TWO PHASE MULTI WAY MERGE SORT

SUBMITTED BY 40093368 | 40083709 | 40091123

INTRODUCTION:

The two phase multiway merge sort is used to sort datasets that are extremely large and cannot fit into the main memory in a single instant of time. In such cases the dataset will be processed in two phases over extended periods of time to sort the data. In this technique we have number of input and output buffers to hold the dataset in execution. Merge sort is used in such scenarios since they are very useful in processing large data. The TPMMS is an external sorting algorithm

PHASE 1:

Let's assume we have M tuples in a relation of N bytes of B blocks size. We have a main memory whose capacity is limited to R bytes. So our first phase will have M*N byte size of data to be processed. R/B is the number of buffers in main memory under usage. During the course of phase 1, N tuples will fill the buffers in R and the execution takes place.

PHASE 2:

After the end of first we will have sub lists of order MN/R. In this phase we will have an output buffer to write multiple sub lists into a single sub list. The number of input buffers to be used is calculated using the relation between main memory and number of blocks in it.

TEST CASE:

Input size	Time taken (Seconds)
923 KB	0.231761099
1 KB	0.031655001

CONCLUSION:

Thus the huge dataset to be sorted is successfully processed. This is an effective way to sort huge amount of data as the overhead in moving data to and from the disk and main memory is significantly less compared to other sorting techniques.

REFERENCES:

- http://www.mathcs.emory.edu/~cheung/Courses/554/Syllabus/4-query-exec/2-pass=TPMMS.html
- https://www.coursehero.com/file/pbbckd/Two-phase-Multiway-Merge-sort-2PMMS-sort-a-relation-R-with-M-available-main/
- https://www.geeksforgeeks.org/external-sorting/