

## CSC443 A2 report

Maximum possible run size is 200MB.

We need  $\text{ceil}(700\text{MB}/200\text{MB})=4$  runs if we want the run size to be 200MB.

```
Rashids-MacBook-Pro:a2 rashid-laptop$ /usr/bin/time -lp ./disk_sort
edges.dat 209715200 8000 4
```

Maximum number of followers: 564512

Average number of followers: 12.876

```
real      16.27
user      13.91
sys        2.11
209633280  maximum resident set size
          0  average shared memory size
          0  average unshared data size
          0  average unshared stack size
92855     page reclaims
          3  page faults
          0  swaps
          0  block input operations
72        block output operations
          0  messages sent
          0  messages received
          0  signals received
120       voluntary context switches
3312     involuntary context switches
```

- 1/2 of the original size is 100MB.

We need  $\text{ceil}(700\text{MB}/100\text{MB})=7$  runs.

```
Rashids-MacBook-Pro:a2 rashid-laptop$ /usr/bin/time -lp ./disk_sort
edges.dat 209715200 8000 7
```

Maximum number of followers: 564512

Average number of followers: 12.876

```
real      16.95
user      14.80
sys        2.04
306556928  maximum resident set size
          0  average shared memory size
          0  average unshared data size
          0  average unshared stack size
74855     page reclaims
          0  page faults
          0  swaps
          0  block input operations
98        block output operations
          0  messages sent
          0  messages received
```

```
0 signals received
90 voluntary context switches
4099 involuntary context switches
```

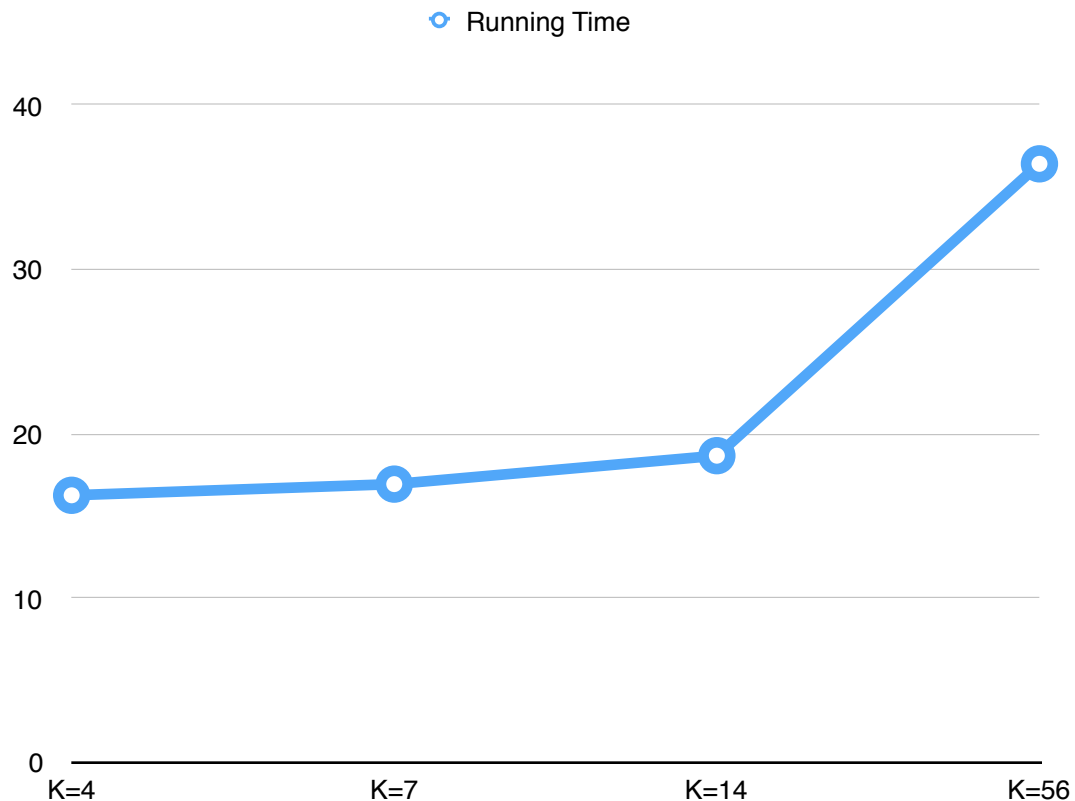
- 1/4 of the original size is 50MB.  
We need  $\text{ceil}(700\text{MB}/50\text{MB})=14$  runs.

```
Rashids-MacBook-Pro:a2 rashid-laptop$ /usr/bin/time -lp ./disk_sort
edges.dat 209715200 8000 14
Maximum number of followers: 564512
Average number of followers: 12.876
real      18.68
user      16.60
sys       1.93
256368640 maximum resident set size
0 average shared memory size
0 average unshared data size
0 average unshared stack size
62602 page reclaims
0 page faults
0 swaps
0 block input operations
107 block output operations
0 messages sent
0 messages received
0 signals received
175 voluntary context switches
2338 involuntary context switches
```

- 1/16 of the original size is 12.5MB.  
We need  $\text{ceil}(700\text{MB}/12.5\text{MB})=56$  runs.

```
Rashids-MacBook-Pro:a2 rashid-laptop$ /usr/bin/time -lp ./disk_sort
edges.dat 209715200 8000 56
Maximum number of followers: 564512
Average number of followers: 12.876
real      36.44
user      33.88
sys       2.10
222117888 maximum resident set size
0 average shared memory size
0 average unshared data size
0 average unshared stack size
54242 page reclaims
0 page faults
0 swaps
0 block input operations
109 block output operations
0 messages sent
```

```
0 messages received
0 signals received
416 voluntary context switches
30894 involuntary context switches
```



I think the reason that the running time is increasing as the number of runs grow is the cost of getting the next smallest element among buffers. My program does a linear search to find the smallest element and it has to search for each record in the file.

Let  $N$  be the number of records in the file.

When the number of runs is 56 we scan 56 buffers  $N$  times which is 14 times more operations than 4 runs. Even though we are increasing the complexity on ram, I think it is sufficient to cause the program to run slower.

I couldn't get the results from the sort function of UNIX, because it doesn't terminate. I see that it uses 1GB of ram, but it doesn't produce any result.