## Rule WLM001: The service class definition may not match workload

**Finding:** CPExpert believes that the service class definition may not match the workload executing under control of the Workload Manager.

**Impact:** This finding should be viewed as a HIGH IMPACT on the performance of your computer system.

**Logic flow:** The following rules cause this rule to be invoked:

| | |
|---|---|
| Rule WLM101: | Service Class did not achieve average response goal |
| Rule WLM102: | Service Class did not achieve percentile response goal |
| Rule WLM103: | Service Class did not achieve execution velocity goal |
| Rule WLM104: | Subsystem Service Class did not achieve average response goal |
| Rule WLM105: | Subsystem Service Class did not achieve percentile response goal |

**Discussion:** Users define service classes by specifying the workload associated with the service class, specifying the performance goal for the service class, and specifying the importance of the performance goal. Optionally, users may define performance periods for the service class and may assign a resource group to the service class.

Service classes are used by the Workload Manager to determine whether performance goals are being met and to control the allocation of system resources to address spaces. A single transaction, job, Started Task, etc. can be controlled by only one service class (defined in the service policy).

The Workload Manager can assign and reassign system resources only to the extent that system resources exist. The Workload Manager may not be able to manage the system to meet performance goals if (1) the goals are too restrictive, (2) too many goals are at a high importance level, (3) the workload is not sufficiently partitioned into service classes, or (4) the system is inadequate to meet management objectives for the workload being processed.

- **Goals too restrictive**. Performance goals may be defined such that the goals are too restrictive with respect to the resource demands of the workload. For example, Rule WLM200 describes a situation in which the CPU requirements of the average transaction is higher than the performance goal specified for the service class to which the transactions are assigned. This situation commonly occurs when users have defined

goals based on prior performance (for example, based on the *RMF Workload Activity Report* in Compatibility Mode), rather than defining the goals based on management requirements. The Workload Manager may not be able to achieve prior performance for all service classes, but management often does not **require** the prior performance.

- **Too many goals are at high importance**. The Workload Manager attempts to balance the performance goals against resource requirements, considering the importance of the goals. If too many goals are defined at a high level, the Workload Manager may be unable to balance the resource requirements of the service classes. This situation often occurs when performance goals have been assigned to most of the workload, rather than associating a discretionary goal for a significant part of the workload.

- **Workload not sufficiently partitioned**. Implicit in the service class concept is that a service class normally describes a homogeneous workload--a workload which has similar performance goals, similar goal importance, and similar resource requirements. If a service class is defined for workloads which are not similar, the Workload Manager may unable to allocate resources to meet widely-varying resource requirements of the workload.

- **System is inadequate**. The system which is processing the workload may be inadequate to satisfy the performance goals of the service classes. The Workload Manager cannot allocate resources to meet performance goals if insufficient resources exist.

CPExpert concludes that the service class definition does not match the workload if the performance goals with the highest importance levels were not met in most of the RMF measurement intervals being analyzed. CPExpert identifies the highest two goal importance levels. For each RMF interval, CPExpert counts the service classes at these goal importance levels which missed their performance goal. CPExpert produces Rule WLM001 if at least two service classes with these important performance goals missed their performance goal in over half of the RMF measurement intervals being analyzed.

The following example illustrates the output from Rule WLM001:

```
RULE WLM001:  THE SERVICE CLASS DEFINITION MAY NOT MATCH WORKLOAD

  CPExpert believes that you may have a "mismatch" between the service
  classes you have defined and the workload executing on your system, or
  your system is insufficient for the workload.  Perhaps you have defined
  too many service classes at high importance levels, perhaps the goals
  are too restrictive, or perhaps you have not sufficiently partitioned
  your workload into service classes.  Service classes at the highest two
  importance levels did not meet their performance goals during 83% of the
  measurement intervals being analyzed.  When service classes do not meet
  their performance goals, the Workload Manager will attempt to allocate
  system resources so the performance goals can be met.  Since the
  Workload Manager was not successful in doing this in your environment,
  you may wish to review the definition of your service classes and the
  workload assigned.
```

**Suggestion:** CPExpert suggests that you review the definition of your service classes, the workload assigned to the service classes, the performance goals for the service classes, and the importance of the performance goals. Since the Workload Manager was unable to manage system resources to meet your performance goals for the workload, some changes may be required.

- **Workload classification**. Work units are assigned to service classes by the workload classification scheme. You may have inadequately specified the workload classification scheme (with corresponding assignment of work units to service classes). The result may be that the workloads conflict with respect to their resource demands as matched against the performance goals.

  Many of the rules CPExpert produces suggest that the workload/service class scheme be reviewed. The context of the suggestions (that is, the problem which caused the rule to be produced) should guide you in selecting a different classification scheme if the scheme needs to be revised.

- **Service class definition**. The service class definition often is a trade-off between defining too few service classes and defining too many service classes.

