SHARE, August 2011, Orlando

WLM Top Ten Things That Confuse You the Most!

Glenn Anderson, IBM Technical Training

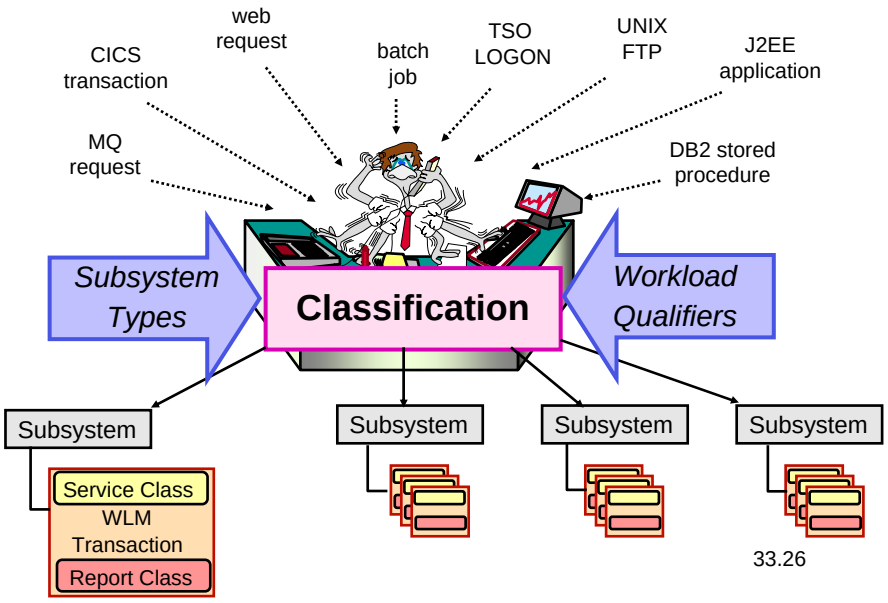
Session 10007

WLM Top 10 Things That Confuse You the Most!

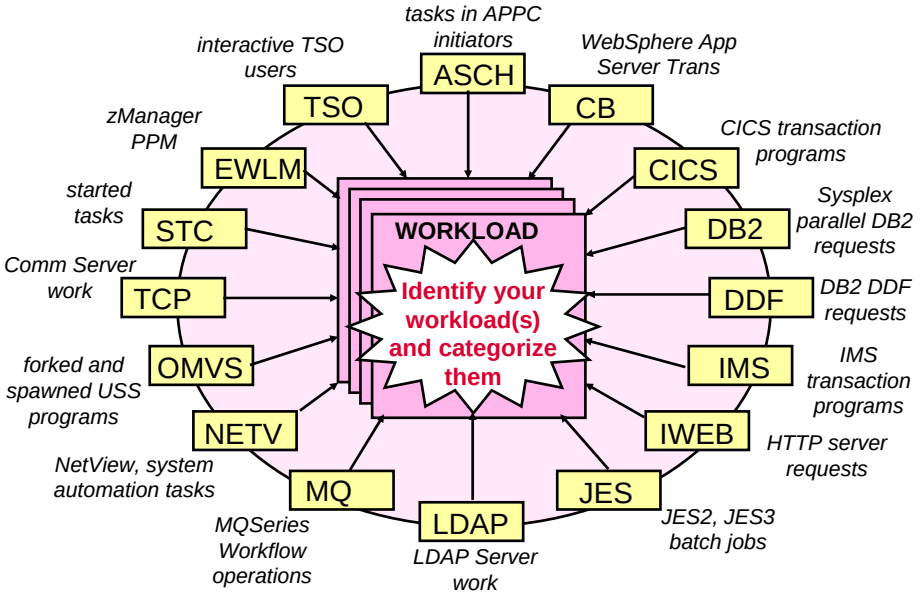
1. How does work get in the correct Service Class?
2. What is the EWLM subsystem all about?
3. What is the Right Number of Service Classes?
4. How Does WLM Manage Work?
5. How is Velocity calculated?
6. What are Enclaves?
7. What is Enclave Server Management?
8. What is Blocked Workload support?
9. How do the new Discretionary controls work?
10. What is the difference between IRD and Hiperdispatch?

