IBM Education Assistance for z/OS V2R1

Item: I/O Priority Group
Element/Component: z/OS WLM and RMF

Material is current as of March 2013

Agenda

- Trademarks
- Presentation Objectives
- Overview
- Usage & Invocation
- Migration & Coexistence Considerations
- Presentation Summary
- Appendix

Trademarks

- See http://www.ibm.com/legal/copytrade.shtml for a list of trademarks.

Presentation Objectives

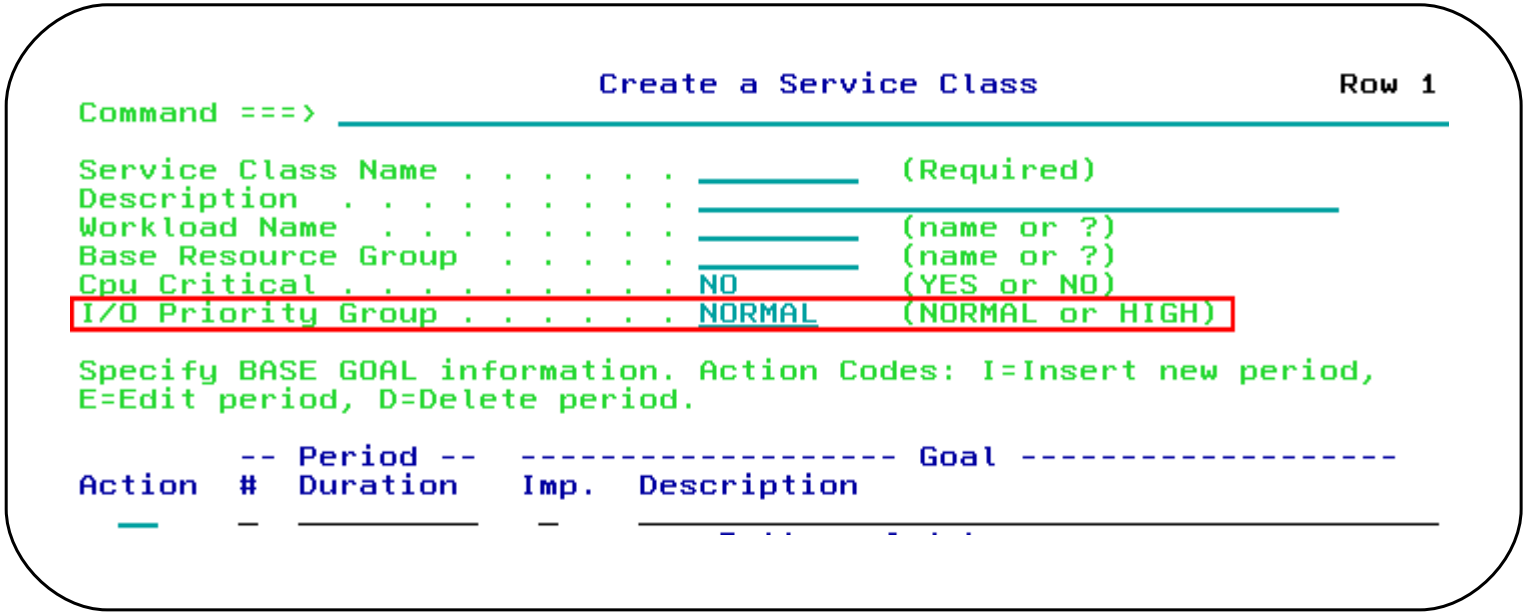
- This line item provides a new option to protect critical work
- Besides long-term Storage and CPU protection, you will now be able to assign long-term **I/O protection** to work which is extremely I/O-sensitive
- You achieve I/O protection by assigning a service class to I/O priority group **HIGH**
  - I/O priority group **NORMAL** is the default for service classes
- When assigning I/O protection, you ensure that less I/O critical work will generally have a lower I/O priority