  - **Defining too few service classes**. A service class normally should describe a workload consisting of applications or transactions with roughly the same performance goals, goal importance, and resource demands. Resources are allocated to service classes, based on service class performance goals and goal importance[1]. If the service classes do not adequate discriminate among the applications and

[1]Some Workload Manager actions may be at the address space level. For example, expanded storage policies are different depending upon the performance goal duration. The Workload Manager applies expanded storage policies at the service class level if the performance goal is 20 seconds or less. The Workload Manager applies expanded storage policies at the address space level if the performance goal is greater than 20 seconds.

transactions, the Workload Manager may not be able to allocate resources to meet the varying resource demands.

- **Defining too many service classes**. The Workload Manager will adjust system resource allocation in an attempt to improve performance of only one service class during a policy adjustment interval[2]. Adjustment to improve performance of only one service class is done because the Workload Manager must observe the results of the adjustment; whether the adjustment helped performance, hurt performance, or had no effect. If adjustments were made to improve performance of more than one service class, it would be impossible to determine which adjustment helped or hurt which service class. Consequently, actions to improve the performance of only one service class is done during each policy adjustment interval.

  The policy adjustment interval is 10 seconds. If too many service classes have been defined, the Workload Manager may be able to adjust system resource allocation to help only a few service classes (the most important service classes with the worst performance). Performance of other service classes may never be improved, or performance improvement actions may take a long time - simply because of the elapsed time necessary for the Workload Manager to make changes, collect data, analyze the effect of the changes, make additional changes, collect more data, and continue the process.

  Perhaps of equal significance is the overhead associated with analyzing service classes. The Workload Manager tries to improve performance of the service class with the worst performance at the highest goal importance. Resources may be taken from the least important service class with the best performance. The Workload Manager will not simply remove and add resources; rather, the Workload Manager will analyze the net value of the planned action.

  The Workload Manager will not add resources unless there is an appreciable net gain to the service class receiving the resources. Within the same goal importance, the Workload Manager will not remove resources from a service class unless the net gain to the receiver outweighs the net loss to the service class the resources are being removed from. The overhead involved with the analysis and decision process increases as the number of service classes becomes large.

[2]The Workload Manager will consider adjustments to improve performance of several service classes (starting with the most important service class which has the highest Performance Index). If performance of the first service class analyzed cannot be appreciably improved, the Workload Manager will select the next worst performing service class, etc. After the Workload Manager has "committed" to a policy adjustment for a service class, it will stop analysis and adjust resources for no other service class.

- **Performance goals**. The performance goals specified for service classes may be too restrictive or they may be specified as averages (rather than percentiles).

  - If the performance goals are too restrictive, the Workload Manager may be unable to allocate system resources to meet the goals. You should specify performance goals with values required to meet management objectives.

    You should not be unduly influenced by pre-Goal mode performance of a particular class of work. Rather, you should establish goals based on organization needs. It is possible that previous performance of certain workloads was at the expense of other workloads or at the expense of relatively poor system utilization.

    The Workload Manager will attempt to achieve high system utilization while meeting performance goals. If the goals are too restrictive, the Workload Manager may have to sacrifice overall system performance to achieve the restrictive goals of the most important service classes.

  - With an average response goal, poor response for only a few transactions can cause the average response goal to be missed. This is because the response time of a few transactions can be **extremely** long. These long response times could cause the average response to be longer than the performance goal, but the resulting average response would not be representative of the performance experienced by the typical transaction.

    With a percentile goal, the Workload Manager would not be as concerned about the few transactions which used significantly more resources and consequently skewed the average response. Rather, the Workload Manager would base its workload management decisions on the percent of transactions which met the response goal.

    You should use percentile response goals (rather than average response goals).

- **Performance goal importance**. The Workload Manager will attempt to allocate system resources so that the most important performance goals are met. In order to accomplish this, the performance goals must be adequately partitioned into highest importance down to lowest importance. Additionally, most organizations have workload which can be classified as discretionary.

  If too many service classes (and too much workload) are assigned high performance goals, the Workload Manager will be unable to meet all performance goals.

- **System resources**. The Workload Manager will attempt to allocate resources to meet performance goals. However, the Workload Manager cannot allocate resources if insufficient resources exist to meet the requirements of the workload. You might have to acquire additional resources, reschedule workload, or shift workload to other systems in order for the Workload Manager to satisfy the performance goals you have defined.

  This problem is listed last because CPExpert believes that you should make all possible tuning efforts before acquiring additional system resources.

**Reference**: MVS/ESA SP Version 5 Planning: Workload Management

MVS/ESA SP Version 5 Programming: Workload Management Services

OS/390 MVS Planning: Workload Management

OS/390 MVS Programming: Workload Management Services

z/OS MVS Planning: Workload Management

z/OS MVS Programming: Workload Management Services