

IBM Education Assistance for z/OS V2R3

WLM shorter response time goal

Agenda

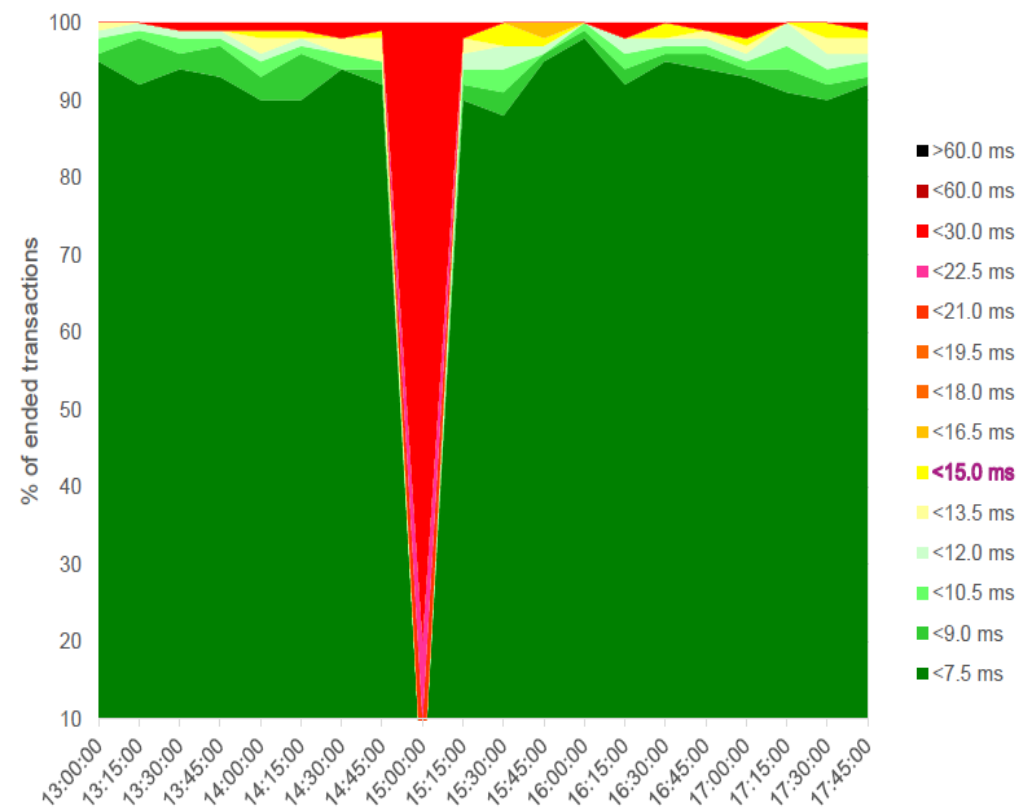
- Trademarks
- Session Objectives
- Overview
- Usage & Invocation
- Interactions & Dependencies
- Migration & Coexistence Considerations
- Session Summary
- Appendix

Trademarks

- See url <http://www.ibm.com/legal/copytrade.shtml> for a list of trademarks.
- Additional Trademarks:
 - None

