CT421 Assignment 2

Leo Chui (20343266), Aoife Mulligan (20307646)

GitHub Repo

Proof of Running

A batch script has been written to aid in the testing of this assignment, and below is a sample output after running the script. The script is included in the ZIP submission.

```
Brute force: 518ms
MapRedue - Mapping phase: 771ms
MapRedue - Grouping phase: 204ms
MapRedue - Reducing phase: 126ms
MapReduce: 1107ms
Distributed MapRedue - Mapping phase: 412ms
Distributed MapRedue - Grouping phase: 210ms
Distributed MapRedue - Reducing phase: 6187ms
Distributed MapReduce: 6811ms
Thread Pool MapRedue - Mapping phase: 130ms
Thread Pool MapRedue - Grouping phase: 138ms
Thread Pool MapRedue - Reducing phase: 211ms
Thread Pool MapReduce: 482ms
```

All results are included in Appendices.

Testing Approach

Testing included running the application with different pool sizes for both map and reduce phases.

Below figure demonstrates the batch script being executed, with all outputs appended to respective output logs.

```
\lchui\ct414-assignment> .\run.bat 10 10
\lchui\ct414-assignment> .\run.bat 5 5
\lchui\ct414-assignment> .\run.bat 5 10
\lchui\ct414-assignment> .\run.bat 10 5
\lchui\ct414-assignment> .\run.bat 3 3
\lchui\ct414-assignment> .\run.bat 5 5
\lchui\ct414-assignment> .\run.bat 10 10
\lchui\ct414-assignment> .\run.bat 1 1
\lchui\ct414-assignment> .\run.bat 15 15
\lchui\ct414-assignment> .\run.bat 50 50
```

Results

The expectations for the results are that higher thread count decreases the duration taken, and vice versa, until thread count reaches 1 per file. We can see that is indeed the case when

looking at the results.

When comparing the last approach's results with previous approaches using results obtained from testing, we can see that all runs of approach 4 yielded better results than the other approaches, with the exception of 1 thread for both pools.

Results differ, even between the same approaches, but generally this is the case.

In terms of best performance, it might be beneficial to have at least half the amount of threads as there are files. The best performance out of the entire testing process is yielded by 15 threads for both phases, however the difference over 1 thread per file is not significant enough, thus can be attributed to simple deviations between runs. Therefore, we believe it is reasonable to say that best performance on average can be yielded by having 1 thread per file.

Appendices

1_1

Brute force: 356ms
MapRedue - Mapping phase: 457ms
MapRedue - Grouping phase: 140ms
MapRedue - Reducing phase: 176ms
MapReduce: 774ms
Distributed MapRedue - Mapping phase: 205ms
Distributed MapRedue - Grouping phase: 169ms
Distributed MapRedue - Reducing phase: 3988ms
Distributed MapReduce: 4363ms
Thread Pool MapRedue - Mapping phase: 158ms
Thread Pool MapRedue - Grouping phase: 84ms
Thread Pool MapRedue - Reducing phase: 216ms
Thread Pool MapReduce: 460ms
Thread Pool MapReduce - Map Pool Size: 1 - Reduce Pool Size: 1

3_3

Brute force: 369ms
MapRedue - Mapping phase: 514ms
MapRedue - Grouping phase: 138ms
MapRedue - Reducing phase: 112ms
MapReduce: 765ms
Distributed MapRedue - Mapping phase: 99ms
Distributed MapRedue - Grouping phase: 91ms
Distributed MapRedue - Reducing phase: 4155ms
Distributed MapReduce: 4348ms
Thread Pool MapRedue - Mapping phase: 88ms
Thread Pool MapRedue - Grouping phase: 107ms
Thread Pool MapRedue - Reducing phase: 160ms
Thread Pool MapReduce: 356ms
Thread Pool MapReduce - Map Pool Size: 3 - Reduce Pool Size: 3