1. How Does Work get in the Correct Service Class?

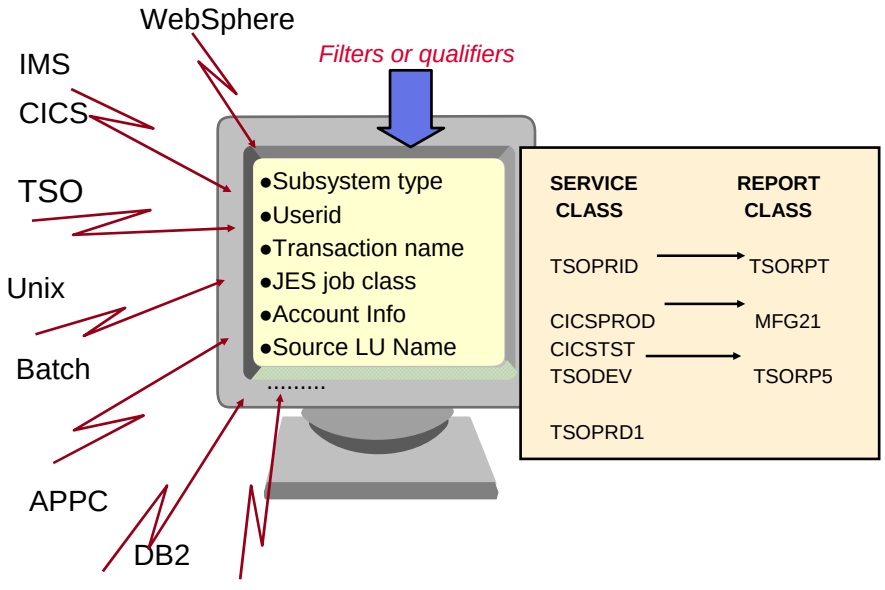
The classification process



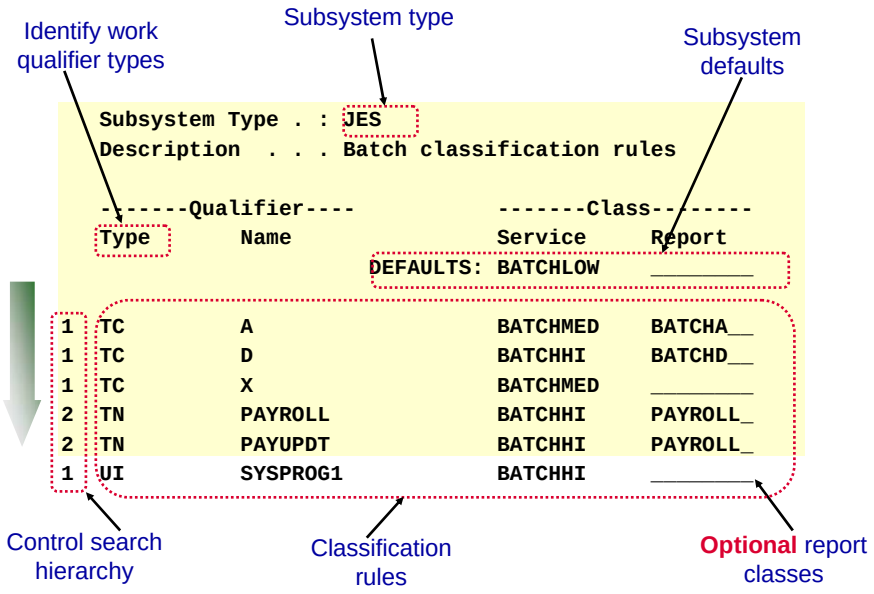
Service definition workload types



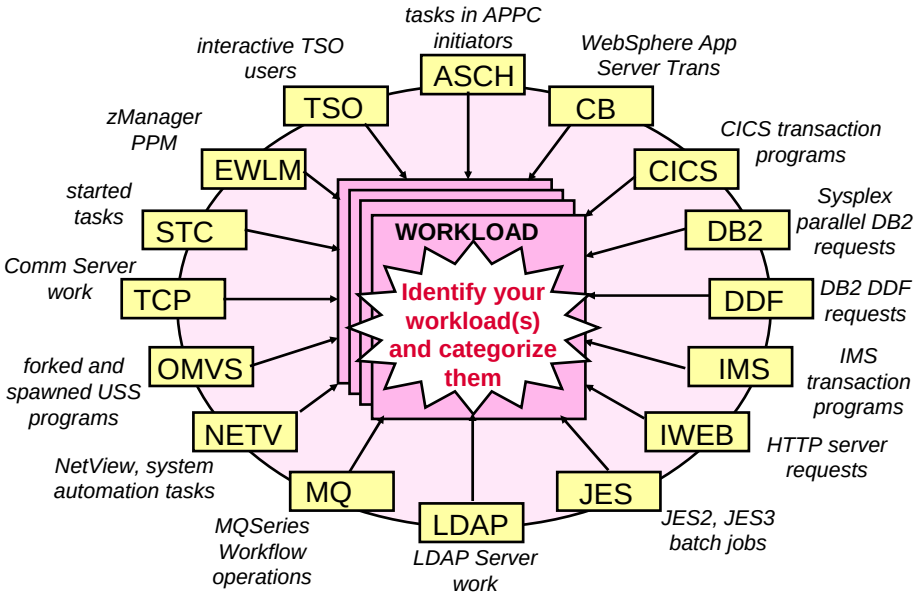
Classification rules



Example of batch classification rules



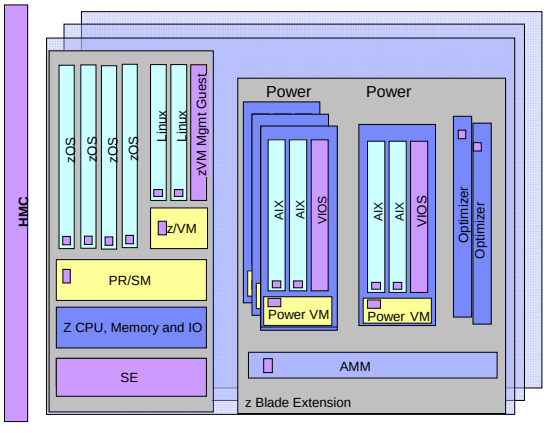
Service definition workload types



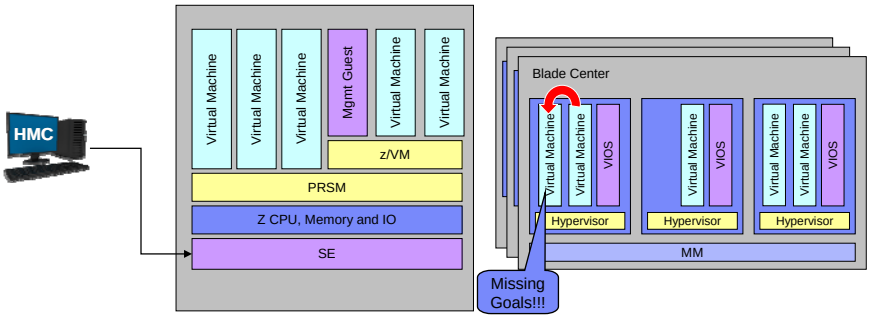
2. What is the EWLM Subsystem All About?

