CSC443 Assignment 1

Zhe Yuan Lin #1002278705

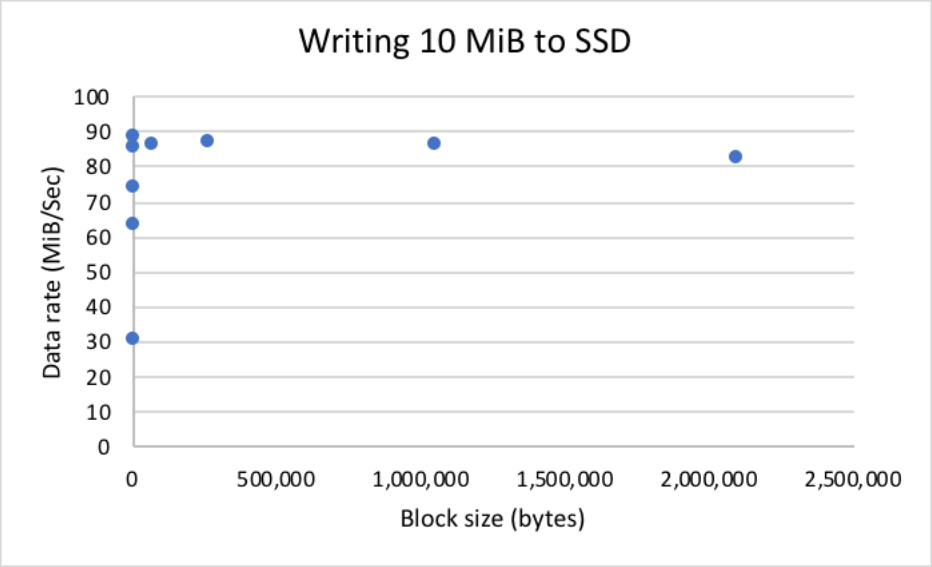
Sequential write to file

Experimented by writing a 10 MiB file to a solid state drive and a USB flash drive, using the following block sizes:

128 B, 512 B, 1024 B, 4096 B, 8 KiB, 64 KiB, 256 KiB, 1 MiB, 2 MiB

1. Solid State Drive

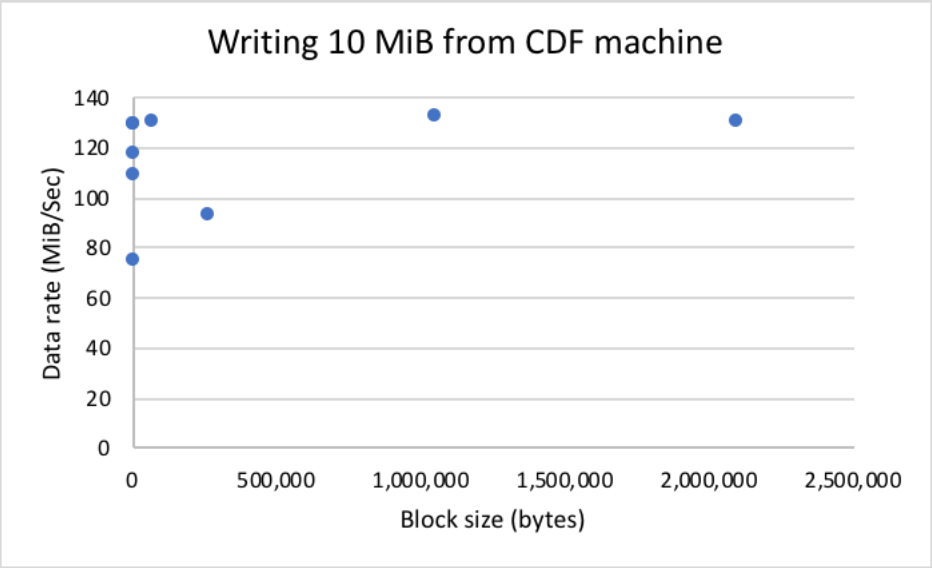
   We ran the Python 3 script create_files_script.py, which writes 10 MiB sized files to a directory, on the solid state drive on a mid-2014 Macbook Pro on MacOs Sierra. The data we got was the following:



   The data rate stabilizes at around block sizes 4096 B (4 KiB) and 8 KiB, after which the data rate is constantly between 82-87 MiB per second. The data rates of block sizes below 4 KiB starts off very low and increases drastically as the block size goes up to around 4 KiB. This can possibly be due to the reason that solid state drives write data in pages, and this solid state drive had a page size of 4 KiB or 8 KiB, thus any block size smaller than the SSD's page size had lower performance, as the SSD has to load the page into memory and write it back several times.

2. Distributed file system

Again, we ran create_files_script.py, but this time on the CDF machines in Bahen 2210.



This curve is similar to that of the solid state drive. The data rate starts of slow at 128 B and increases until 4 KiB or 8 KiB. However, after running the script multiple times, the block size of 256 KiB seems to cause a dip in performance. The reason for this is unknown, but otherwise the curve does mimic that of the solid state drive, making is a good possibility that the CDF machines run on a solid state drive.

