DB2 work at low dispatch priorities in environment with high CPU utilization to be more aggressive with blocked workload support
 - Protects environment by ensuring work holding DB2 locks gets CPU service
 - Recommends setting BLWLINTHD in IEAOPTxx from default of 20 to 3-5 secs

Session 11612: Workload Management of Transactional Workloads, Tue, 3:00 PM

Session 11500: Introduction to WLM Management of CICS and IMS Workloads, Wed, 8:00 AM

Session 11605: z/OS Workload Manager: What are you Thinking?, Tue, 4:30 PM

Predicting DB2 10 zIIP Eligibility for Prefetch and Deferred Writes

- In DB2 V9 workloads executing under 'non-preemptable' SRBs in DBM1 include:
 - Prefetch
 - Deferred write
 - Castout
 - Pseudo close
 - Write scheduling

- In DB2 10, prefetch and deferred write are executed under 'preemptable' SRB processing in DBM1
 - These workloads are marked as zIIP eligible
 - Represents the vast majority of the DB2 9 'non-preemptable' SRB processing in DBM1 address space
 - Will depend on workload mix

V9 and V10 Sample CPU Times from SMF 100

■ V9 example:

| CPU,TIMES | TCB TIME | PREEMPT SRB | NONPREEMPT SRB | TOTAL TIME | PREEMPT IIP SRB | /COMMIT |
|---------------------------------------|----------|-------------|----------------|------------|-----------------|----------|
| SYSTEM SERVICES ADDRESS SPACE | 6.186504 | 0.001054 | 13.465251 | 19.652809 | N/A | 0.001272 |
| DATABASE SERVICES ADDRESS SPACE(DBM1) | 0.292939 | 0.013103 | 41.596518 | 41.902560 | 0.000000 | 0.002712 |
| IRLM | 0.000128 | 0.000000 | 0.420866 | 0.420994 | N/A | 0.000027 |
| DDF ADDRESS SPACE | 0.003241 | 0.000000 | 0.005809 | 0.009051 | 0.000000 | 0.000001 |

■ V10 example:

| CPU,TIMES | TCB TIME | PREEMPT SRB | NONPREEMPT SRB | TOTAL TIME | PREEMPT IIP SRB | /COMMIT |
|---------------------------------|----------|-------------|----------------|------------|-----------------|----------|
| SYSTEM SERVICES ADDRESS SPACE | 5.388198 | 9.265621 | 1.226727 | 15.880545 | N/A | 0.001083 |
| DATABASE SERVICES ADDRESS SPACE | 0.144509 | 0.838230 | 0.022309 | 1.005048 | 25.877390 | 0.000069 |
| IRLM | 0.000060 | 0.000000 | 0.266525 | 0.266585 | N/A | 0.000018 |
| DDF ADDRESS SPACE | 0.086311 | 0.000138 | 0.004407 | 0.090856 | 0.000000 | 0.000006 |
| TOTAL | 5.619079 | 10.103988 | 1.519967 | 17.243034 | 25.877390 | 0.001176 |

Calculating zIIP Offload

- For the workload changed to pre-emptible SRBs 100% of the CPU time is zIIP eligible

| CPU,TIMES | TCB TIME | PREEMPT SRB | NONPREEMPT SRB | TOTAL TIME | PREEMPT IIP SRB | /COMMIT |
|---------------------------------------|----------|-----------------|------------------|------------|-----------------|----------|
| SYSTEM SERVICES ADDRESS SPACE | 6.186504 | 0.001054 | 13.465251 | 19.652809 | N/A | 0.001272 |
| DATABASE SERVICES ADDRESS SPACE(DBM1) | 0.292939 | 0.013103 | 41.596518 | 41.902560 | 0.000000 | 0.002712 |
| IRLM | 0.000128 | 0.000000 | 0.420866 | 0.420994 | N/A | 0.000027 |
| DDF ADDRESS SPACE | 0.003241 | 0.000000 | 0.005809 | 0.009051 | 0.000000 | 0.000001 |