zEnterprise and Unified Resource Manager

- zEnterprise Unified Resource Manager (zManager)
  - HMC is management console
  - Ensemble-Wide scope of responsibility
  - Hardware configuration and operational control
  - Virtual server life cycle management
  - Virtual network and storage provisioning
  - Energy Management
  - Goal-oriented performance management



Managing Resources across Virtual Servers on P7 blade



- Manage resources across virtual servers to achieve workload goals
  - Detect that a virtual server is part of Workload not achieving goals
  - Determine that the virtual server performance can be improved with additional resources
  - Project impact on all effected Workloads of moving resources to virtual server
  - If good trade-off based on policy, redistribute resources
  - Initially support CPU management

Co-operative management with z/OS WLM

- z/OS provides differentiated service to PPM classified work
- Transaction coming to z/OS needs to be ARM instrumented
- WLM service definition needs to map PPM service classes to z/OS WLM service classes
- PPM service class associated with transaction is used by WLM to classify work unit to a different WLM service class.
- WLM manages the resources based on the goal assigned to this specific service class.

WLM support for Unified Resource Manager

- z/OS V1R12 introduces WLM functionality level LEVEL025 to support zManager
- zManager Service Classes can be classified to WLM service and report classes by specifying classification rules for subsystem EWLM
  - Work qualifier type ESC (EWLM service class name) is used to correlate zManager service classes with WLM service or report classes

```
          --------Qualifier--------                  -------Class--------
Action    Type       Name     Start                  Service     Report
                                          DEFAULTS:  ________    ________
____   1  ESC        Booking  ___                    ________    ________
____   2    ESC        System   9                    ________    ________
____   3      ESC        GoldServ 15                 ________    ________
____   4        ESC        ice      23               SERVCLS3    ________
```

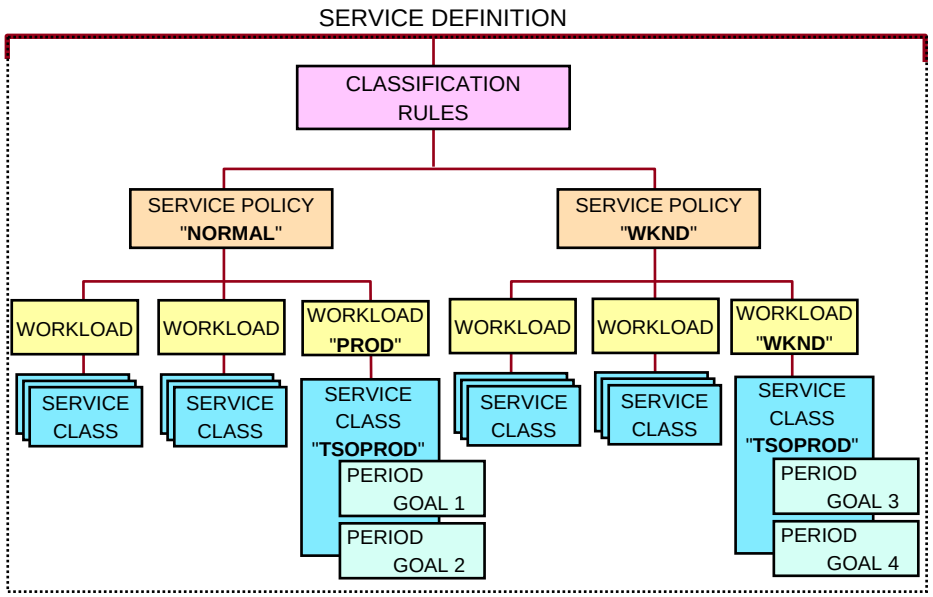
3. What is the Right Number of Service Classes?

The Correct Question Is:

What is the right number of active (in an LPAR) Service Class periods.

(Answer: 20-35)

Service definition hierarchy



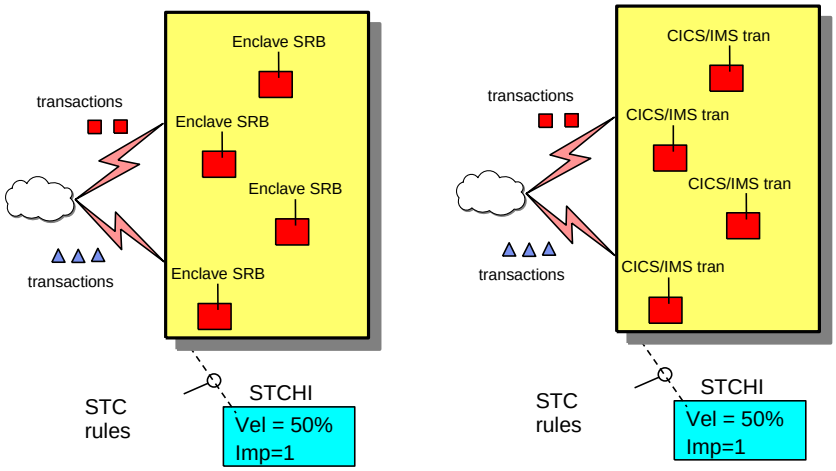
4. How Does WLM Manage Work?

What is a WLM Transaction?

