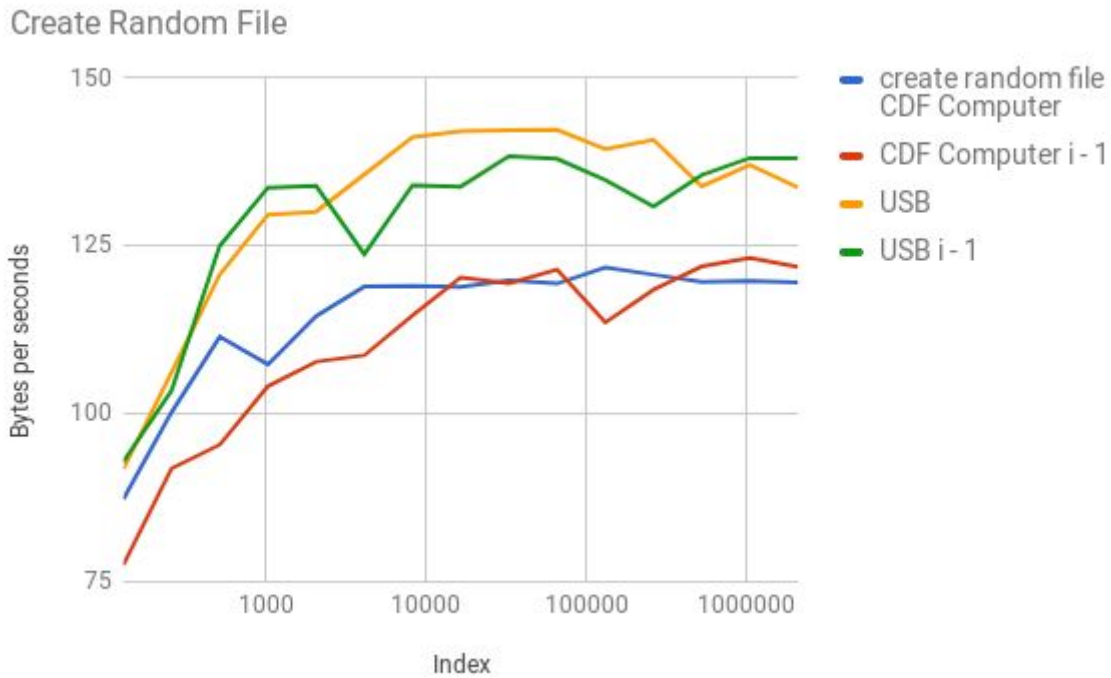


# Assignments 1

## Write Data Rate Versus Block Size



### Discussion on methods

For each write buffer size we performed a write of 30000000 characters 30 times and took the average time and plotted it above. The choice of block sizes was to have them be powers of 2. That choice was because we expect that the computer will have a write buffer that is also a power of 2. We started with a buffer of size 128 going to a buffer size of 2097152. The lines that have i - 1 mean that we started with buffer 127 and doubled it each time. I additionally performed this experiment with a buffer size starting at 129 but failed to see any noticeable pattern and so omitted the data in the chart above.

The point of writing with a reduced buffer size was to see if writing less than the buffer would have a noticeable effect. You can see that directly from the computer on CDF that this did have an effect. As the buffer becomes larger this effect starts to disappear. I additionally performed this experiment with

