CSc 841/ 641: Computer Performance Evaluation Dr. Jozo Dujmović
Project: Benchmarking disk performance

The purpose of this project is to develop your own disk benchmark program. The goal is to measure computer performance in the area of file input/output workloads. Your DISKben benchmark should be similar to the CPUben benchmark (HW#1).

Disk performance depends on several main factors: disk access time, data transfer rate, buffering, caching, and disk controller organization and performance. All disk workloads combine two types of operations:

- Write
- Read

These operations can be performed with two fundamental forms of access:

- Sequential access
- Random access

Therefore, four fundamental activities of file input/output are:

- Sequential write
- Sequential read
- Random write (frequently omitted to avoid the "read before write" problems)
- Random read

In this project you have to develop a disk benchmark program DISKben (preferably in C++) that measures the disk file I/O speed as follows:

a. Write a function DSEQ that creates a sequential binary file with arbitrary contents. Then read sequentially the whole file. That includes sequential write, sequential read and the corresponding file open and close operations. If you need to repeat these operations you may delete the created file using the system( ) function from <process.h>. Adjust your function so that it runs approximately 1 second or more. Let T_{seq} be its run time. The corresponding processor speed in the area of sequential disk access is $V_{seq} = 60/T_{seq}$ [operations per minute, or megabytes transferred per second].

b. Write a function DRAN that creates a binary file for direct access with arbitrary contents. Then read randomly the whole file. That includes the corresponding file open and close operations. If you need to repeat these operations you may delete the created file using the system( ) function from <process.h>. Adjust your function so that it runs approximately 1 second or more. Let T_{ran} be its run time. The corresponding processor speed in the area of random disk access is $V_{ran} = 60/T_{ran}$ [operations per minute, or megabytes per second].

c. Adjust your programs so that approximately $V_{seq} = V_{ran}$. Write a main program DISKben that runs DSEQ for 10 seconds and measures V_{seq}, and then runs DRAN for 10 seconds and measures V_{ran}. Measure V_{seq} and V_{ran} as precisely as possible. Compute the average overall disk speed $V_{disk}=2/(1/V_{seq} + 1/V_{ran})$. Display V_{seq}, V_{ran}, and V_{disk}.

Use DISKben to compare at least three different computers. Make sure that all benchmarks are compiled and built as release versions with the maximum level of optimization. Perform the following experiments: (1) measurement of differences between release and debug versions of your programs, and (2) program execution in various environments (Windows, Cygwin, and Linux) using the same hardware. Write a report about design of your benchmark, comparison of competitive computers, experiments that you performed, and your experience with benchmarking. Additional (extra credit) work might be to write DISKben in various languages and/or using various compilers, and analyze how benchmarking results depend on the selected language or compiler. Take into account that modern operating systems use the whole free memory as a disk cache, and in such cases the resulting speed reflects the performance of processor, memory, and the disk caching efficiency yielding rather high values of the disk speed indicators.