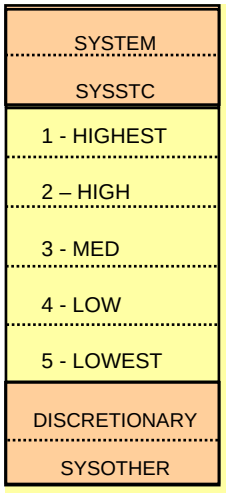
- **A WLM transaction represents a WLM "unit of work"**
  - basic workload entity for which WLM collects a resource usage value
  - foundation for statistics presented in workload activity report
  - represents a single subsystem "work request"
- **Subsystems can implement one of three transaction types**
  - **Address Space:**
    - WLM transaction measures all resource used by a subsystem request in a **single address space**
    - Used by JES (a batch job), TSO (a TSO command), OMVS (a process), STC (a started task) and ASCH (single APPC program)
  - **Enclave:**
    - Enclave created and destroyed by subsystem for each work request
    - WLM transaction measures resources used by a single subsystem request across **multiple address spaces and systems**
    - Exploited by subsystems - Component Broker(WebSphere), DB2, DDF, IWEB, MQSeries Workflow, LDAP, NETV, TCP
  - **CICS/IMS Transactions**
    - Neither address space or enclave oriented - special type
    - WLM transaction measures resource used by a single CICS/IMS transaction program request

The WLM View

Address Spaces, and the transactions inside

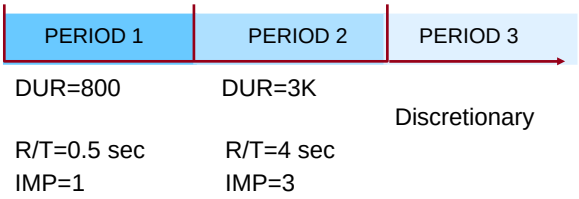


Importance

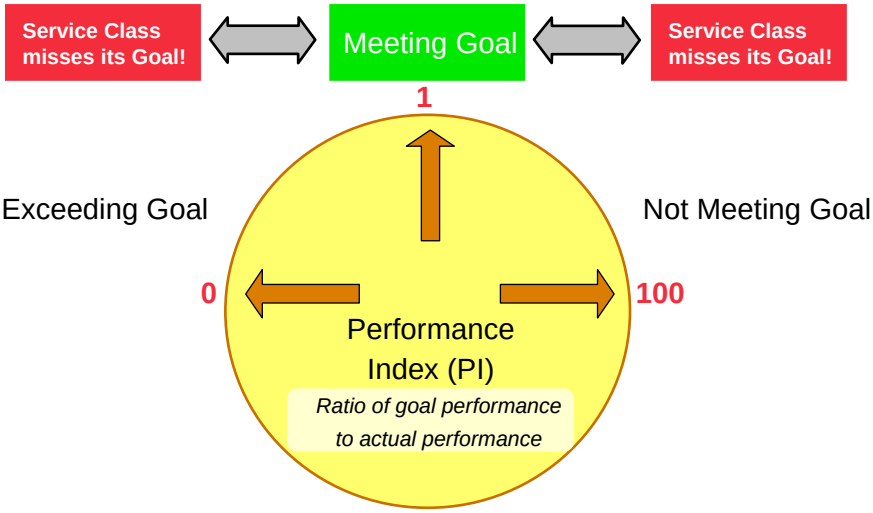


Workload Importance

- Importance is relevant when system is overloaded
- WLM uses it to decide which workload goals are most important to satisfy
- Generally importance decreases across multiple periods



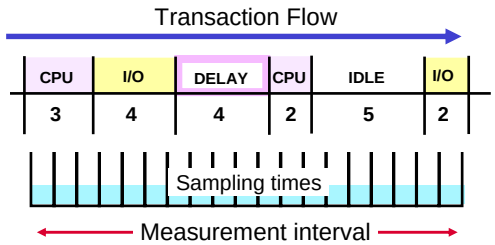
Performance index



- Separate PI is calculated for each service class period, across all sysplex hosts
- Different calculations produce a comparable value from different goal types

5. How Is Velocity Calculated?

Velocity goals



- Delayed I/O requests queued by goal achievement, not DP
- I/O Using
  - Includes non-paging DASD I/O only
  - Device connect
- I/O Delay
  - IOS queue
  - Subchannel pending
  - CU queue

$$\text{Velocity} = \frac{\text{(CPU Using + I/O Using) x 100}}{\text{CPU Using + I/O Using + WLM Delay*}}$$

$$= \frac{11 \times 100}{11 + 4} = 73\%$$

*Delay = CPU Delay + I/O Delay + Paging Delay + MPL Delay + A/S Delay

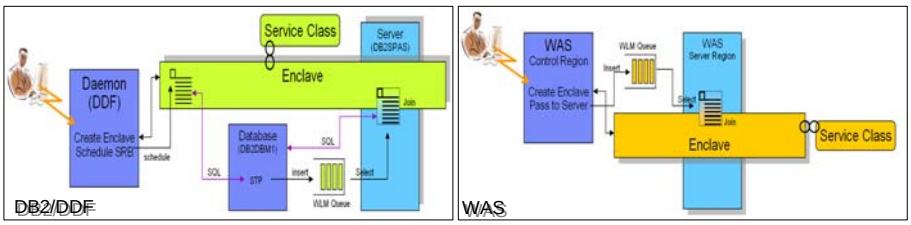
Response Time Distribution for Velocity Goals