5_5

Brute force: 360ms
MapRedue - Mapping phase: 467ms

MapRedue - Grouping phase: 143ms
MapRedue - Reducing phase: 107ms
MapReduce: 717ms
Distributed MapRedue - Mapping phase: 220ms
Distributed MapRedue - Grouping phase: 207ms
Distributed MapRedue - Reducing phase: 3823ms
Distributed MapReduce: 4250ms
Thread Pool MapRedue - Mapping phase: 70ms
Thread Pool MapRedue - Grouping phase: 90ms
Thread Pool MapRedue - Reducing phase: 116ms
Thread Pool MapReduce: 276ms
Thread Pool MapReduce - Map Pool Size: 5 - Reduce Pool Size: 5

5_10

Brute force: 341ms
MapRedue - Mapping phase: 515ms
MapRedue - Grouping phase: 99ms
MapRedue - Reducing phase: 145ms
MapReduce: 760ms
Distributed MapRedue - Mapping phase: 159ms
Distributed MapRedue - Grouping phase: 304ms
Distributed MapRedue - Reducing phase: 4042ms
Distributed MapReduce: 4505ms
Thread Pool MapRedue - Mapping phase: 68ms
Thread Pool MapRedue - Grouping phase: 90ms
Thread Pool MapRedue - Reducing phase: 115ms
Thread Pool MapReduce: 273ms
Thread Pool MapReduce - Map Pool Size: 5 - Reduce Pool Size: 10

10_5

Brute force: 339ms
MapRedue - Mapping phase: 474ms
MapRedue - Grouping phase: 160ms
MapRedue - Reducing phase: 111ms
MapReduce: 746ms
Distributed MapRedue - Mapping phase: 174ms
Distributed MapRedue - Grouping phase: 253ms
Distributed MapRedue - Reducing phase: 3845ms

Distributed MapReduce: 4273ms
Thread Pool MapRedue - Mapping phase: 66ms
Thread Pool MapRedue - Grouping phase: 88ms
Thread Pool MapRedue - Reducing phase: 120ms
Thread Pool MapReduce: 275ms
Thread Pool MapReduce - Map Pool Size: 10 - Reduce Pool Size: 5

10_10

Brute force: 354ms
MapRedue - Mapping phase: 513ms
MapRedue - Grouping phase: 127ms
MapRedue - Reducing phase: 190ms
MapReduce: 830ms
Distributed MapRedue - Mapping phase: 230ms
Distributed MapRedue - Grouping phase: 104ms
Distributed MapRedue - Reducing phase: 3906ms
Distributed MapReduce: 4240ms
Thread Pool MapRedue - Mapping phase: 63ms
Thread Pool MapRedue - Grouping phase: 87ms
Thread Pool MapRedue - Reducing phase: 135ms
Thread Pool MapReduce: 285ms
Thread Pool MapReduce - Map Pool Size: 10 - Reduce Pool Size: 10

15_15

Brute force: 383ms
MapRedue - Mapping phase: 467ms
MapRedue - Grouping phase: 164ms
MapRedue - Reducing phase: 187ms
MapReduce: 818ms
Distributed MapRedue - Mapping phase: 156ms
Distributed MapRedue - Grouping phase: 87ms
Distributed MapRedue - Reducing phase: 4013ms
Distributed MapReduce: 4258ms
Thread Pool MapRedue - Mapping phase: 72ms
Thread Pool MapRedue - Grouping phase: 78ms
Thread Pool MapRedue - Reducing phase: 118ms
Thread Pool MapReduce: 268ms
Thread Pool MapReduce - Map Pool Size: 15 - Reduce Pool Size: 15

50_50

Brute force: 401ms
MapRedue - Mapping phase: 477ms
MapRedue - Grouping phase: 153ms
MapRedue - Reducing phase: 180ms
MapReduce: 810ms
Distributed MapRedue - Mapping phase: 225ms
Distributed MapRedue - Grouping phase: 241ms
Distributed MapRedue - Reducing phase: 3898ms
Distributed MapReduce: 4364ms
Thread Pool MapRedue - Mapping phase: 59ms
Thread Pool MapRedue - Grouping phase: 93ms
Thread Pool MapRedue - Reducing phase: 143ms
Thread Pool MapReduce: 295ms
Thread Pool MapReduce - Map Pool Size: 50 - Reduce Pool Size: 50