Rule WLM102: Service Class did not achieve percentile response goal

Finding: CPExpert has detected that a service class period did not achieve the percentile response goal that was specified in the Service Policy in effect. This finding applies to performance goals that specify **percentile response time** as the performance goal.

Impact: This finding can have a HIGH IMPACT on performance of your computer system.

Logic flow: This is a basic finding. There are no predecessor rules.

Discussion: Service classes can be defined with a "percentile" response performance goal. A "percentile" response performance goal means that the performance goal is defined as "x%" of the transactions should complete within "y" time. For example, a typical percentile response goal is that **90% of the transactions should complete within 200 milliseconds**.

MVS accounts for each transaction executing in the system and determines the transaction's response time[1]. MVS maintains fourteen counters for each service class that has a response goal. The counters represent a response time distribution with respect to the response goal.

For response goals, RMF includes in SMF Type 72 records a count of transactions that completed in varying percentages of the response goal. These transaction counts are recorded by RMF as the "Response Time Distribution Count Table" contained in SMF Type 72(Subtype 3) records[2].

The Workload Manager periodically assesses the performance of each service class, comparing the performance achieved by the service class against the performance goals specified for the service class[3]. This assessment is referred to as the "policy adjustment" interval. During the policy adjustment interval, the Workload Manager decides whether to adjust resource policies based on whether service classes are meeting performance goals.

---

[1]This response time applies only to the time the transaction was in the system; it does not apply to response time delays experienced in the network.

[2]Please refer to Exhibit 4-11 in Section 4 for a description of the response time distributions.

[3]Please see Section 4 for a more detailed description of this process.

©Copyright 1994, Computer Management Sciences, Inc.

For service classes that have a **percentile response time goal**, the Workload Manager determines whether the specified percent of transactions are achieving the response time specified by the response goal for the service class. If more than the specified percent of transactions achieved a response greater than the specified response goal, the system is not meeting performance goals for the service class period. If the importance of the service class is sufficiently high, the Workload Manager may re-allocate system resources in an attempt to meet performance goals.

CPExpert analyzes the SMF Type 72 records to determine whether service class periods met their performance goals during each RMF measurement interval. For service class periods that have a percentile response performance goal specified, the performance goal is specified as "x% of the transactions completing within y time." CPExpert simply sums the transaction count in the first six counters to determine the number of transactions ending within 100% or less of the response goal. This value is divided by the total number of transactions ending to yield the percent of transactions ending within 100% or less of the response goal. If the resulting percentage is less than the performance goal percentage, CPExpert can conclude that the performance goal was not met.

CPExpert produces Rule WLM102 when CPExpert detects that a service class period did not meet its percentile response goal for an entire RMF measurement interval. CPExpert reports the total transactions that ended during the interval, the number of transactions that met the response goal, the percentage of transactions that met the goal, and the primary and secondary causes of response delay.

Additionally, CPExpert computes the contribution that the primary and secondary causes of delay made to the average transaction response time.

For example, suppose that an installation specified that 90% of the transactions should complete within 200 milliseconds for a service class period serving interactive TSO transactions. CPExpert might detect that only 80% of the transactions completed within 200 milliseconds, and the performance goal was not achieved. CPExpert would report the number of ending transactions, the number of transactions that met the 200 millisecond goal, and that only 80% of the transactions met the goal.

CPExpert would analyze the causes of delay to TSO transactions and report the primary and secondary causes of delay. CPExpert might compute that the primary cause of delay to TSO transactions was that they were denied access to a processor for 35% of their active time, and that

they were waiting for "unknown" causes[4] for another 30% of their active time. CPExpert would report both these causes, and their respective percentages in Rule WLM102. CPExpert would continue analysis to assess which service classes might deprive TSO transactions from access to a processor and to assess the likely causes of "unknown" delays.

CPExpert analyzes the following possible delays to response time[5]:

- **CPU Using delay**
- **Denied CPU delay**
- **CPU Capping delay**
- **Swap-in delay**
- **MPL delay**
- **Page-in delay**
- **Non-paging DASD delay**
- **Non-DASD delay**
- **Queue delay**
- **Unknown delay**

For the purposes of identifying primary and secondary causes of response delay, CPExpert combines all auxiliary storage page-in delays into "page-in delay" to reflect the impact of auxiliary storage on response.

Additionally, CPExpert computes the average Performance Index for the service class during any measurement interval in which the performance goal was not achieved. The Performance Index is computed as the actual response divided by the performance goal, but is a more detailed algorithm than the algorithm described in Rule WLM101[6].

[4]Recall from Section 4 that the "unknown" cause is unknown as far as the System Resources Manager is concerned. The SRM identifies causes of delay only for those categories over which it has control. Delays over which the SRM has no control are grouped together into an "unknown" category. These delays typically are I/O delay, ENQ delay, waiting for cross-memory services, etc.

[5]Please see Section 4 (Chapter 3.3) for a description of these delays.

[6]Please refer to Section 4 for a description of how the Performance Index is computed for percentile performance goals.

The Performance Index gives an indication of how seriously the performance goal was missed: a Performance Index of less than 1 indicates that response was less than the performance goal; a Performance of greater than 1 indicates that response was worse than the performance goal.

The following example illustrates the output from Rule WLM102:

```
RULE WLM102:  SERVICE CLASS DID NOT ACHIEVE PERCENTILE RESPONSE GOAL

   Service Class TSOUSERS (Period 1) did not achieve its response goal
   during the measurement intervals shown below.  The response goal was
   80.00 percent of the transactions completing within 0.500 seconds,
   with an importance level of 2.  The percentages with the primary/
   secondary causes of delay are computed as a function of the average
   address space active time.

                         --------LOCAL SYSTEM------
                                TRANS     %
                        TOTAL MEETING MEETING PERF PLEX PRIMARY,SECOND
MEASUREMENT INTERVAL    TRANS   GOAL    GOAL  INDX  PI  CAUSES OF DELAY
12:59-13:14,14MAR2001     97     47    48.5  2.00 4.00 I/O USING(34%),CPU USING(24%)
13:14-13:29,14MAR2001    100     44    44.0  4.00 4.00 I/O USING(39%),CPU USING(26%)
13:29-13:44,14MAR2001    114     44    38.6  4.00 4.00 I/O USING(31%),CPU USING(29%)
13:44-13:59,14MAR2001    106     54    50.9  4.00 4.00 UNKNOWN(58%),I/O USING(18%)
```

The information associated with Rule WLM102 is shown based on data collected by the *local system,* which is the system being analyzed for performance purposes.

CPExpert also computes and reports a *sysplex* Performance Index. The WLM maintains both a “sysplex Performance Index” and a “local system Performance Index.” Briefly, the WLM first examines the sysplex Performance Index to determine whether a service class period is missing its performance goal and whether action should be taken. After the sysplex Performance Index is examined at a particular Goal Importance level, the WLM then examines the local system Performance Index. Rule WLM140 explains this WLM logic in more detail, and describes the implications of the WLM logic.

Suggestion: There are no suggestions with this finding. CPExpert will continue analysis and other rules will be produced to provide more information.