- Calculate the potential zIIP busy by:

$$\text{NonPreempt SRB Time} / \text{Interval Time} = \% \text{ of a CP Eligible for zIIP}$$

Example of Potential Offload

- Eight Data Sharing Members during Online Peak

| Total DB2 Started Task Avg. GCPs used minus DDF Threads | DBM1 NonPreempt Avg. GCPs | Total DBM1 Avg. GCPs | Estimated Average zIIP CPs Offloaded in V10 | Estimated zIIP Offload as % of Total DB2 Started Tasks |
|---|---------------------------|----------------------|---|--|
| 1.08 | 0.49 | 0.57 | 0.49 | 46% |
| 1.56 | 0.66 | 0.80 | 0.66 | 42% |
| 1.01 | 0.41 | 0.50 | 0.41 | 41% |
| 0.62 | 0.52 | 0.52 | 0.52 | 82% |
| 0.30 | 0.26 | 0.26 | 0.26 | 86% |
| 0.10 | 0.08 | 0.08 | 0.08 | 77% |
| 1.08 | 0.51 | 0.58 | 0.51 | 47% |
| 0.84 | 0.29 | 0.49 | 0.29 | 34% |

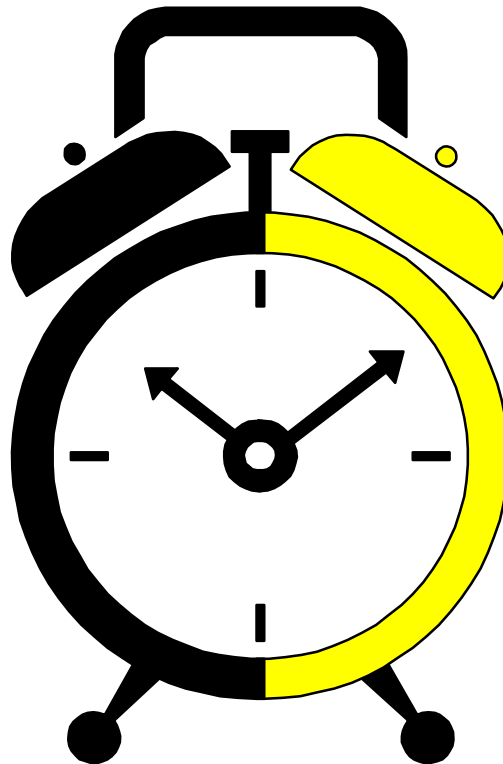
Example of Potential Offload

- Eight Data Sharing Members during Batch Peak

| Total DB2 Started Task Avg. GCPs used minus DDF Threads | DBM1 NonPreempt Avg. GCPs | Total DBM1 Avg. GCPs | Estimated Average zIIP CPs Offloaded in V10 | Estimated zIIP Offload as % of Total DB2 Started Tasks |
|---|---------------------------|----------------------|---|--|
| 0.57 | 0.42 | 0.45 | 0.42 | 73% |
| 1.04 | 0.71 | 0.76 | 0.71 | 68% |
| 0.58 | 0.36 | 0.38 | 0.36 | 63% |
| 0.28 | 0.22 | 0.23 | 0.22 | 81% |
| 0.23 | 0.14 | 0.21 | 0.14 | 62% |
| 0.12 | 0.10 | 0.10 | 0.10 | 82% |
| 0.62 | 0.44 | 0.46 | 0.44 | 70% |
| 0.52 | 0.24 | 0.41 | 0.24 | 45% |

Addendum

- Older flashes which should still be understood, or make you go Hmmm.
- APARs which are still causing issues, even though they are old.



