Improvement obtained with a new file matching program in GiAPA V04M03H

For jobs that are active when data collection is started GiAPA will now assume that the number of I/Os found at the first interval triggering a HotSpot represents the value from before GiAPA was started. Consequently, GiAPA will subtract that initial number from the total number of I/Os shown for the file.

Let us illustrate this issue with an example:

A job running across several days has for a given file used 73,000,000 I/Os by the end of Day 1. The job only receives transactions during working hours (07:00 – 18:00), but is left running 24/7 awaiting further input.

To keep the information for all active jobs in memory tables GiAPA must restart every 24 hours to not overflow the tables. GiAPA data collection normally restarts at midnight. The job in question was on Day 2 not showing any activity prior to 07:00, where it again started using sufficient CPU to generate a GiAPA HotSpot collecting open file information.

Prior to this release such a file, not showing any I/O activity close to beginning of the data collection, has been identified as a new file where all I/Os should be shown as belonging to the actual day.

If this file on Day 2 showed e.g. 73,000,111 I/Os at 07:00:15, and 75.333.555 I/Os in the last HotSpot for the day, GiAPA would report all 75,333,555 I/Os as belonging to Day 2, and the total for the file for the two days would be 146,333,555. This was obviously not correct.

The ideal solution – i.e. to pass all file statistics automatically from one day to the next – would be too complicated and resource consuming to implement.

Therefore, we recoded GiAPA for jobs running across several days to automatically subtract the first I/O value found. In the above example GiAPA will now report that the file on Day 2 used 2.333.444 I/Os, which is much more accurate. Collecting data "only" every 15 seconds this value cannot be 100 % correct, however, GiAPA now provides a value sufficient for a reliable impression of the file usage even for jobs active several days.

An implication of the new way of calculating is that if the job happens to open a new file after waking up at 07:00 then the first I/O count for the file is tallied as belonging to the previous day i.e. it will be subtracted from the final number of I/Os for the file. This is not correct, but does not impact the overall picture given that only I/Os found by the first HotSpot are subtracted.

The above described change does not influence data originating from several collections within the same date. Even if GiAPA was restarted during a day, all data collected the same date will be analyzed as one set of data.