Overview

- **Problem Statement / Need Addressed:**
  - I/O priority queueing is used to control DASD I/O requests that are queued because the device is busy.
  - You can direct WLM to manage I/O priorities in the sysplex based on service class goals and importance. WLM dynamically adjusts the I/O priority based on how well goals are met and whether the device can contribute to achieve the goal.
  - When the I/O priority of certain work should be increased, WLM needs to understand what other work is affected.
    - WLM determines device sets being most used by each service class to provide this access pattern information.
    - The device sets are refreshed periodically every 10 minutes.
  - Typically, different workloads use distinct device sets and WLM changes I/O priorities between service classes using the same device set.
  - If work starts to use a device which crisscrosses the access pattern currently used, WLM observes I/O delay immediately but does not start I/O priority management until the device sets are refreshed. In the worst case, this might take up to 10 minutes.
    - Thus, the I/O priority is not raised immediately and goals might be missed.
- **Solution:** Introduce concept of I/O priority groups for long-term I/O protection
- **Benefit/Value:** I/O-sensitive service classes are assigned to group HIGH and always have higher I/O priority than service classes in group NORMAL

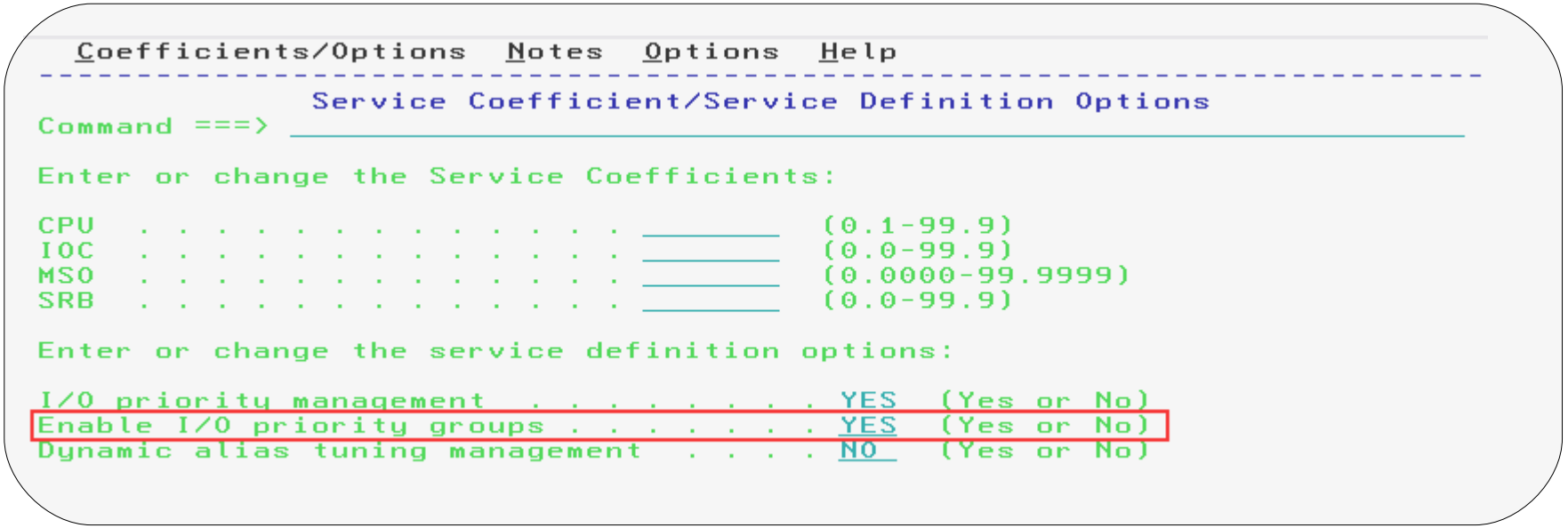
Usage & Invocation

- Use the I/O Priority Group field on following panels of the WLM ISPF Application
  - Create a Service Class
  - Modify a Service Class
  - Override attributes for a Service Class



Usage & Invocation

- I/O priority group HIGH is only honored when I/O priority groups are enabled on the Service Definition Options panel
- In addition, dynamic I/O priority management must also be enabled, that is, workload management dynamically manages your I/O priorities based on service class goals and importance