TD105930 - Hiperdispatch and SAP DB Servers

Typically, customers run SAP DB Servers in their own z/OS LPARs and this leads to unique recommendations for Hiperdispatch for these LPARs

- z/OS R13: Use HiperDispatch (z10 and z196)
- z/OS R12 or earlier: Depends on the hardware configuration
 - z196 – Enable HiperDispatch
 - Main goal is to cluster each LPAR's work to a limited number of chips in the same book as this provides hardware cache benefits.
 - Installations with a very large LPAR processing SAP work may need to do more analysis on whether or not HiperDispatch should be used.
 - A "large" installation would be an LPAR which has more than five logical zIIPs defined AND the logical zIIPs are more than 80% busy.
 - For environments which meet this definition of "large", the amount of IIPCP time incurred in the environment may warrant running the environment with HiperDispatch disabled
 - z10 - Disable HiperDispatch
 - With the z10 the main goal is to cluster each LPAR's work to a specific book
 - For an SAP environment where the LPAR fits on one book, the benefits of HiperDispatch in an SAP environment are nominal
 - If the SAP LPAR is on a processor with purchased CPs (GCPs, zIIPs) on more than one book, additional analysis may be needed with respect to the HiperDispatch recommendation

z/OS 1.13 Performance Line Item Review

- Response Time Distribution for Velocity Goals
- RMF GRS & Supervisor Delay Monitoring
- SMF IFASMF DL to stop reading before end of logstream
- zFS Direct I/O Support
- RMF Integrated Ensemble Performance Monitoring

z/OS V1.13 - Response Time Distribution for Velocity Goals

- Currently WLM reporting does not provide a response time distribution (ended transactions) for workloads with velocity goals

- Need to provide a response time distribution for all transactional workloads, even if they have a velocity goal
 - ▶ More data to analyze workload behavior and to detect problems
 - ▶ Better support for migration of goal definitions to response time goals

- IWMRCOLL to be updated to provide a response time distribution for service class periods with an execution velocity goal

- RMF Postprocessor Workload Activity report will displays the new response time distributions

z/OS 1.13 RMF Report: Velocity R/T Distribution

```

REPORT BY: POLICY=POLICY01 WORKLOAD=STC SERVICE CLASS=STCDEF RESOURCE GROUP=*NONE PERIOD=1 IMPORTANCE=5
          CRITICAL =NONE
-TRANSACTIONS-  TRANS-TIME  HHH.MM.SS.TTT  --DASD I/O--  ---SERVICE---  SERVICE TIME  ---APPL %---  --PROMOTED--  ----STORAGE----
AVG      28.04  ACTUAL          16.629  SSCHRT  89.0  IOC   524944  CPU   1.453  CP    0.22  BLK   0.000  AVG   1143.34
MPL      28.04  EXECUTION          15.724  RESP   0.2  CPU   649332  SRB   0.277  AAPCP  0.00  ENQ   0.000  TOTAL 32056.00
ENDED    2      QUEUED           904  CONN   0.1  MSO   14840  RCT   0.010  IIPCP  0.00  CRM   0.000  SHARED 200.56
END/S    0.00  R/S AFFIN          0  DISC   0.0  SRB   123890  IIT   0.197          LCK   0.000
#SWAPS   100  INELIGIBLE          0  Q+PEND 0.1  TOT   1313K  HST   0.000  AAP    0.00  -PAGE-IN RATES-
EXCTD    0      CONVERSION          0  IOSQ   0.0  /SEC   1459  AAP   0.000  IIP    0.00  SINGLE  0.0
AVG ENC  0.00  STD DEV            0          ABSRPTN  52          IIP   0.000  BLOCK  0.0
REM ENC  0.00          TRX SERV  52          SHARED  0.0
MS ENC   0.00          HSP      0.0
    
```

