

Data Systems

Assignment - 2

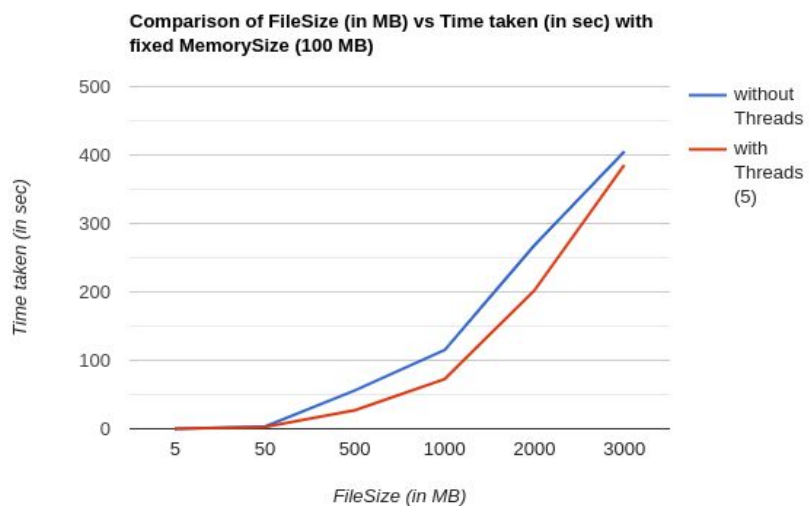
System Configuration:

- **Processor** - Intel® Core™ i5-7200U CPU @ 2.50GHz × 4
- **Memory** - 15.5 GiB
- **Disk Capacity** - 1.0 TB
- **OS Name** - Ubuntu 20.04.1 LTS
- **OS Type** - 64-bit

Observations:

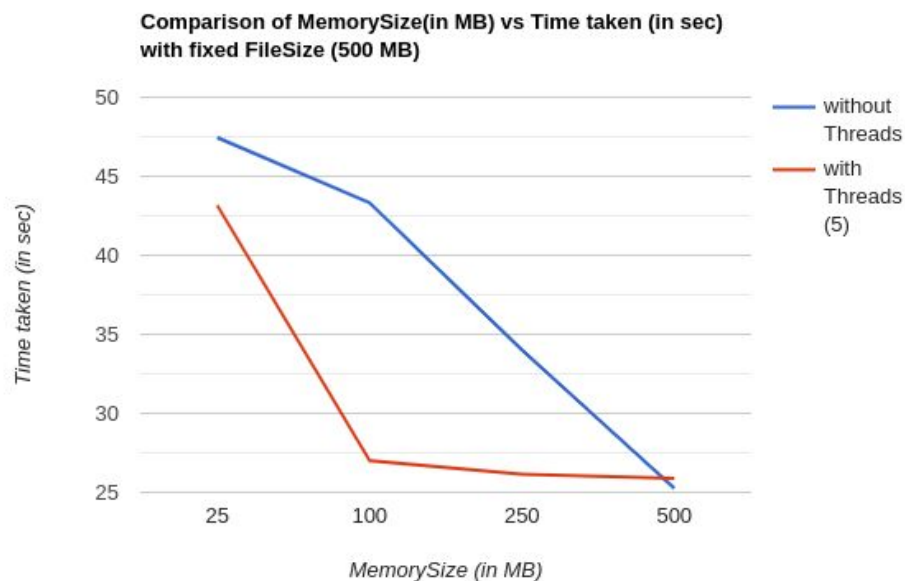
- **Varying FileSize with constant memory (100 MB):**

<u>File Size (MB)</u>	<u>Time Taken (sec)</u> (without threads)	<u>Time Taken (sec)</u> (with threads = 5)
5	0.23	0.22
50	3.25	2.57
500	55.98	27.00
1000	115.06	72.46
2000	268.30	202.41
3000	405.10	385.34



- **Varying Memory Size with constant FileSize (500 MB):**

Mem. Size (MB)	Time Taken (sec) (without threads)	Time Taken (sec) (with threads = 5)
25	47.45	43.14
100	43.32	27.00
250	34.02	26.15
500	25.26	25.89



- **Explanation -**

1. Part 1 - It is observed that as the file size increases with the memory size as constant (100 MB), the time taken for sorting also increases as the number of intermediate files generated also increases and hence sorting them one by one and writing them to the output file adds up to an increase in the total time taken.

In case of threads, the time is lesser as compared to when not using threads because of the parallelization of phase-1 but that's not always the case specially when coming to the bigger file sizes. According to me this is because the threads are sharing and acting on a common resource, i.e

input.txt. The partitions are generated using only a single input file which slows the process and eventually results in more time.

2. **Part 2** - It is observed that as the main memory size increases with the input file size kept constant, i.e 500 MB. The time taken for two phase merge sort decreases gradually. This is because as the main memory size increases, the size of the intermediate files which can be brought to the main memory for sorting increases and eventually the number of intermediate files to be generated decreases significantly.

In case of threads, the time is less as compared to no threads because of the parallelization of the first phase, but as the main memory size increases the difference becomes shorter.