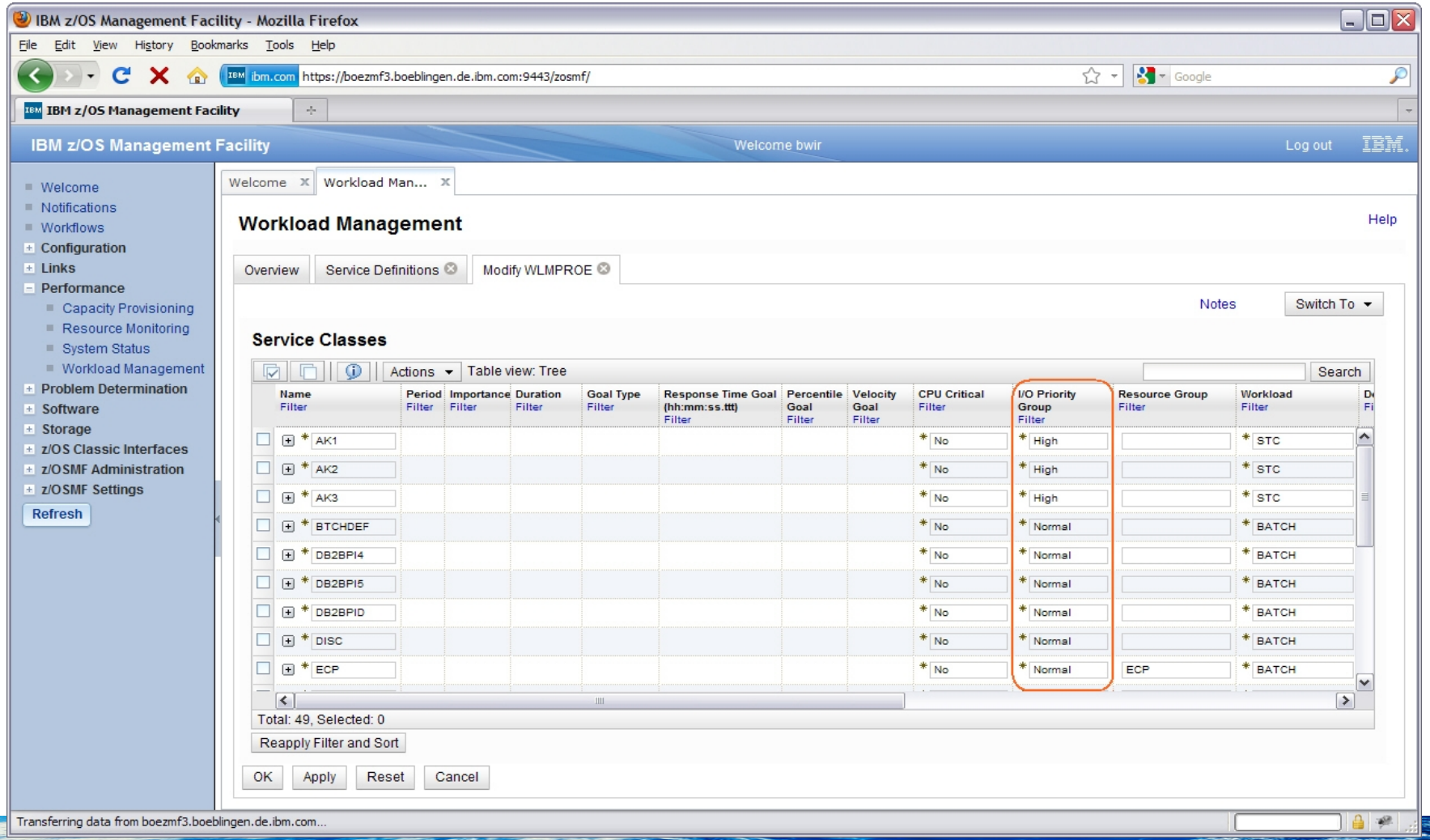


Usage & Invocation

- Group HIGH is only honored by WLM if dynamic I/O priority management is enabled and I/O priority groups are enabled, see the Coefficients/Options panel
- **Validate definition** option can be used to check whether service classes assigned to I/O priority group HIGH although above options not enabled
  - Warning message **IWMAM918W** will be displayed:
    "Service class(es) assigned to I/O priority group HIGH but I/O priority management or I/O priority groups are not enabled. The I/O priority group will not be honored."

