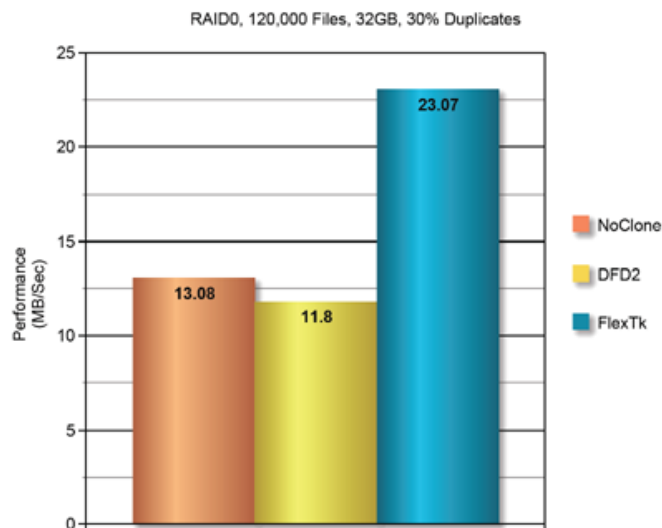


## FlexTk Duplicate Files Search Performance Review

Searching for duplicate files is a computing and I/O intensive operation that is requiring both a fast hard disk and a powerful CPU. Depending on the number of files that should be processed and the number of existing duplicates, a search operation may take from a couple of minutes for a few hundreds of files to several hours in the case that you need to scan many thousands of files located on multiple disks or enterprise storage systems.

The main purpose of the performance review is to provide our customers with an estimate performance and an expected scalability level of the FlexTk's built-in duplicate files finder on different hardware configurations and data sets. In addition, we have compared our own software to latest versions of other popular duplicate file finders - NoClone and Duplicate Files Detective 2 (DFD).



All performance tests were performed using FlexTk v2.6.12 on a PC machine equipped with a dual-core 2.4 GHz E6600 Intel CPU and 2 GB of system memory running Windows XP Professional (32-Bit). In order to analyze performance on different types of files we have prepared the following three data sets:

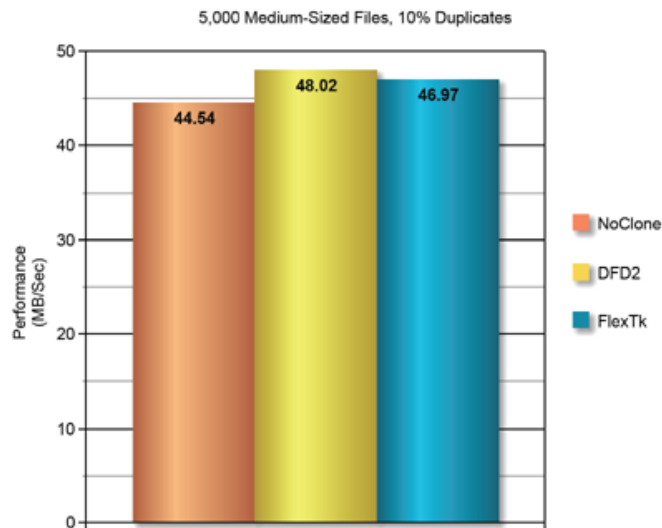
- File Set #1 - 15GB, 5,000 Medium-Sized files with 10% duplicates
- File Set #2 - 3GB, 55,000 Small-sized files with 10% duplicates
- File Set #3 - 32GB, 120,000 Files of various sizes with 30% duplicates

In order to analyze duplicate files search performance on different hardware architectures, we have replicated all the three data sets to the following storage devices:

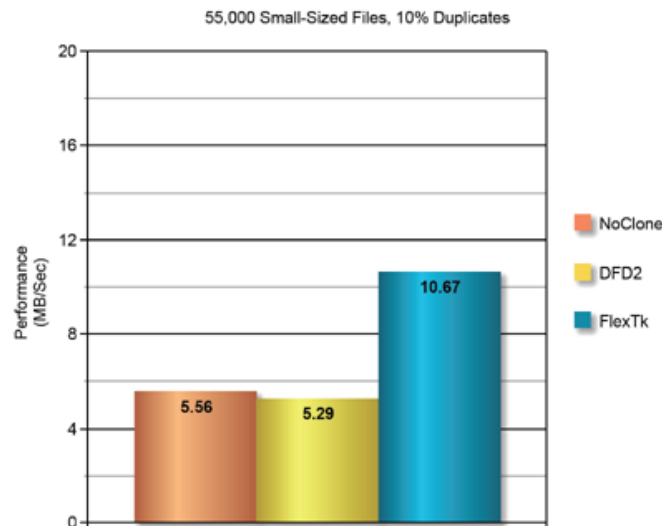
- Storage Device #1 - 150GB, Western Digital Raptor
- Storage Device #2 - 2x150GB, Western Digital Raptor in RAID0 configuration
- Storage Device #3 - 2TB NAS Storage connected through a Gigabit Ethernet link
- Storage Device #4 - 500GB, Western Digital USB Disk

Each software tool was executed once for each data set on each hardware configuration with a system reboot before each benchmark - 12 benchmarks per tool and 36 benchmarks in total. Individual benchmark results from all four different hardware configurations were averaged and finally we have prepared three graphs representing average duplicate files search performance for small files, medium-sized files and mixed files.

As it is clear from the first graph, all three tools are delivering very close performance results while processing a small amount of medium-sized (2MB-5MB) files. From the performance point of view, this is the best case scenario and if you need to process a couple of hundreds of files any software tool will do the job.

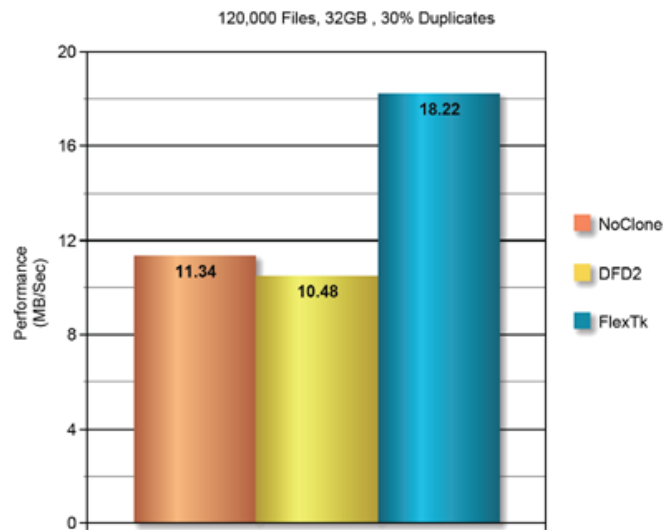


The second data set contains 55,000 small (1KB-200KB) files, mostly Word documents and images. Here the situation changes dramatically and we can see the performance of NoClone and DFD dropping very significantly. During performance tests we have identified a common pattern that exists in NoClone, DFD and other tools, which are not mentioned in this review. The issue is that all tools that we have tested are starting the duplicate files search operation very fast, but getting slower and slower as more files should be processed.



From the beginning, the FlexTk's built-in duplicate files finder was designed as a scalable solution capable of processing millions of files at a sustainable speed. In addition, the FlexTk's built-in duplicate files finder is capable of parallelizing all processing operations and effectively utilizing modern hardware architectures including multi-core CPUs, disk RAIDs and Gigabit Ethernet networks.

The last data set confirms the identified pattern - the FlexTk's built-in duplicate files finder is capable of processing large data sets significantly faster and more effectively.



*\* This performance review has been prepared for information purposes only and we strongly advise you to make your own performance evaluations using your specific hardware components and datasets.*