CS246: Database Management Systems Lab

Lab # 01 (1 Questions, 50 Points)
Timings: 14:00 to 17:00 Hours Pages: 3

IIT Guwahati

07 Jan 2020 (Tue)

Question 1: (50 points)

Implement in C or C++ Programming Language the following problems. Use of any programming language other than C or C++ is not allowed and will lead to awarding 0 marks.

Problem Statement Given input data files containing one data item per line, your task is to:

Read each input file (input file generator specification is detailed shortly)

Sort the data using (a) Quick sort and (b) Merge sort

Write the sorted data to a new output file (file naming: qi10k.txt, qs10k.txt, qf10k.txt for quick sort on 10K integer data file, 10K string data file and 10K floating point data file respectively; mi10k.txt, ms10k.txt, mf10k.txt for merge sort on 10K integer data file, 10K string data file and 10K floating point data file respectively)

Measure The time taken for (a) sorting alone (b) sorting and writing into a file

Output Your program output should be of the following format

Quick Sort		
Description	Only sorting	Sorting and writing into file
Integers 10K	0.100 sec	0.200 sec
Integers 20K	0.200 sec	0.400 sec
...	...	...
Strings 10K	1.200 sec	1.800 sec
Strings 20K	1.800 sec	2.300 sec
...	...	...
Float 10K	0.200 sec	0.400 sec
Float 20K	0.400 sec	0.800 sec
Merge Sort		
Integers 10K	0.100 sec	0.200 sec
Integers 20K	0.200 sec	0.400 sec
...	...	...
Strings 10K	1.200 sec	1.800 sec
Strings 20K	1.800 sec	2.300 sec
...	...	...
Float 10K	0.200 sec	0.400 sec
Float 20K	0.400 sec	0.800 sec

Your task is to write input generator program having three distinct functions to generate several sets of input files as per the following specification:

Integer input generator Randomly generate

1. 10,000 integers between 0 and 1000000. Each line containing one integer.
2. Save them in a file i10k.txt
3. Repeat steps (1) and (2) by increasing the input size by a factor of 2. That is 20,000 integers; 40,000 integers; etc. Save the output into files i20k.txt, i40k.txt etc.
4. Generate 10 such inputs.

String input generator Randomly generate

1. 10,000 strings each having length 10. Each line containing one string.
2. Save them in a file s10k.txt
3. Repeat steps (1) and (2) by increasing the input size by a factor of 2. That is 20,000 strings; 40,000 strings; etc. Save the output into files s20k.txt, s40k.txt etc.
4. Generate 10 such inputs.

Float input generator Randomly generate

1. 10,000 floating point numbers between 0.00 and 10.00. Each line containing one floating point number.
2. Save them in a file f10k.txt
3. Repeat steps (1) and (2) by increasing the input size by a factor of 2. That is 20,000 floating point numbers; 40,000 floating point numbers; etc. Save the output into files f20k.txt, f40k.txt etc.
4. Generate 10 such inputs.

Instructions Adhere to the following

File naming Prepend C/C++ program file names with your roll number. Adhere to the input and output file naming convention as given in the problem description.

Independent efforts You should make an honest and independent effort in obtaining the solution to the above problem. You are also encouraged to bring one data structures and algorithms text book and one programming language text book of your choice.

Discussions with fellow students are not allowed.

Intenet Use of internet during lab hours is not allowed.

Evaluation At the end of 17:00 hours, TAs will come and evaluate your program. Leave the lab once your evaluation is completed.

Marking Scheme The evaluation criteria is as follows:

25 Marks For input generator

Input file names 5 Marks for constructing correct input file names.

Integer inputs 5 marks for generating 10 sets of integer inputs.

String inputs 10 marks for generating 10 sets of strings.

Float inputs 5 marks for generating 10 sets of floating point numbers.

5 Marks For reading input data into appropriate data structures.
3 Marks Correct implementation of quick sort.
2 Marks Correct implementation of merge sort.
5 Marks For constructing the correct output file names.
10 Marks For obtaining the output in the given output format.