Usage & Invocation

- z/OSMF Workload Management task is enhanced accordingly



Usage & Invocation … callable services

- When using the WLM services **IWMDEXTR** or **IWMDINST** to extract or install a service definition, the service definition is either in XML format or a data area mapped by the IWMSERVD mapping macro.
- The layout of the XML service definition (DTD) is extended as follows. The entire DTD is described in Appendix C of the WLM Services Guide.

```
<!ELEMENT ServiceClasses ( ServiceClass* ) >
<!ELEMENT ServiceClass ( Name, Description?, CreationDate, CreationUser, ModificationDate, ModificationUser,
  CPUCritical?, IOPriorityGroup?, ResourceGroupName?, Goal ) >

<!ELEMENT ServiceClassOverrides ( ServiceClassOverride* ) >
<!ELEMENT ServiceClassOverride ( ServiceClassName, CPUCritical, IOPriorityGroup?, ResourceGroupName?, Goal ) >

<!ELEMENT ServiceClassName ( #PCDATA ) >
<!ELEMENT CPUCritical ( #PCDATA ) >
<!ELEMENT IOPriorityGroup ( #PCDATA ) >

<!ELEMENT ServiceOptions ( IOPriorityManagement, DynamicAliasManagement?, IOPriorityGroupsEnabled? ) >

<!ELEMENT IOPriorityManagement ( #PCDATA ) >
<!ELEMENT DynamicAliasManagement ( #PCDATA ) >
<!ELEMENT IOPriorityGroupsEnabled ( #PCDATA ) >
```

- Within the IWMSERVD mapping, the general service definition data area mapped by IWMSVDEF is extended
  - There is a new flag in the service class attributes section. If SVDEFIPG is ON, the service class is assigned to the I/O priority group
  - There is a new flag in the service defintion constants section. If SVDEFIOE is ON, I/O priority groups are enabled

Usage & Invocation … callable services

- The RASD parameter list of **SYSEVENT REQASD** and **REQFASD** is extended to return information about the I/O priority group of the address space. Additional flags are added to field RASDFLAGS1.
  - RasdCHighIO (Bit 6): Service class designated by classification or RESET SRVCLASS is assigned to I/O priority group HIGH
  - RasdIOPrioHigh (Bit 7): I/O priority group HIGH was assigned either to the address space (see RasdCHighIO) or to transaction service classes being served by the space
- **IWMRQRY** is the interface reporting products should use to obtain address space related general execution delays. The answer area mapped by IWMWRQAA is enhanced according to REQFASD. An additional flag is added to field RQAEFLG1.
  - RqaeIOPrioHigh (Bit 6): Same as RasdIOPrioHigh
- **IWMPQRY** is the interface to return a representation of the active policy. The answer area mapped by IWMSVPOL is extended
  - There is a new flag in the service policy information section SVPOLSP. If SVPOLIOE is on, I/O priority groups are enabled
  - There is a new flag in the service class definition section SVPOLCD. If SVPOLIPG is on, the service class is assigned to the I/O priority group

Usage & Invocation … SMF record type 99

- SMF record type 99 subtype 2 (WLM service class data) is extended. Field SMF99_FLAGS of the Period Data Section is updated. If bit 6 is set, the period belongs to a service class which is assigned to I/O priority group HIGH.

Offsets	Name	Length	Format	Description
201 C9	SMF99_FLAGS	1	binary	Flags. Bit 0 to 5: Meaning not changed **Bit 6: Period belongs to a service class which was assigned to I/O priority group.** Bit 7: Reserved.