Possibility of optimal block size for write

There doesn't seem to be an optimal block size for write, as it depends on the situation (i.e. page size of the SSD, etc.). In my experiments, the block size of 64 KiB seems to do well, however, the larger block sizes do just as well.

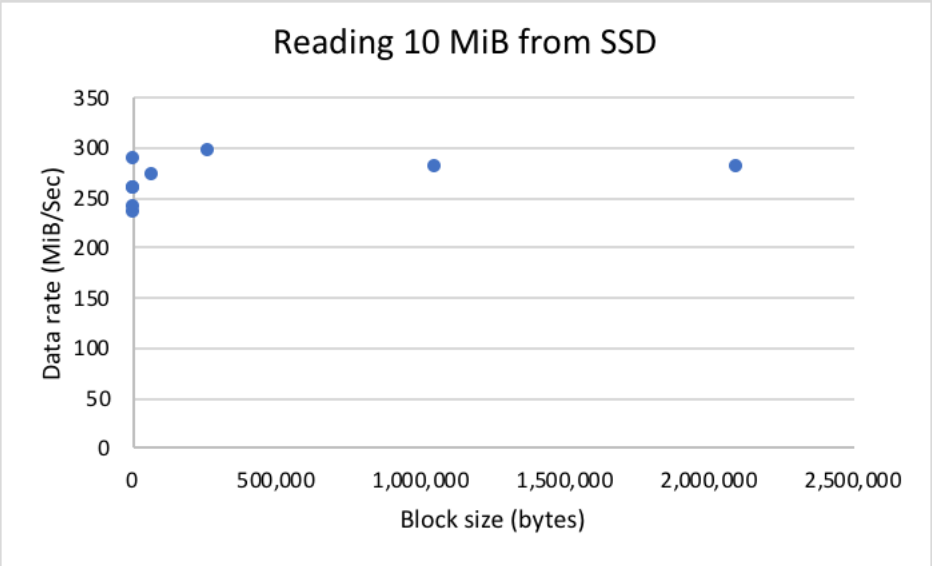
Sequential read from file

Experimented by reading a 10 MiB file called TestFile to a solid state drive and a USB flash drive, using the following block sizes:

128 B, 512 B, 1024 B, 4096 B, 8 KiB, 64 KiB, 256 KiB, 1 MiB, 2 MiB

1. Solid State Drive

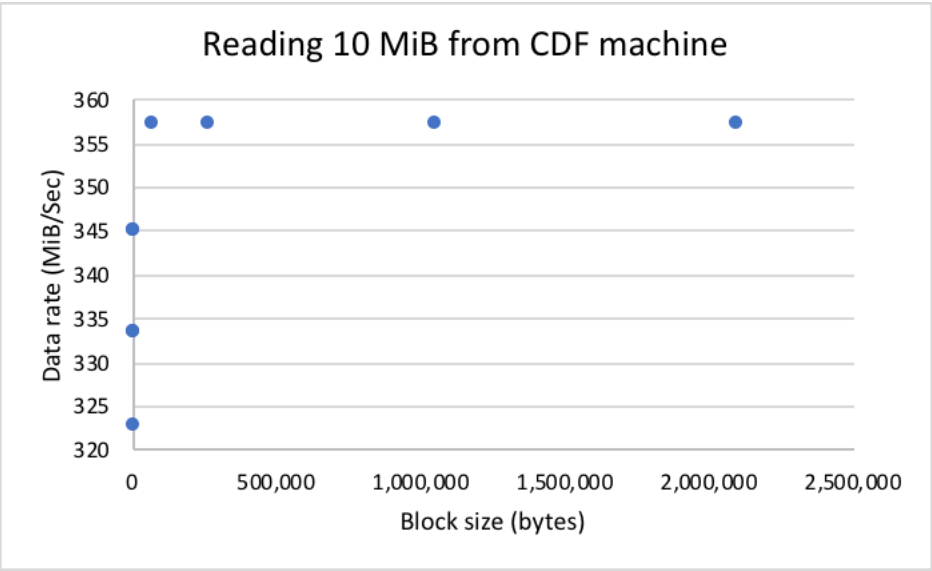
We are reading from the same SSD as used in the sequential write experiment.



Again, just like the writes, the data rates of reading from the SSD seem to increase and stabilize up to 4096 B. This is due to the fact that, just like how SSDs write, they also read in pages. As we discussed before, the page size of my SSD seems to be either 4 KiB or 8 KiB, therefore, as shown in the graph, reads with a block size smaller than 4 KiB is more costly than those that are larger.

2. Distributed File System

We are reading from the CDF machines in Bahen 2210, just like the writes.



Again, this behavior is very similar to that of the SSD. Data rates increase up to 64 KiB and stabilizes after that at around 357 MiB/S. This further confirms our suspicion that the CDF machines runs a solid state drive.

Possible optimal block size for read

Again, just like the optimal block size for write, I don't believe this exists and it depends on the situation. For example, in SSDs, anything bigger than the page size seems to do good.