

This folder contains the code to sort large files with limited RAM using heap sort.

It contains the following folder-

1. Output_file= Contains the final sorted text file obtained at the end of the execution of the code.
2. Input_file= Contains the input file which is unsorted. This file consists of strings of length 10 per line.
3. code- Consists of .java code for sorting.

Idea- The whole point sorting algorithms is to sort and make it easy for humans to read and use data for further analysis. In this assignment, I have used the same idea of sorting to sort a very large file with limited RAM. The code functions in the following ways-

A. My code generates the file (unsorted) which needs to get sorted. Since I am targeting sorting strings, it generates a large file of specified size with unsorted string of length 10 on each line. The file size to be generated needs to be mentioned in advance before running the code.

[LARGE_FILE_SIZE_IN_MB](#) value in my code is 200MB for demo purpose. Please feel free to change it to 4GB if you want. This file that is generated gets deleted at the end of the code so there is no storing file on the system taking place in this case.

B. After the file is created, it is distributed into chunks of size again specified by the user depending on the available RAM space.

C. Each chunk is then sorted using heap sort and stored in temp files. If there are 5 chunks to be sorted, there will be 5 temp files created during this phase.

D. Each chunk is then merged into one file which contains the sorted elements.

The following output was obtained when I ran my code for testing purpose. You can change the file size and run it again to check.

```
<terminated> heapsort (1) [Java Application] /Library/Java/JavaVirtualMachines/jdk1.8.0_71.jdk/Contents/Home/bin/java (04-Dec-2017, 8:25:38 pm)
Created file of size 100MB at '/var/folders/80/0vpx615j41j8wgt1s42txvrc0000gn/T/largeDemofile8148150200479710291.txt'
Number of chunks to be sorted are = 5
bytesToReadPerChunk = 4194304 bytes
Created sorted file at '/var/folders/80/0vpx615j41j8wgt1s42txvrc0000gn/T/ResultingSortedfile8022618577705717082.txt'
The program takes 242 seconds to run
```