- SMF record type 99 subtype 6 (service class period data) is extended. Field SMF996_FLAGS is updated. If bit 1 is set, the period belongs to a service class which was assigned to I/O priority group HIGH.

Offsets	Name	Length	Format	Description
80 50	SMF996_FLAGS	1	binary	Flags. Bit 0: Meaning not changed **Bit 1: Period belongs to a service class which was assigned to I/O priority group.** Bit 2-7: Reserved.

Usage & Invocation … SMF record type 72.3 and 79

- In addition to WLM's SMF type 99 record, RMF's record types 72 subtype 3 and SMF 79 subtypes 1 and 2 are extended to indicate assignment to the I/O priority group.

SMF record 72 subtype 3 (Workload activity) – Workload manager control section				
Offsets	Name	Length	Format	Description
0 0	R723MSCF	1	Binary	Service/Report class flags. Bit 0-6: Meaning not changed **Bit 7: Indicator for I/O priority group HIGH**

SMF record 79 subtype 1 (Address space state data) – ASD data section				
Offsets	Name	Length	Format	Description
236 EC	R791FLG3	1	Binary	Additional flags. **Bit 0: Service class assigned by classification or RESET SRVCLASS belongs to I/O priority group HIGH in the active policy** **Bit 1: I/O priority group HIGH was assigned either to the address space or to transaction service classes served by the space** Bit 2-7: Reserved
237 ED		3		Reserved
SMF record 79 subtype 2 (address space resource data) – ARD data section				
224 E0	R792FLG3	1	binary	Additional flags. **Bit 0: Service class assigned by classification or RESET SRVCLASS belongs to I/O priority group HIGH in the active policy** **Bit 1: I/O priority group HIGH was assigned either to the address space or to transaction service classes served by the space** Bit 2-7: Reserved
225 E1		3		Reserved

Usage & Invocation … RMF Postprocessor Workload Activity Report

- The **Postprocessor Workload Activity (WLMGL)** report is extended. As soon as the service class is assigned to I/O priority group HIGH, an appropriate indication is displayed in the SERVICE CLASS(ES) and SERVICE CLASS PERIODS sections. If the service class is not assigned to group HIGH or no assignment took place because backlevel data is processed, the I/O PRIORITY GROUP=HIGH indication is omitted.

```
---------------------------------------------------------------------------------------------------------------------- SERVICE CLASS(ES)

REPORT BY: POLICY=WLMPOL     WORKLOAD=BATCH      SERVICE CLASS=BATCH      RESOURCE GROUP=*NONE
                                                 CRITICAL     =CPU+STORAGE                        I/O PRIORITY GROUP=HIGH
                                                 DESCRIPTION  =Batch Workload

-TRANSACTIONS-  TRANS-TIME HHH.MM.SS.TTT  --DASD I/O--  ---SERVICE---   SERVICE TIME  ---APPL %---  --PROMOTED--  ----STORAGE----
AVG       0.74  ACTUAL                 0  SSCHRT   0.0  IOC          0   CPU    6.429  CP       0.66  BLK     0.000  AVG     7663.01
MPL       0.74  EXECUTION              0  RESP     0.0  CPU     287332   SRB    0.000  AAPCP    0.00  ENQ     0.000  TOTAL   5698.61
ENDED        0  QUEUED                 0  CONN     0.0  MSO     537297   RCT    0.002  IIPCP    0.00  CRM     0.000  SHARED     0.00
END/S     0.00  R/S AFFIN              0  DISC     0.0  SRB         11   IIT    0.000                 LCK     0.000
#SWAPS      15  INELIGIBLE             0  Q+PEND   0.0  TOT     824640   HST    0.000  AAP      0.00  SUP     0.000  -PAGE-IN RATES-
EXCTD        0  CONVERSION             0  IOSQ     0.0  /SEC       916   AAP    0.000  IIP      0.06                 SINGLE       0.0
AVG ENC   0.00  STD DEV                0                                  IIP    0.502                               BLOCK        0.0
REM ENC   0.00                                          ABSRPTN  1232                                               SHARED       0.0
MS ENC    0.00                                          TRX SERV 1232                                               HSP          0.0


---------------------------------------------------------------------------------------------------------------------- SERVICE CLASS PERIODS

REPORT BY: POLICY=WLMPOL     WORKLOAD=BATCH      SERVICE CLASS=BATCHHI    RESOURCE GROUP=*NONE      PERIOD=2 IMPORTANCE=4
                                                 CRITICAL     =NONE                                I/O PRIORITY GROUP=HIGH

-TRANSACTIONS-  TRANS-TIME HHH.MM.SS.TTT  --DASD I/O--  ---SERVICE---   SERVICE TIME  ---APPL %---  --PROMOTED--  ----STORAGE----
AVG       0.74  ACTUAL                 0  SSCHRT   0.0  IOC          0   CPU    6.429  CP       0.66  BLK     0.000  AVG     7663.01
MPL       0.74  EXECUTION              0  RESP     0.0  CPU     287332   SRB    0.000  AAPCP    0.00  ENQ     0.000  TOTAL   5698.61
ENDED        0  QUEUED                 0  CONN     0.0  MSO     537297   RCT    0.002  IIPCP    0.00  CRM     0.000  SHARED     0.00
END/S     0.00  R/S AFFIN              0  DISC     0.0  SRB         11   IIT    0.000                 LCK     0.000
#SWAPS      15  INELIGIBLE             0  Q+PEND   0.0  TOT     824640   HST    0.000  AAP      0.00  SUP     0.000  -PAGE-IN RATES-
EXCTD        0  CONVERSION             0  IOSQ     0.0  /SEC       916   AAP    0.000  IIP      0.06                 SINGLE       0.0
AVG ENC   0.00  STD DEV                0                                  IIP    0.502                               BLOCK
REM ENC   0.00                                          ABSRPTN  1232                                               SHARED
MS ENC    0.00                                          TRX SERV 1232                                               HSP
```