RMF WLMGL Enhancement

```
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                                IIP    0.000                               BLOCK        0.0
REM ENC  0.00                                        ABSRPTN    52                                               SHARED       0.0
MS ENC   0.00                                        TRX SERV   52                                               HSP          0.0

GOAL: EXECUTION VELOCITY 20.0%     VELOCITY MIGRATION:   I/O MGMT  88.2%     INIT MGMT 88.2%
          RESPONSE TIME EX   PERF  AVG   --EXEC USING%--  -------------- EXEC DELAYS % -----------  -USING%-  --- DELAY % ---     %
SYSTEM                    VEL% INDX ADRSP  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 -----INTERVAL: 14.59.998 -----MRT CHANGES:  0 ---     SYSTEM: SYSE -----INTERVAL: 01.22.123 -----MRT CHANGES:  1 ---
   ----TIME----  -NUMBER OF TRANSACTIONS-  ------PERCENT------        ----TIME----  -NUMBER OF TRANSACTIONS-  ------PERCENT------
   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
```

6. What are Enclaves?

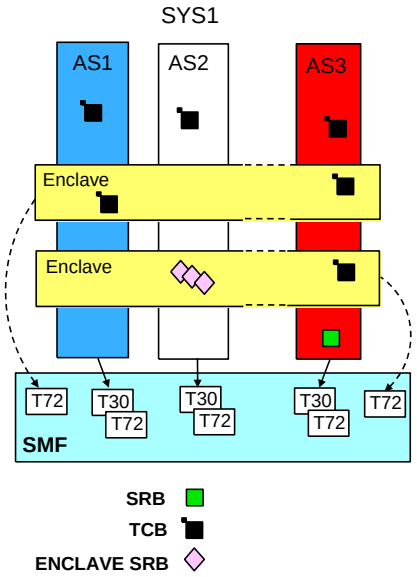
Enclave Management – Introduction



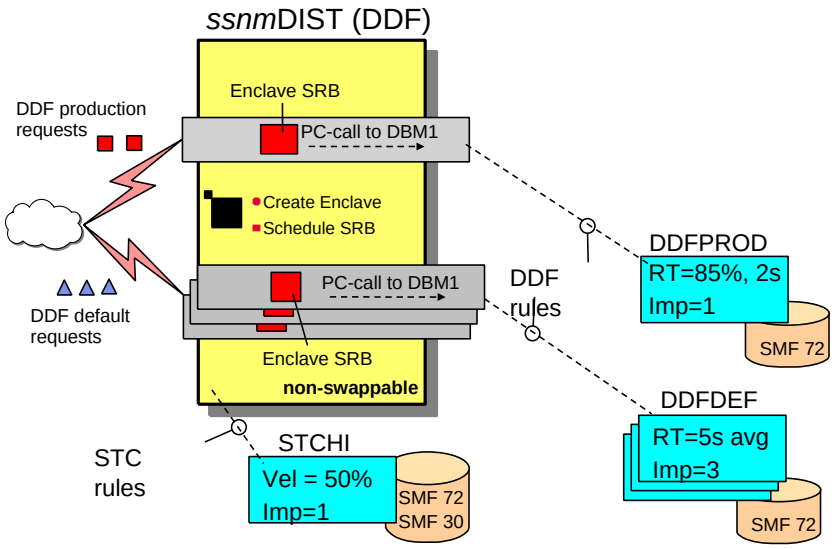
- An **enclave** is a transaction that can span multiple dispatchable units (SRBs and TCBs) in one or several address spaces and is reported on and managed as one unit
- The enclave is managed separately from the address spaces it runs in
  - CPU and I/O resources associated with processing the transaction represented by the enclave are managed by the transaction's performance goal
  - Storage (MPL level, paging) of the address space is managed to meet the goals of the enclaves it serves (if enclave server address space) or to the performance goal of the address space (if no server address space)

Enclave Characteristics

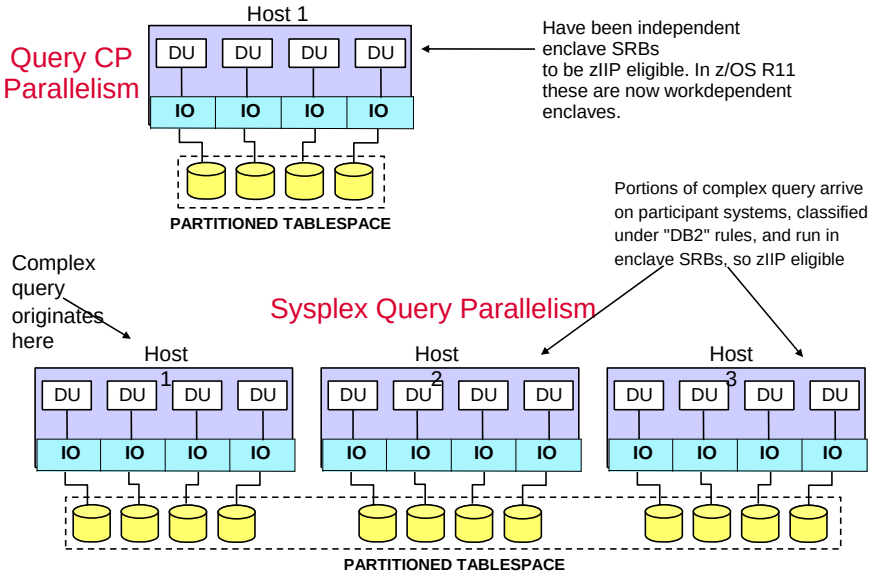
- Created by an address space (the "owner")
- One address space can own many enclaves
- One enclave can include multiple dispatchable units (SRBs/tasks) executing concurrently in multiple address spaces (the "participants")
  - Enclave SRBs are preemptible, like tasks
  - All its dispatchable units are managed as a group
- Many enclaves can have dispatchable units running in one participant address space concurrently
- RMF produces separate T72 SMF records for independent enclaves