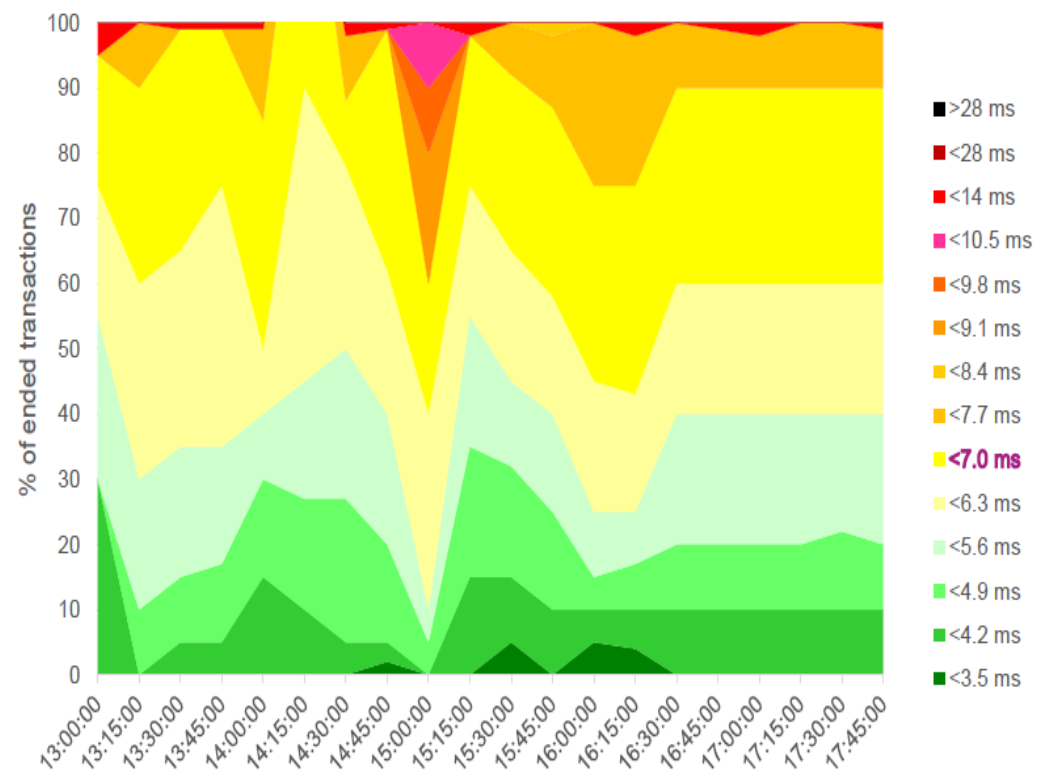
Session Objectives

- WLM enhances the definition of response time goals for service classes
- For service class periods with an *average response time goal* or a *response time goal with percentile*, the lowest goal which can be specified changes from 15 milliseconds to one millisecond
- This figure illustrates an exemplary response time distribution for a service class with a 15 msec response time goal and very fast running transactions.
- During the presented time period, approximately 95% of the transactions complete with an average response time below 7.5 msec (dark green area).
- Thus, the service class considerably overachieves its goal allowing it to run at a low dispatching priority. This makes the service class susceptible to workload spikes where the average response time of all transactions rises far above the goal (red area with the outlier around 15:00).



Session Objectives

- WLM enhances the definition of response time goals for service classes
- For service class periods with an *average response time goal* or a *response time goal with percentile*, the lowest goal which can be specified changes from 15 milliseconds to one millisecond
- A more meaningful goal of 7 milliseconds will cause a better balanced response time distribution as illustrated in the following figure.
- The workload always runs at reasonable dispatch priority preventing bad response times even with workload spikes.



Overview

- Problem Statement / Need Addressed

- Currently, the shortest average or percentile response time that can be specified in the WLM service definition is 15 milliseconds (0.015 sec).
- However, transactions on current z System hardware may complete well below 15 milliseconds, especially in DB2 Distributed Data Facility (DDF) and CICS environments.
- This makes it difficult to manage such transactional work effectively.

- Solution

- With z/OS V2R3, the current lower bound of 15 milliseconds for a response time goal is replaced by one millisecond allowing to specify meaningful goal values for very fast running transactions.
- Both the WLM ISPF Application and the Workload Management task in the IBM z/OS Management Facility (z/OSMF) are enhanced to specify an average or percentile response time goal starting at one millisecond.
- Accordingly, MVS workload management is adapted to fulfill attainment of response time goals below 15 milliseconds.
- WLM's workload reporting is updated to support response time goals lower than 15 milliseconds. The response time and distribution information returned by the IWMRCOLL (collecting workload activity data) service is extended to properly report on transactions which run faster than one millisecond on the average.

- Benefit / Value

- Important work, such as the most important DDF or CICS transactions that have response times below 15 milliseconds, can be managed more effectively.