### Discuss the existence of the optimal block size for write

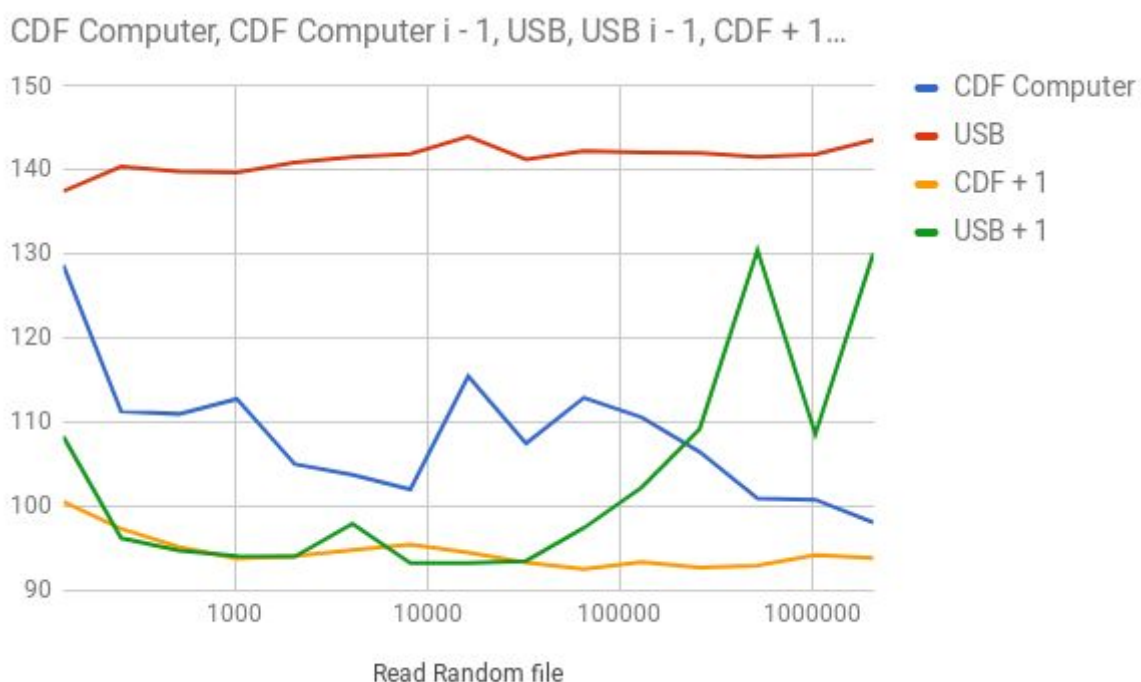
When looking at the chart above we can see that the bytes written per second increases between the start and around 8192 then it plateaus becoming only slightly better with bigger

buffer sizes. In the context of a DBMS the longer you wait to fill up a large buffer size to write the larger the chance that something unexpected will happen and you'll have to recover from your last write point. Meaning that we have a preference on smaller buffers. Knowing that we prefer smaller buffers and that we don't see large performance gains after 8192 we can conclude that 8192 would be the optimal size buffer for the machine used.

## Comparison on different storage medium

When comparing the performance of the USB and the memory used by the cdf machine we found that the USB did noticeably better. This can likely be attributed to the way that CDF Stores its memory. The memory that would normally be on disk is on a server and not at the local machine I was using. So the USB being within the machine does not have to go through the extra steps of sending the data to and ensuring that the server receives the data and is able to write to that USB on its own.

## Read Data Rate Versus Block Size



## Discussion on methods

For each read, we performed a write of the file it was to read and then read the file and repeated this 30 times and took the average time taken and plotted it above. I chose block sizes to be powers of 2. This choice was because we expect that the computer will have a read buffer that is also a power of 2. We started with a buffer of size 128 and went to a buffer of size 2097152. The lines that have i - 1 mean that we started with buffer 127 and doubled it each time.

Reading with an increased buffer size showed noticeable improvement with a larger buffer in comparison to a smaller buffer. But still was noticeable worse than the buffer size that was a power of 2.

## Discuss the existence of the optimal block size for read

Reading from the file does not seem to have any obvious patterns. It seems fairly random with a trend of improvement for the USB and a slight trend of degradation for the CDF machine. From the chart above I would conclude that no buffer size outperformed any other as long as the buffer was a power of 2 and that due to the downward constraint on size mentioned above that the best buffer size is 128.

Although concluding this from the chart it is absurd to say that the IO is okay with any buffer size. I expect that the results above are due to the IO caching the file. So although the file is read at a fast rate it is only because we are reading it contiguously and hence can take advantage of caching. We can see this as when we have a larger buffer size and the IO can reasonably prefetch the upcoming data blocks we see a large improvement on the USB.

In the context of a DBMS we often only need to read data contiguously on given pages. So it isn't the case that the simulation in the above paragraph would matter. It is fine to assume that any read size that is a power of 2 is okay due to caching of the pages. But we would still expect to see problems if you have too small a buffer size and the IO is unable to cache the pages you need.

## Comparison on different machines and different storage medium

The USB did far better than the CDF machine. This can be attributed to the same facts discussed above in the same section on writing to a USB.