Migration & Coexistence Considerations

- Toleration/coexistence **APAR OA37824** needed on z/OS V1R12 and z/OS V1R13 systems because dynamic I/O priority management is a sysplex-wide function
  - Pre-V2R1 systems are enabled by OA37824 to recognize I/O priority group HIGH and manage I/O priorities consistently in the mixed-release sysplex
- I/O priorities are also recognized at the disk system level. Thus...
  - I/O Priority Groups must be explicitly enabled in the WLM service definition options
  - Turn on this option only if all systems sharing disk systems run on z/OS V2R1 or on z/OS V1R12 / R13 with OA37824
  - As soon as the *Enable I/O Priority Groups* option is turned on in one sysplex, turn it also on in other sysplexes even if they do not exploit I/O priority group HIGH. This ensures that all systems sharing a disk system work with an identical range of I/O priorities
- Assigning service classes to I/O priority group HIGH is only possible with the z/OS V2R1 WLM ISPF Application or z/OSMF V2R1
- As soon as one service class is assigned to I/O priority group HIGH, the functionality level of the service definition is changed to **LEVEL029**
- RMF support is only available with z/OS V2R1

Presentation Summary

- With **I/O Priority Groups** a new option is introduced to define special protection for critical work
- If you have work which is extremely **I/O-sensitive**, assign its service class to **I/O priority group HIGH** when creating or modifying the WLM service definition
  - The I/O priority group can be set by using the WLM ISPF Application or the Workload Management task in z/OSMF
  - Work in service classes of I/O priority HIGH **always have a higher I/O priority** than work managed by service classes in I/O priority group NORMAL
- Although applicable to all work, CICS and IMS work will particulary benefit from this enhancement
  - Especially if your CICS/IMS work accesses occasionally data sets residing on volumes which are usually used by other work, for example, batch load
- For reporting purposes, use **SMF records** and the **RMF Postprocessor Workload Activity Report (WLMGL)**

Appendix

- Publication references
  - z/OS V2R1 MVS Planning: Workload Management, SA22-7602
  - z/OS V2R1 MVS System Management Facilities (SMF), SA22-7630
  - z/OS V2R1 RMF Report Analysis, SC33-7991