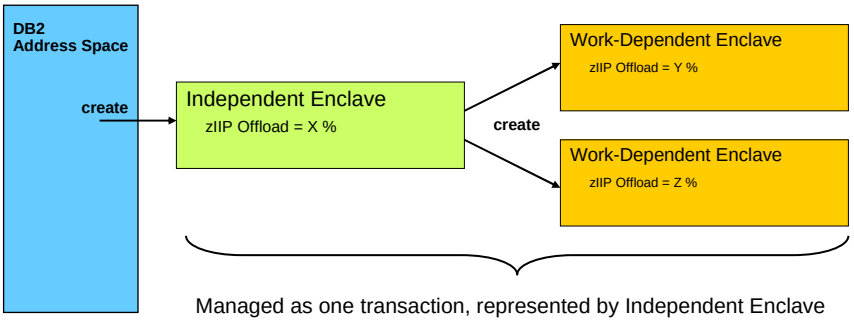
DDF and Enclave SRBs



DB2 Parallel Query and Enclave SRBs

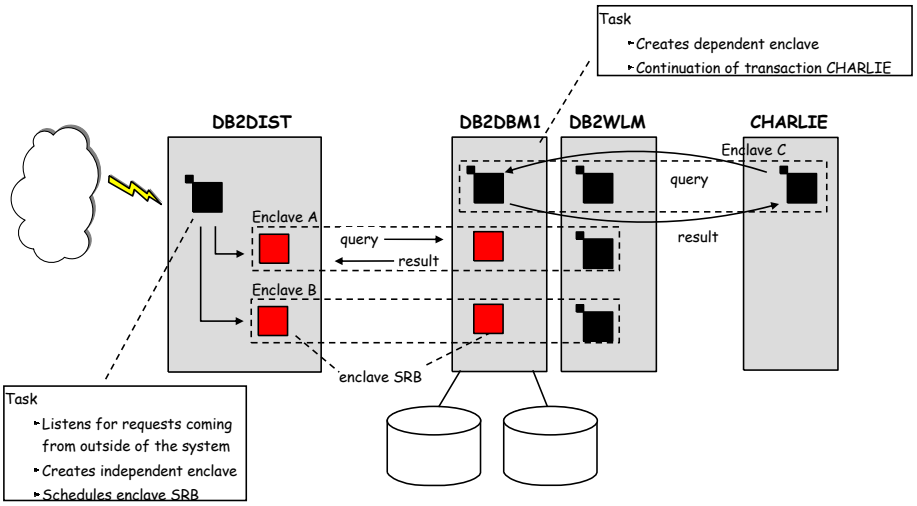


Work-Dependent Enclaves



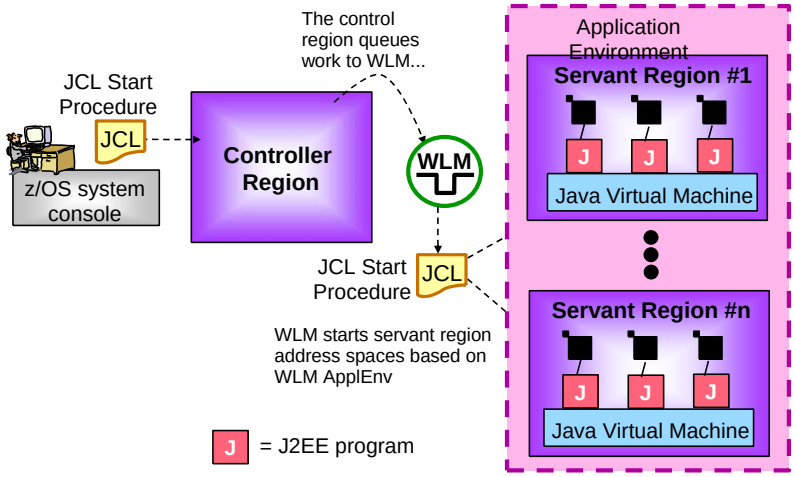
Implement a new type of enclave named “Work-Dependent” as an extension of an Independent Enclave. A Work-Dependent enclave becomes part of the Independent Enclave’s transaction but allows to have its own set of attributes (including zIIP offload percentage)

DB2 Stored Procedures and Enclaves



WebSphere App Server Use of Enclaves

WAS on z/OS transactions, arriving at the Control Region, each run in an enclave that is classified under the "CB" rules.