Usage & Invocation

- The WLM Administrative Application is enhanced to allow the definition of service period response time goals below 15 milliseconds which is the minimum response time goal with z/OS V2R2 and below. The new minimum goal value is 0.001 seconds (one millisecond) and can be defined for base goals or when overriding attributes for a service class.
- When specifying an average response time goal, the total response time can be between 0.001 seconds (one millisecond) and 24 hours.
- With z/OS V2R3, the WLM Application level displayed on the Definition Menu panel changes to LEVEL035. As soon as at least one service class period is defined with a response time goal below 0.015 seconds (15 milliseconds), the Functionality level is raised to LEVEL035.

```

IWMAP34      Average response time goal

Enter a response time of up to 24 hours for period 2

Hours . . . . . 00 (0-24)
Minutes . . . . 00 (0-99)
Seconds . . . . 00.005 (0-9999)

Importance . . 2 (1=highest, 5=lowest)
Duration . . . 10000 (1-999,999,999, or
                    none for last period)

F1=Help  F2=Split  F5=KeysHelp  F9=Swap  F12=Cancel
  
```

```

IWMAP35      Response time with percentile goal

Enter a percentile and response time goal for period 3

Percentile . . 90 (1-99)

Hours . . . . . 00 (0-24)
Minutes . . . . 00 (0-99)
Seconds . . . . 00.008 (0-9999)

Importance . . 2 (1=highest, 5=lowest)
Duration . . . 10000 (1-999,999,999, or
                    none for last period)

F1=Help  F2=Split  F5=KeysHelp  F9=Swap  F12=Cancel
  
```

Usage & Invocation

- The Workload Management task in IBM z/OS Management Facility is updated to allow the definition of response time goals below 15 milliseconds. Response time goals can be specified in the **Service Classes or Service Class Overrides** dialog.
- The supported values for an average or percentile response time are now between 24:00:00.000 (24 hours) and 00:00:00.001 (0.001 second).
- When the response time goal specified is out of the supported range, message **IZUW630E** **The response time goal is not valid** is displayed.
- The explanation for IZUW630E is updated: The response time goal can range from one millisecond to 24 hours.
- As soon as one service class period is defined with a response time goal below 00:00:00.015 (15 milliseconds), the functionality level displayed in the **Service Definition Details** is raised to 035.

Usage & Invocation

- The Collect Workload Activity Data service IWMRCOLL is extended
- The RCAE Response Time Data section of the IWMRCOLL answer returns 7 new double-word fields containing transaction times in units of microseconds for a higher precision.
- Transaction time fields rcaeTET, rcaeXET, rcaeETS, rcaeQDT, rcaeADT, rcaeCVT, and rcaeIQT are still provided in units of 1024-microseconds.

| Offset DEC | Offset HEX | Type | Length | Name | Description |
|------------------|------------|-----------|--------|----------------|---|
| 0 | 0 | Structure | 240 | RCAERST | RCAE response time data |
| ... | | | | | |
| 176 | B0 | Unsigned | 8 | rcaeTranTime | Total transaction elapsed time. Same as rcaeTET but in microsecond units |
| 184 | B8 | Unsigned | 8 | rcaeTranTimeSq | Sum of transaction elapsed times squared. Same as rcaeETS but in microsecond units |
| 200 | C8 | Unsigned | 8 | rcaeExecTime | Total transaction execution time. Same as rcaeXET but in microsecond units |
| 208 | D0 | Unsigned | 8 | rcaeQDlyTime | Total queue delay time. Same as rcaeQDT but in microsecond units |
| 216 | D8 | Unsigned | 8 | rcaeADlyTime | Total time batch jobs were ineligible to run because a resource the job had affinity to was unavailable. Same as rcaeADT but in microsecond units |
| 224 | E0 | Unsigned | 8 | rcaeConvTime | Total time batch jobs spent in JCL conversion. Same as rcaeCVT but in microsecond units |
| 232 | E8 | Unsigned | 8 | rcaeQueTime | Total time batch jobs spent on job queue after JCL conversion. Same as rcaeIQT but in microsecond units |
| Constants | | | | | |
| | | Decimal | 1 | Rcae_Version24 | RCAE version 24 (24 = HBB77B0) |
| | | Decimal | 1 | Rcaa_Version24 | RCAA version 24 (24 = HBB77B0) |