GOAL: EXECUTION VELOCITY 20.0% VELOCITY MIGRATION: I/O MGMT 88.2% INIT MGMT 88.2%

| SYSTEM | RESPONSE TIME EX VEL% | PERF INDX | AVG ADRSP | --EXEC USING%-- | | | | EXEC DELAYS % | | | -USING%- | | --- DELAY % --- | | | % | |
|--------|--------------------------|--------------|--------------|-----------------|-----|-----|-----|---------------|-----|--|----------|-----|-----------------|-----|-----|-----|-----|
| | | | | CPU | AAP | IIP | I/O | TOT | | | CRY | CNT | UNK | IDL | CRY | CNT | QUI |
| *ALL | --N/A-- | 88.2 | 0.2 | 47.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | | 0.0 | 0.0 | 38 | 62 | 0.0 | 0.0 | 0.0 |
| SYSD | | 88.2 | 0.2 | 15.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.0 | | 0.0 | 0.0 | 40 | 60 | 0.0 | 0.0 | 0.0 |
| SYSE | | 88.6 | 0.2 | 17.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.0 | | 0.0 | 0.0 | 35 | 64 | 0.0 | 0.0 | 0.0 |

-----RESPONSE TIME DISTRIBUTIONS-----

| SYSTEM: SYSD | | | | | SYSTEM: SYSE | | | | |
|--------------------------|-----------|--------------------------|-----------|-----------|--------------------------|-----------|--------------------------|-----------|-----------|
| -----INTERVAL: 14.59.998 | | | | | -----INTERVAL: 01.22.123 | | | | |
| -----MRT CHANGES: 0 | | | | | -----MRT CHANGES: 1 | | | | |
| ----TIME---- | | -NUMBER OF TRANSACTIONS- | | | ----TIME---- | | -NUMBER OF TRANSACTIONS- | | |
| HH.MM.SS.TTT | CUM TOTAL | IN BUCKET | CUM TOTAL | IN BUCKET | HH.MM.SS.TTT | CUM TOTAL | IN BUCKET | CUM TOTAL | IN BUCKET |
| < 00.00.00.200 | 581 | 581 | 94.2 | 94.2 | < 00.00.00.300 | 581 | 581 | 94.2 | 94.2 |
| <= 00.00.00.240 | 584 | 3 | 94.7 | 0.5 | <= 00.00.00.360 | 584 | 3 | 94.7 | 0.5 |
| <= 00.00.00.280 | 586 | 2 | 95.0 | 0.3 | <= 00.00.00.420 | 586 | 2 | 95.0 | 0.3 |
| <= 00.00.00.320 | 586 | 0 | 95.0 | 0.0 | <= 00.00.00.480 | 586 | 0 | 95.0 | 0.0 |
| <= 00.00.00.360 | 588 | 2 | 95.3 | 0.3 | <= 00.00.00.640 | 588 | 2 | 95.3 | 0.3 |
| <= 00.00.00.400 | 591 | 3 | 95.8 | 0.5 | <= 00.00.00.600 | 591 | 3 | 95.8 | 0.5 |
| <= 00.00.00.440 | 592 | 1 | 95.9 | 0.2 | <= 00.00.00.660 | 592 | 1 | 95.9 | 0.2 |
| <= 00.00.00.480 | 592 | 0 | 95.9 | 0.0 | <= 00.00.00.720 | 592 | 0 | 95.9 | 0.0 |
| <= 00.00.00.520 | 593 | 1 | 96.1 | 0.2 | <= 00.00.00.780 | 593 | 1 | 96.1 | 0.2 |
| <= 00.00.00.560 | 596 | 3 | 96.6 | 0.5 | <= 00.00.00.840 | 596 | 3 | 96.6 | 0.5 |
| <= 00.00.00.600 | 596 | 0 | 96.6 | 0.0 | <= 00.00.00.900 | 596 | 0 | 96.6 | 0.0 |
| <= 00.00.00.800 | 599 | 3 | 97.1 | 0.5 | <= 00.00.01.200 | 599 | 3 | 97.1 | 0.5 |
| <= 00.00.01.600 | 604 | 5 | 97.9 | 0.8 | <= 00.00.02.400 | 604 | 5 | 97.9 | 0.8 |
| > 00.00.01.600 | 617 | 13 | 100 | 2.1 | > 00.00.02.400 | 617 | 13 | 100 | 2.1 |