Service Class with Enclave Transactions

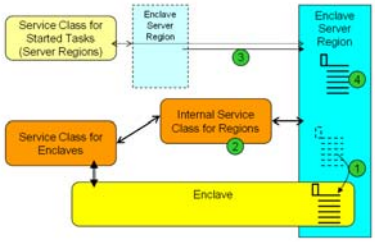
```
REPORT BY: POLICY=WLMPOL01   WORKLOAD=WAS       SERVICE CLASS=WI180%01   RESOURCE GROUP=*NONE      PERIOD=1 IMPORTANCE=1
                                                   CRITICAL     =NONE

  -TRANSACTIONS-  TRANS-TIME HHH.MM.SS.TTT  --DASD I/O--  ---SERVICE---   SERVICE TIME  ---APPL %---  --PROMOTED--  ----STORAGE----
  AVG       1.04  ACTUAL                43  SSCHRT   0.0  IOC         0   CPU  225.586  CP      9.36  BLK    0.000  AVG        0.00
  MPL       1.04  EXECUTION             41  RESP     0.0  CPU     62663K  SRB    0.000  AAPCP   0.13  ENQ    0.000  TOTAL      0.00
  ENDED    44604  QUEUED                 1  CONN     0.0  MSO         0   RCT    0.000  IIPCP   0.00  CRM    0.000  SHARED     0.00
  END/S    24.78  R/S AFFIN              0  DISC     0.0  SRB         0   IIT    0.000                 LCK    0.000
  #SWAPS       0  INELIGIBLE             0  Q+PEND   0.0  TOT     62663K  HST    0.000  AAP     3.18                 -PAGE-IN RATES-
  EXCTD        0  CONVERSION             0  IOSQ     0.0  /SEC    34813   AAP   57.172  IIP     0.00                 SINGLE      0.0
  AVG ENC   1.04  STD DEV              135                                 IIP    0.000                               BLOCK       0.0
  REM ENC   0.00                                              ABSRPTN    34K                                         SHARED      0.0
  MS ENC    0.00                                              TRX SERV   34K                                         HSP         0.0

             RESP  ------------------------------ STATE SAMPLES BREAKDOWN (%) ------------------------------  ------STATE------
  SUB    P   TIME  --ACTIVE-- READY IDLE  ----------------------------WAITING FOR----------------------------  SWITCHED SAMPL(%)
  TYPE       (%)   SUB  APPL             TYP3                                                                 LOCAL SYSPL REMOT
  CB    BTE   0.0   0.0   0.0   0.0  0.0   0.0                                                                   0.0   0.0   0.0
  CB    EXE  93.6   1.5  98.3   0.0  0.0   0.3                                                                   0.0   0.0   0.0

  GOAL: RESPONSE TIME 000.00.01.000 FOR  80%

           RESPONSE TIME EX   PERF  AVG   --EXEC USING%--  -------------- EXEC DELAYS % ----------  -USING%-  --- DELAY % ---    %
  SYSTEM      ACTUAL%    VEL% INDX ADRSP  CPU AAP IIP I/O  TOT CPU AAP   Q                          CRY CNT  UNK IDL CRY CNT  QUI
                                                                        MPL
  JC0            100     78.7  0.5   0.9   11 2.5 0.0 0.0  3.7 2.3 1.2 0.1                          0.0 0.0   83 0.0 0.0 0.0  0.0
```

Service Class with Enclave Transactions

```
-TRANSACTIONS-  TRANS-TIME HHH.MM.SS.TTT
  AVG       1.04  ACTUAL                43
  MPL       1.04  EXECUTION             41
  ENDED    44604  QUEUED                 1
  END/S    24.78  R/S AFFIN              0
  #SWAPS       0  INELIGIBLE             0
  EXCTD        0  CONVERSION             0
  AVG ENC   1.04  STD DEV              135
  REM ENC   0.00
  MS ENC    0.00
```

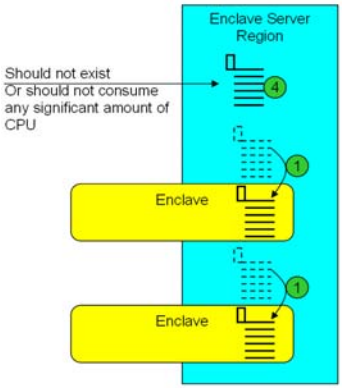
7. What is Enclave Server Management?

Enclave Server Management – Today



- After enclave has been created
  1. TCB joins an enclave
     (alternatively an SRB can be schedule into the enclave and the SRB uses SYSEVENT ENCASSOC to associate the enclave with the region)
  2. An internal service class is created (unless it already exists) to associate the region with the enclave service class (this is required for storage management)
  3. The server region is moved from its external service class to the internal service class
- Result
  - All TCBs (1)which have joined an enclave are managed towards the goals of the external service class for the enclaves
  - All TCBs (4) which have not joined an enclave
    - Those are not really managed at all
    - But they inherit attributes (Dispatch Priority, I/O Priority) of the enclave service class

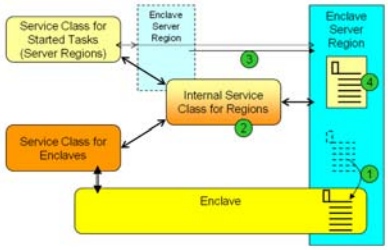
Enclave Server Management – Ramifications



- Assumption (Programming Model)
  - All work being executed within the server region is related to enclaves
    - That means there is no TCB which consumes a significant amount of resources not related to enclaves
- What if the programming model does not hold true?
  - What happens if there is significant work running in TCBs not associated with enclaves?
    - Example: Garbage collection for a JVM (WAS)
    - Example: 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 in server regions ?? (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 that the swap in is only for a queue server task which wants to select a unit of work and then joins the enclave. If no enclave is joined, the address space is again swapped out
  - 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

Enclave Server Management – New

Specify: MANAGENONENCLAVEWORK=YES



- After enclave has been created
  1. TCB joins an enclave
     (alternatively an SRB can be schedule into the enclave and the SRB uses SYSEVENT ENCASSOC to associate the enclave with the region)
  2. An internal service class is created (unless it already exists) to associate the region with the enclave service class (this is required for storage management)
  3. The server region is moved from its external service class to the internal service class
     But the internal service class is now associated with the STC service class for the region and the enclave service class
- Result
  - All TCBs (1)which have joined an enclave are managed towards the goals of the external service class for the enclaves
  - All TCBs (4) which have not joined an enclave are now managed towards the goals of the STC service class for the server regions

