
Rule DAS260: DEVICE DISCONNECT WAS MAJOR CAUSE OF I/O DELAY

Finding: CPExpert determined that device disconnect (DISC) time was the major cause of delay in delay in DASD response to critical applications for the device.

Impact: This finding may have a MEDIUM IMPACT or HIGH IMPACT on the performance of the critical application, on the device, and on the performance of other volumes attached to the cache controller.

Logic flow: The following rules cause this rule to be invoked:
 DAS200: Volume with the worst overall performance

Discussion: DISC means that there is some delay that is often (but not always) associated with a mechanical movement during which the device disconnects from the control unit (or the control unit disconnects from the channel).

With legacy systems (e.g., 3380 drives attached to 3990-2 control units), the DISC time of most concern was associated with seek (arm movement) and rotational position sensing (time waiting for the disk platter to rotate to the location where desired data resides). Considerable performance improvement efforts were directed at reducing the seek activity and reducing the rotational position sensing (RPS)¹ delays for the legacy systems. These two mechanical delays still exist for most modern *redundant array of independent disks* (RAID)² systems, but their impact can not be directly reduced with normal methods.

With modern disks, data is cached into device cache buffers that contain data read from a track on the disk platter. Using device cache buffers containing the track data eliminated the multiple-RPS delays caused by a path busy when the device tried to reconnect. Required data is read into the device cache buffer during a single rotation and stored until a path is available to transfer the data.

In addition to the cache buffer design, modern control units such as the 3990-6 or 2105 have very large cache memory installed. With cache in the

¹RPS delays were caused by a path not being available when the required data came under a device read head. Since a path was not available, the device could not reconnect to the channel or control unit. Consequently, data could not be read and transmitted, and another rotation of the platter was experienced until the data again came under the device read head. Multiple rotations might be required, depending on the busy level of the path.

²An array is an ordered collection of physical devices (disk drive modules) that are used to define logical volumes or devices.

control units, data to be read can be transferred in a variety of ways, depending on where the data resides.

For a read operation, desired data often is found in the control unit cache. If the required data is in cache, the data can be transferred between the control unit cache and the channel, and this transfer is done at channel speed. If the required data is not in cache, the data can be transferred between the device and channel (and concurrently placed into the control unit cache for subsequent access).

For write operations, data can be placed into Non-volatile Storage (NVS) as a part of the control unit. Write operations normally end as the data to be written is placed in the NVS; and the storage processor writes the data to the device asynchronous with other activity (as a “back end” staging operation). See subsequent discussion for more detail about read and write operations.

The storage director can simultaneously transfer data between the channel and device and manage the data transfer of different tracks between the cache and channel, and the cache and the device. With large amounts of cache memory, a high percent of data accesses normally will be resolved from the fast cache memory and the relatively slow device will not cause significant delays.

As a result of the above improvements, DISC time for modern systems is a result of *cache read miss* for read operations, back-end staging delay for write operations, peer-to-peer remote copy (PPRC) operations, and other miscellaneous reasons³. DISC time often can be very small with adequate cache. For example, there would be zero disconnect time for a cache read hit (the record was found in the cache). However, DISC time can be large and can cause serious delay to I/O operations.

Suggestion: Please refer to Rule DAS160 for further information about DISC time.

³Artis has described a “sibling PEND” condition that results from collisions within the physical disk subsystem of RAID devices. See “Sibling PEND: Like a Wheel within a Wheel,” www.cmg.org/cmgap/int449.pdf. While this condition is titled “sibling PEND,” the time properly belongs in DISC time, rather than PEND time .