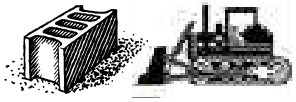
Blocked Workloads



#23 zNibbler (Blocked Workloads)
zTidBits Series

Generally, self-optimization means *self-tuning*. That is, the ability of the system to recognize that some function can help improve how it works and take the appropriate action.

In z/OS, Workload Manager (WLM) continues to help address the requirements of blocked workloads. If your system runs for extended periods at 100% processor utilization, blocked workloads can benefit from new function run through parameters in the IEAOPTxx parmlib member. With the new support, blocked workloads can proceed even when the system is 100% active for an extended period.

During periods of 100% processor utilization, **discretionary workloads** (workloads defined with a very low dispatch priority which implements the Mean-Time-To-Wait algorithm: DP=192-201) might not be dispatched. These discretionary workloads can hold serially reusable resources that other workloads require, thereby blocking the progress of workloads with a higher dispatch priority. In z/OS, you can specify that work in a given address space or enclave that is ready to run, but that does not get processor service "within a certain period of time" (starved work), is *temporarily promoted* to a higher dispatch priority.

Various enhancements such as improved enqueue and lock promotion already exist. Nevertheless, all these methods require resource monitors to recognize the contention situation and signal it to WLM. This is not always possible, because the resource managers require comprehensive instrumentation. It is possible to have a priority inversion loop, which means one resource holder blocks the work that is waiting for another (different) resource holder. These are cases where WLM does not know what work is holding the important resources and where the new function can help.

WLM gives a small amount of processor (called a *trickle*) to starved work. This support periodically examines the work coming into the queue and identifies which work units are "starved" for processor service over an extended period. WLM then "promotes" these work units to the ENQUEUE HOLD dispatch priority for the length of a time slice. When the time slice is complete, the work units have to wait for a defined time interval before they are considered for another promotion.
Two parameters in IEAOPT*xx* call this support:
* The BLWLINTHD parameter specifies the amount of time blocked work in an address space or enclave must wait before it is considered for promotion. The default for this parameter is 20 seconds.
* The BLWLTRPCT parameter specifies how much of the processor capacity to use to promote blocked workloads. This means, every second, the system uses a percentage of the processor capacity you define to help blocked work. The default for this parameter is 0.5%.

RMF supports this function by reporting relevant measurements. The Postprocessor CPU Activity report provides a new section with the following information about blocked workloads:
PROMOTE RATE the number of blocked work units that can be promoted in the dispatching priority per second. This value results from the BLWLTRPCT option settings.
PROMOTE RATE - USED (%) the utilization of the defined promote rate during the reporting interval. This rate is calculated per RMF cycle and averaged for the whole RMF interval. It demonstrates how many trickles were given away (in percent of the allowed maximum) for the RMF interval. *WAITERS FOR PROMOTE* average number of address spaces and enclaves found blocked during the interval that were not promoted according to the BLWLINTHD option.
WAITERS FOR PROMOTE - PEAK the maximum number of address spaces and enclaves found blocked during the interval according to OPT parameter BLWLINTHD. The AVG value might be quite low even when there were considerable peaks of blocked workload. Thus, the peak value is also listed.
When WAITERS FOR PROMOTE is greater than zero, this means that work is being blocked for a longer period than BLWLINTHD specifies. In this case, you might want to increase BLWLTRPCT. If you have a problem with blocked work holding resources for too long and you see no waiters in the RMF data, you might want to decrease BLWLINTHD.

The Postprocessor Workload Activity report (WLMGL) indicates the amount of processor time used by transactions of a certain service or report class that were running at a promoted dispatching priority. This new field is called PROMOTED and is displayed in the SERVICE TIMES field.

It's important that you enable this function for all production systems and examine the new information in RMF at regular intervals. The default settings guarantee a small portion of the available processor capacity is spent helping blocked workloads.

Ref: OA18244 - RMF SUPPORT FOR SRM'S BLOCKED WORKLOAD FUNCTION.