z/OS 1.13 SMF Logstream Support

- Currently IFASMFDL will always read until the end of the logstream regardless of specified end date and time
- Two new options:
 - SMARTENDPOINT
 - First introduced with DUMP option in z/OS 1.12 with APAR OA31737 and OA34374
 - z/OS 1.13 extends support to ARCHIVE and DELETE
 - SMARTEPOVER(XXXX)
 - Specifies a value between 0000 and 0200 (2 hrs)
 - Default is 0200
 - SMARTEPOVER is added to SMARTENDPOINT to determine logical end point

z/OS 1.12 Performance Items

■ WLM Enhancements

- ▶ WLM Managed Initiators will consider the impact of resource group maximums when starting initiators
 - SMF 99 records updated to show reason for not starting
- ▶ Improve Discretionary Work Throughput
 - Run discretionary work for a longer period of time before dispatching other discretionary work, while still interrupting it after short periods for non-discretionary work

■ RMF Enhancements

- ▶ RMF changed to be able to read SMF records directly from SMF log stream improving ability to run reports with current data
- ▶ Include information in the CPU Activity Report about how many units of work are running or waiting for a processor (CP, zIIP, or zAAP)
 - Same information is added to SMF Type 70 records

z/OS 1.12 Performance Items

- Shutdown and Restart Improvements
 - ▶ Address spaces allocating large numbers of data sets (e.g. DB2, batch) should see substantial reductions in shutdown and restart times
 - ▶ Changing subsystem initialization from serial to parallel for initialization routines listed in IEFSSNxx and a new BEGINPARALLEL keyword
 - ▶ XCFIPL time improved when using very large sysplex couple data sets

- RAS Enhancements
 - ▶ Improve capture performance for SVC dumps with substantial amounts of data on Auxiliary Storage
 - Internal IBM laboratory tests reduced capture time by over 60%
 - ▶ SADUMP will better prioritize data capture for address spaces, and dump a number of system address spaces first irrespective of their ASID numbers
 - Capture data needed most to diagnose system problems with a partial dump
 - Allow specification of additional address spaces to be added to the predefined list using a new ADDSUMM option
 - [z/OS Best Practices: Large Stand-Alone Dump Handling Version 2](http://www-03.ibm.com/support/techdocs/atmastr.nsf/WebIndex/TD103286)
<http://www-03.ibm.com/support/techdocs/atmastr.nsf/WebIndex/TD103286>

z/OS 1.12 Enhanced Reporting of Work Units

- New in-ready distribution of work units provides a more detailed view of the CPU demand than the in-ready distribution of address spaces
- Number of work units is presented per processor type (CP, zAAP, zIIP)
- Data is added to the SMF 70 records

```

z/OS V1R12                                SYSTEM ID SYSD
                                           RPT VERSION V1R12 RMF
SYSTEM ADDRESS SPACE AND WORK UNIT ANALYSIS
-----NUMBER OF ADDRESS SPACES-----
QUEUE TYPES                MIN          MAX          AVG

IN                          73          74          73.4
IN READY                    6           9           8.8

OUT READY                   0           0           0.0
OUT WAIT                    0           0           0.0

LOGICAL OUT RDY             0           0           0.0
LOGICAL OUT WAIT           24          25          24.6

ADDRESS SPACE TYPES

BATCH                       10          10          10.0
STC                         85          85          85.0
TSO                         1           1           1.0
ASCH                        0           0           0.0
OMVS                        2           2           2.0

-----NUMBER OF WORK UNITS-----
CPU TYPES                MIN          MAX          AVG
CP                       5           60          9.3
AAP                      0           0           0.0
IIP                      0           2           0.6
    
```