

Database System Assignment 2

Twoway Merge Sort

Chaitanya Patel - 201501071

Configuration of the system

- HP Pavilion p26x
- Intel i5 5th Gen 2.2 GHz
- 8 GB RAM
- 1 TB hard drive

Observations : Memory size 100 MB with varying file size

- All files were generated by `gensort` utility.
- Each record has 3 fields of length 10, 32 and 52.
- First field is used as sort key.
- Observations are as below.

File Size	User Time	Real Time
5 MB	0m0.864s	0m0.890s
50 MB	0m9.520s	0m9.904s
500 MB	1m39.364s	1m41.895s
1 GB	3m15.704s	3m19.731s
2 GB	6m33.312s	7m58.771s
3 GB	9m45.205s	10m57.105s

Note

- Reading of file is done in the units of whole chunk i.e. the whole needed chunk of file is read at once. If the file is read line by line, then there is little performance decrease.

- On contrary to that, writing should be more fast if it is done by chunks. But due to inbuilt buffer management system of the OS itself, there was quite little performance difference if the writing is done line by line.

Analysis

- Each sorting routine read and write twice in whole process.
- We can observe that twoway merge sorting is I/O bound process. Time taken to sort the file increase linearly with increasing file size (not by $n \log n$ as in the case of normal merge sort).
- With limited 100 MB main memory size, towway sort is able to sort file as big as 3 GB in just 9 minutes.

Observations : File size 500 MB with varying main memory

Main Memory	User Time	Real Time
5 MB	1m21.032s	1m27.467s
25 MB	1m29.608s	1m35.901s
100 MB	1m37.296s	1m46.986s
250 MB	1m43.012s	1m51.165s
500MB	1m47.132s	2m8.492s

Note

- Here inbuilt OS buffering system will again play a major role in performance. Reading or writing chunk by chunk will only slightly be faster than reading or writing line by line.

Analysis

- Here we observe that, with varying main memory size, the time taken for sorting doesn't change much.
- It again reconfirms the statement that twoway merge sort is I/O bound process.