Usage & Invocation

- The WLM service definition mapping IWMSVDEF is provided with new version constants for z/OS V2R3
- Accordingly, new constants new constants for functionality level 35 and WLM version number 35 are provided in the following WLM service definition mapping macros:
 - IWMSVAEA (WLM service definition application environment mapping)
 - IWMSVDCR (WLM service definition classification rule mapping)
 - IWMSVNPA (WLM service definition notepad mapping)
 - IWMSVPOL (IWMPQRY answer area mapping)

| Type | Length | Name | Value HEX | Value DEC | Description |
|---------|--------|-----------------------|-----------|-----------|---|
| Decimal | 1 | SVDEF_SP7B0 | 23 | 35 | WLM version number introduced with WLM z/OS V2R3 |
| Decimal | 1 | SVDEF_LEVEL035 | 23 | 35 | Functionality level introduced with WLM z/OS V2R3 |
| Decimal | 1 | SVDEF_Current_Version | 23 | 35 | Current version level used when checking functionality within WLM product |

Usage & Invocation

- Structure of XML service definition (DTD) - A new name space and corresponding functionality level is added (see Appendix C in Workload Management Services)

| Spaces and Corresponding Functionality | |
|---|-------|
| Name Space | Level |
| ... | |
| http://www.ibm.com/xmlns/prod/zwlm/2015/12/ServiceDefinition.xsd | 030 |
| http://www.ibm.com/xmlns/prod/zwlm/2017/09/ServiceDefinition.xsd | 035 |

Interactions & Dependencies

- Software Dependencies
 - None
- Hardware Dependencies
 - None
- Exploiters
 - z/OS Resource Measurement Facility (RMF) V2R3 exploits the new transaction response time fields in units of microseconds returned by the IWMRCOLL service

Migration & Coexistence Considerations

- This support is provided with z/OS V2R3. Since z/OS V2R2 and/or V2R1 may run in the same sysplex and workload management operates on a sysplex scope, coexistence/toleration PTFs (**APAR OA50824**) are required for z/OS V2R2 and V2R1.
- The WLM Administrative Application level changes to 35, which is only available with z/OS V2R3.
- The functionality level of a service definition changes to 35 as soon as a service class period is defined with a response time goal (average or with percentile) below 15 milliseconds. This applies to both the base goal information as well as overridden goals.
- A service definition of functionality level 35 cannot be extracted, displayed, modified, installed or activated in a back-level WLM Administrative Application. Since toleration of functionality level 35 is provided down to z/OS V2R1, a service definition's policy specifying response time goals below 15 milliseconds can be successfully activated in a mixed z/OS V2R3/R2/R1 release sysplex.
 - However, response time goals below 15 milliseconds are not honored. Instead, WLM applies a 15 milliseconds goal on all systems in the sysplex (even on z/OS V2R3).
 - This "goal conversion" takes place as long as at least one sysplex member is below z/OS V2R3.
 - As soon as all systems are on z/OS V2R3, WLM applies the original goal as specified in the activated service policy.

Session Summary

- The shortest average or percentile response time that can currently be specified in the WLM service definition is 15 milliseconds (0.015 sec).
 - This is insufficient since transactions on current z System hardware may complete well below 15 milliseconds, especially in DB2 Distributed Data Facility (DDF) and CICS environments.
 - This makes it difficult to manage such transactional work in an effective and reliable manner.
- With z/OS V2R3, the current lower bound of 15 milliseconds for a response time goal is replaced by one millisecond allowing to specify meaningful goal values for very fast running transactions.
 - Thus, important work such as the most important DDF or CICS transactions that have response times below 15 milliseconds, can be managed more effectively.

Appendix

- Publication references
 - z/OS MVS Programming: Workload Management Services
 - z/OS MVS Planning: Workload Management
 - z/OS MVS Data Areas
 - IBM Knowledge Center - z/OSMF messages IZUW0001-IZUW99999