DEPARTMENT OF COMPUTER SCIENCE, HANYANG UNIVERSITY

Sorting

Concurrent Programming

Programming Project #1

Final due date: October 02, 2019 (HARD DEADLINE)

1 TASK OVERVIEW

Sorting remains a key primitives for many uses of database systems. The task is to sort three datasets of varying size, skew, and data types. The key constraints here are that memory usage is limited, as only 10 GB of memory is allowed while the datasets vary in size from 1 GB to 4 GB.

Input to your program will be provided by a path to an input file to partition or sort and output must be output to a specified output file.

2 TASK DETAIL

Our test harness will first give the path to the input file to sort to your program's first argument, and the path to the output file to your program's second input. The files to be

used are in the format generated by gensort, where the first 10 bytes are the key and the last 90 bytes the payload. We will iteratively give your program each of the three files one at a time once we have verified the sorting of the current file is correct. Sorting will be validated by valsort. The data sizes are as follows: 1 GB, 2 GB, 4 GB. 2 GB will be in ASCII, 4 GB will have skewed keys. We will start by supporting small evaluation, and expand to medium, large through the assignment.

The main event of this assignment is ***external sorting*** that sorts large data with limited memory size.

Your solution will be evaluated for correctness and execution time.

3 Test Protocol

Our testing infrastructure will evaluate submission based on your latest gitlab commit. Your code should have ***Makefile*** to compile the submitted code, and then running a series of tests. Your executive file name must be $'$***run***$'$. Executive file should get two parameters, ***inputfilepath*** and ***outputfilepath***.

./run inputfilepath outputfilepath

Each test has its own time limit, for all sizes of the data - small: 10s, medium: 20s, large: 100s - . If the submission exceeds that time, we will stop the test as a failure. These might be relaxed if they prove too constricting .

4 Assignment Score

Scoring of this assignment will be determined by your ranking. Rank will be decided by running time and correctness of your submission. Your last commit should be your best version because final validation after assignment deadline will proceed with your last commit.

5 TEST ENVIRONMENT

Submissions will be tested in a server with characteristics shown in the following environment:

Processor	Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz
Configuration	20 Cores / 40 Hyperthreads
Main Memory	2 GiB / no swap space
Storage	Samsumg ssd 850 pro 1TB

6 SUBMISSION

Assignment must be submitted to GitLab, and the ***project*1** directory should be created to submit the tasks. Makefile must be created within project1 directory. The executable file name is $'\boldsymbol{run}'$. It must locate at the same path with Makefile.

project1/Makefile

project1/run