

Program Structures and Algorithms

Fall 2021

Assignment No. 5 – Parallel Sorting

- Tasks in the assignment:

Modify merge sort so that each partition of the array is sorted in parallel and then merged. Parallelizing the recursive calls until the subsequences are sorted and merging them back. CompletableFuture implementation along with ForkJoinPool is used to achieve this parallelism. Several parameters were modified here

Part 1: Deciding a cut-off value

1. Modify the cutoff values for sufficiently large arrays while parallelizing sort.
2. Switch to system sort when the cutoff value is reached.
3. Observe the change in sorting time.

Part 2: Deciding recursion depth

1. Recursion depth is the number of available threads. These threads sort the array partitions in parallel.
2. Decide on an ideal number of threads to improve performance.
3. Prevent partition after depth of $\log n$ is reached where n is the separate number of threads.

Part 3: Combination of the two

1. Modify the number of threads using the ForkJoinPool.
2. Check variation in running time for various array sizes and cutoff values.
3. Here I have modified the number of threads along with array size and cutoff values to achieve a combination of these parallelization schemes.

- Conclusion

Parallelizing allows load balancing and proper use of the multiple processors decreasing the running time. We also need to ensure the recursion depth to be used to avoid the out of memory exception. However the cutoff value plays a major role in deciding an efficient parallel sort. Irrespective of the degree of parallelism if the cutoff is low the running time increases. The cut-off number affects the sorting time. The cut-off value determines when to switch to the system sort. Thus a higher cutoff values means more contribution of the time taken by system sort. Hence its important to determine the correct cut-off value to achieve a best case scenario of using merge sort and the system sort.

- Evidence to support conclusions

1. Snapshot of result of experiments

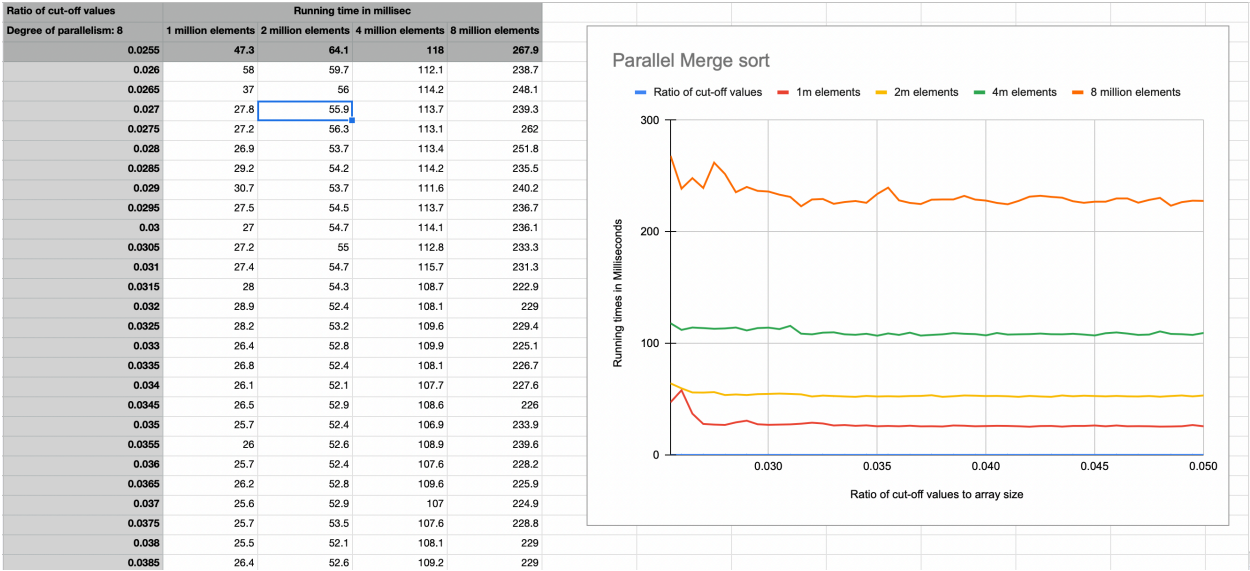
The screenshot shows an IDE window titled "INFO6205-Assignments" with a project named "src". The file "result_arraySize_1000000.csv" is open. The output shows the results of an experiment with a degree of parallelism of 2 and an array size of 1,000,000. The output is a list of cutoff values and their corresponding times, showing a general trend of increasing time as the cutoff value increases.

```
Run: Main
/Library/Java/JavaVirtualMachines/jdk-16.0.2.jdk/Contents/Home/bin/java ...
Degree of parallelism: 2, Arraysize: 1000000
cutoff: 25500 10times Time:724ms
cutoff: 26000 10times Time:426ms
cutoff: 26500 10times Time:342ms
cutoff: 27000 10times Time:331ms
cutoff: 27500 10times Time:297ms
cutoff: 28000 10times Time:287ms
cutoff: 28500 10times Time:288ms
cutoff: 29000 10times Time:308ms
cutoff: 29500 10times Time:293ms
cutoff: 30000 10times Time:298ms
cutoff: 30500 10times Time:287ms
cutoff: 31000 10times Time:288ms
cutoff: 31500 10times Time:298ms
cutoff: 32000 10times Time:298ms
cutoff: 32500 10times Time:296ms
cutoff: 33000 10times Time:295ms
cutoff: 33500 10times Time:298ms
cutoff: 34000 10times Time:293ms
cutoff: 34500 10times Time:298ms
cutoff: 35000 10times Time:298ms
cutoff: 35500 10times Time:289ms
cutoff: 36000 10times Time:289ms
cutoff: 36500 10times Time:289ms
cutoff: 37000 10times Time:298ms
cutoff: 37500 10times Time:291ms
cutoff: 38000 10times Time:295ms
cutoff: 38500 10times Time:291ms
cutoff: 39000 10times Time:298ms
cutoff: 39500 10times Time:291ms
cutoff: 40000 10times Time:298ms
cutoff: 40500 10times Time:298ms
cutoff: 41000 10times Time:291ms
cutoff: 41500 10times Time:287ms
```

