# COMP 6751 Project Report

Group 4
40054328 Ming Ma
40079681 Tiancheng Xu
40054463 Daniel Zakeri

## Project introduction

The project can sort a large input file of positive integers with the restrict of limited memory. The function of the program is to compare the tuples in the input, and to output a file with the same tuples sorted in ascending order.

## Modules

The program uses Two-Phase Multiway Merge-Sort algorithm to do sorting. The progress can be divided into two phases.
At the beginning of the program, you can setup the memory limitation as run arguments. The program will get the file name and memory information at his time.

In the first phase, we have a method phaseOne() which takes the number of available integers and scanner to read the file. We sort the integers in array with the method Array.sort(). In the end of phase one, it may output multiple subfiles.

In the second phase, we first calculate the best combination of number of buffers with the buffer size. In the class MPMMS, we do this calculation and generate respectively input buffers. After, the method phaseTwo() read tuples into input buffers and pick the smallest integer to output buffer. As soon as the output buffer cannot write more integers, it writes to the file.

## Test cases

We test the project with 1,000,000 tuples and 500,000 tuples. And main memory restriction be 5MB and 10MB respectively.

1,000,000 tuples

```
sampleSize = 1000000 samples
freeMemory = 3860 KB
numOfIntInMemory = 494150 integers

PHASE 1 START
PHASE 1 END (2986 ms)

PHASE 2 START
Free memory = 4053 KB
numOfIntInMemory = 518907
numOfInputBuffers = 3
numOfIntInAInputBuffer = 109658 integers
numOfIntInOutputBuffer = 189933 integers

        PHASE 2 STAGE 0 START
        PHASE 2 STAGE 0 END (3020 ms)
PHASE 2 END (3041 ms)
```

500,000 tuples

```
sampleSize = 500000 samples
freeMemory = 7883 KB
numOfIntInMemory = 1009092 integers

PHASE 1 START
PHASE 1 END (660 ms)

PHASE 2 START
PHASE 2 END (0 ms)
```