Enclave Server Management – Summary

- When should you use MANAGENONENCLAVEWORK=YES?
  - This is really a question of priorities and what do you want to achieve in your environment!
    - For WAS environments tasks like garbage collection can consume a significant amount of CPU and they can last a significant amount of time
      - →If your enclave work is too low prioritized such tasks may take longer then actually necessary and therefore may impact your productive work
      - →In such cases it is worthwhile to think of using the new option and prioritize the regions high enough so that the maintenance can complete fast
  - If you ever experienced the “swapping” problem or if you are vulnerable to such a problem because your enclave work runs very infrequently
    - It is also worthwhile to use this option
- When shouldn’t you use the new OPT parameter?
  - There is no reason not to use it!
    - But if your application environment exploiting enclaves fits to our original programming model (as most DB2 Stored procedure work does) there is also no reason that you have to use the new option.

8. What is Blocked Workload Support?

Blocked Workload Support

- **Problem**
  - Work competes for resources, serialized by locks and latches
    - Low import work may hold a resource and high important work may have to wait for it
- **WLM Blocked Workload Support**
  - Recognizes blocked work
    - Work which doesn‘t show any progress for an elongated period of time
  - Allows this work to use a small amount of CPU periodically
    - With the hope to resolve existing (potential) resource contentions

Blocked Workload Support: User Interface: IEAOPT

BLWLTRPCT	Percentage of the CPU capacity of the LPAR to be used for promotion - Specified in units of 0.1% - Default is 5 (=0.5%) - Maximum is 200 (=20%) - Would only be spent when enough units of work exist which need promotion
BLWLINTHD	Specifies threshold time interval for which a blocked address space or enclave must wait before being considered for promotion. - Minimum is 5 seconds. Maximum is 65535 seconds. - Default is 20 seconds.

Blocked Workload Support: User Interface: RMF

```
                                    C P U   A C T I V I T Y
…

BLOCKED WORKLOAD ANALYSIS

OPT PARAMETERS: BLWLTRPCT (%)   0.5   PROMOTE RATE:  DEFINED    50000   WAITERS FOR PROMOTE:  AVG   0.001
                BLWLINTHD        60                 USED (%)      95                         PEAK     15
```

- **Extensions of RMF Postprocessor CPU Activity and WLMGL reports with information about blocked workloads and the temporary promotion of their dispatching priority**
- **SMF record 70-1 (CPU activity) and SMF 72-3 (Workload activity)**

9. How Do the New Discretionary Controls Work?

New z/OS 1.12 Discretionary Batch Improvements

- TIMESLICES=1-255 (IEAOPTxx)
- Specifies number of timeslices a CPU-intensive address space or enclave with a discretionary goal should be given before a dispatchable unit of equal importance is dispatched
- Increasing this parameter might:
  - Increase processor delay for some CPU-intensive work
  - Decrease the number of context switches between equal priority work and therefore increase the throughput of the system
- Parameter only affects discretionary work that is CPU-intensive as determined by significant mean time to wait (MTTW)
  - As controlled by the CCCSIGUR parameter
- Default: 1

New z/OS 1.12 Discretionary Batch

- CCCSIGUR=0-32767 (IEAOPTxx)
- Specifies the minimum mean-time-to-wait (MTTW) threshold value in milliseconds for heavy CPU users
  - Used to determine the range of MTTW values which are assigned to each of the ten MTTW dispatching priorities - x'C0' to x'C9'
  - Specified real time value is adjusted by relative processor speed to become SRM time to give consistent SRM control across various processors
  - Default Value: 45
- Used to differentiate Dispatch Priority of discretionary work
  - Work clumps at x'C9'
    - Appears all address spaces have short MTTW
    - CCCSIGUR is too large and should be decreased
  - Work clumps at x'C0'
    - Appears all work has large MTTW
    - CCCSIGUR is too small and should be increased
- Recommendation: start by doubling or halving the value

10. What is the Difference Between IRD and HiperDispatch?

HiperDispatch vs IRD logical processor management

- **Intelligent Resource Director (IRD) LP management**
  - Automatically adjusts the number of **online** CPs to achieve the minimum required to run the work of the partition.
    - applies to CPs only – not zAAPs or zIIPs
  - Actually varies the CPUs online/offline
    - LP shares recalculated to those online
- **HiperDispatch mode**
  - Minimizes the number of **active** LPs to achieve the partition share and the partition share of unused resources (when the unused is of value)
    - supports zAAPs and zIIPs
  - **"Parks"** unused processors. Parking is simply placing a LP in a long-term wait until it is again needed to run work.
  - WLM can park/unpark CPUs over a relatively short time scale (seconds)
- **IRD LP management is disabled in HiperDispatch mode**

WLM Top 10 Things You Should Now Understand!

1. How does work get in the correct Service Class?
2. What is the EWLM subsystem all about?
3. What is the Right Number of Service Classes?
4. How Does WLM Manage Work?
5. How is Velocity calculated?
6. What are Enclaves?
7. What is Enclave Server Management?
8. What is Blocked Workload support?
9. How do the new Discretionary controls work?
10. What is the difference between IRD and Hiperdispatch?

z/OS Tuning Courses from IBM Training

- Basic z/OS Tuning Using the Workload Manager (WLM)
  - ES545
  - 4.5 Days, Hands-on Lab Exercises
- Advanced z/OS Performance: WLM, Sysplex, Unix Services, Web
  - ES851
  - 4.5 Days
- ibm.com/training