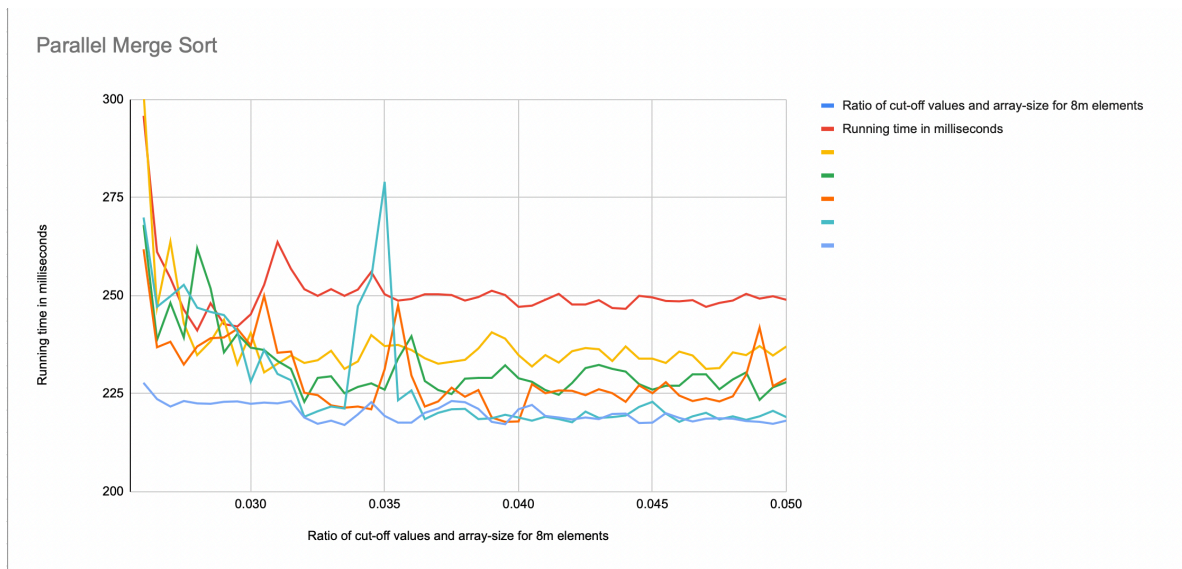
The screenshot shows an IDE window titled "INFO6205-Assignments" with a project named "src". The file "result_arraySize_1000000.csv" is open. The output shows the results of an experiment with a degree of parallelism of 64 and an array size of 1,000,000. The output is a list of cutoff values and their corresponding times, showing a general trend of increasing time as the cutoff value increases.

```
Run: Main
cutoff: 190000 10times Time:1074ms
cutoff: 200000 10times Time:1068ms
Degree of parallelism: 64, Arraysize: 8000000
cutoff: 204000 10times Time:2277ms
cutoff: 208000 10times Time:2236ms
cutoff: 212000 10times Time:2217ms
cutoff: 216000 10times Time:2231ms
cutoff: 220000 10times Time:2225ms
cutoff: 224000 10times Time:2224ms
cutoff: 228000 10times Time:2229ms
cutoff: 232000 10times Time:2230ms
cutoff: 236000 10times Time:2224ms
cutoff: 240000 10times Time:2227ms
cutoff: 244000 10times Time:2225ms
cutoff: 248000 10times Time:2231ms
cutoff: 252000 10times Time:2189ms
cutoff: 256000 10times Time:2173ms
cutoff: 260000 10times Time:2181ms
cutoff: 264000 10times Time:2170ms
cutoff: 268000 10times Time:2197ms
cutoff: 272000 10times Time:2228ms
cutoff: 276000 10times Time:2193ms
cutoff: 280000 10times Time:2176ms
cutoff: 284000 10times Time:2176ms
cutoff: 288000 10times Time:2281ms
cutoff: 292000 10times Time:2212ms
cutoff: 296000 10times Time:2231ms
cutoff: 300000 10times Time:2228ms
cutoff: 304000 10times Time:2211ms
cutoff: 308000 10times Time:2178ms
cutoff: 312000 10times Time:2172ms
cutoff: 316000 10times Time:2210ms
cutoff: 320000 10times Time:2221ms
cutoff: 324000 10times Time:2193ms
cutoff: 328000 10times Time:2189ms
```

Observations:



Ratio of cut-off values and array-size for 8m elements	Running time in milliseconds						
	Degree of parallelism: 2	Degree of parallelism: 4	Degree of parallelism: 8	Degree of parallelism: 16	Degree of parallelism: 32	Degree of parallelism: 64	
0.0255							
0.026	295.7	301.6	267.9	261.7	269.8	227.7	
0.0265	261.1	247	238.7	236.8	247.1	223.6	
0.027	254.4	263.7	248.1	238.2	249.8	221.7	
0.0275	246.4	242.9	239.3	232.4	252.7	223.1	
0.028	241.1	234.8	262	237	246.9	222.5	
0.0285	248	238.3	251.8	239.1	245.8	222.4	
0.029	242.7	243.8	235.5	239.3	245	222.9	
0.0295	242.1	232.5	240.2	241.5	240.8	223	
0.03	245.2	240.5	236.7	237.3	228	222.4	
0.0305	252.7	230.4	236.1	249.9	236.1	222.7	
0.031	263.6	232.6	233.3	235.4	230	222.5	
0.0315	256.8	234.7	231.3	235.7	228.4	223.1	
0.032	251.6	232.8	222.9	225.2	219.1	218.9	
0.0325	249.9	233.5	229	224.6	220.5	217.3	
0.033	251.6	235.9	229.4	222	221.7	218.1	
0.0335	249.9	231.3	225.1	221.4	221.2	217	
0.034	251.5	233.2	226.7	221.7	247.3	219.7	
0.0345	256	239.9	227.6	221	254.4	222.8	
0.035	250.3	237.1	226	231.2	278.9	219.3	
0.0355	248.7	237.4	233.9	247.5	223.3	217.6	
0.036	249.1	236.1	239.6	229.6	225.8	217.6	
0.0365	250.3	234	228.2	221.7	218.5	220.1	
0.037	250.3	232.6	225.9	223	220.1	221.2	
0.0375	250.1	233.1	224.9	226.5	221	223.1	
0.038	248.7	233.6	228.8	224.2	221.1	222.8	



The cutoff values when increased with the number of elements give a similar pattern of running times. Increase in parallelization does not have a very significant improvement in running time.