Program Structure and Algorithms (INFO 6205)

Fall – 2021 Final Project

Report

Keyur Ashokbhai Barot – 001568664

I. Tasks:

Following tasks were performed in the project:

1. Implement MSD radix sort for a natural language which uses Unicode characters.
2. Sort Simplified Chinese words and compare the results of MSD radix with Timsort, Dual Pivot Quicksort, Huskysort and LSD radix sort.

II. Results:

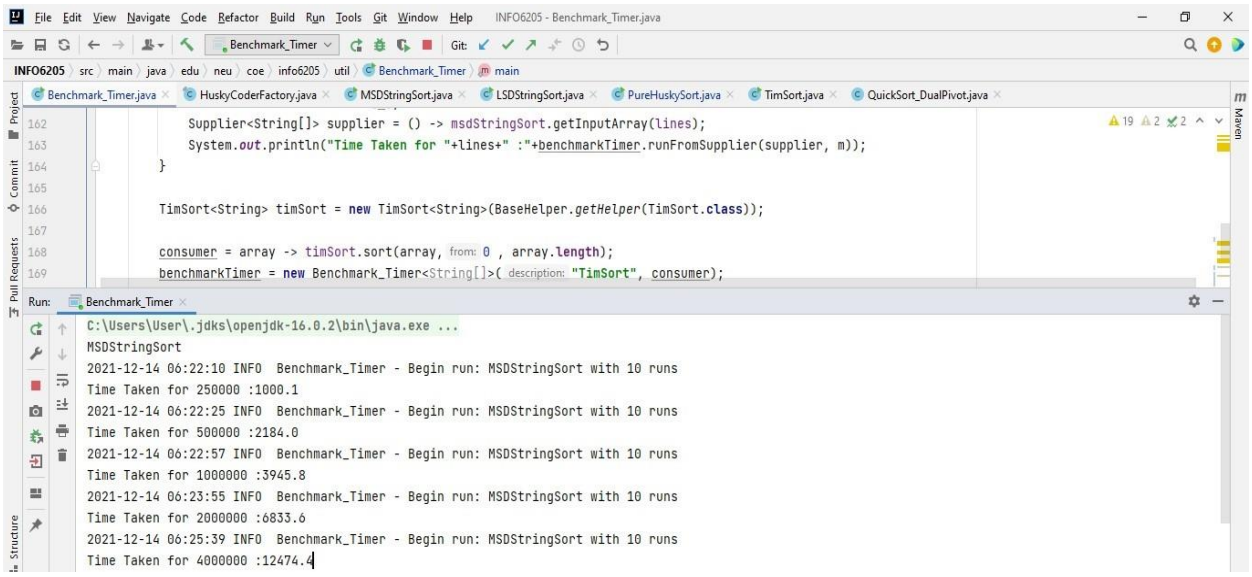
The code for the benchmarking can be found here:

https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/util/Benchmark_Timer.java

1. MSD Radix Sort:

https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/counting/MSDStringSort.java

The benchmarking results for the MSD radix sort are as follows:

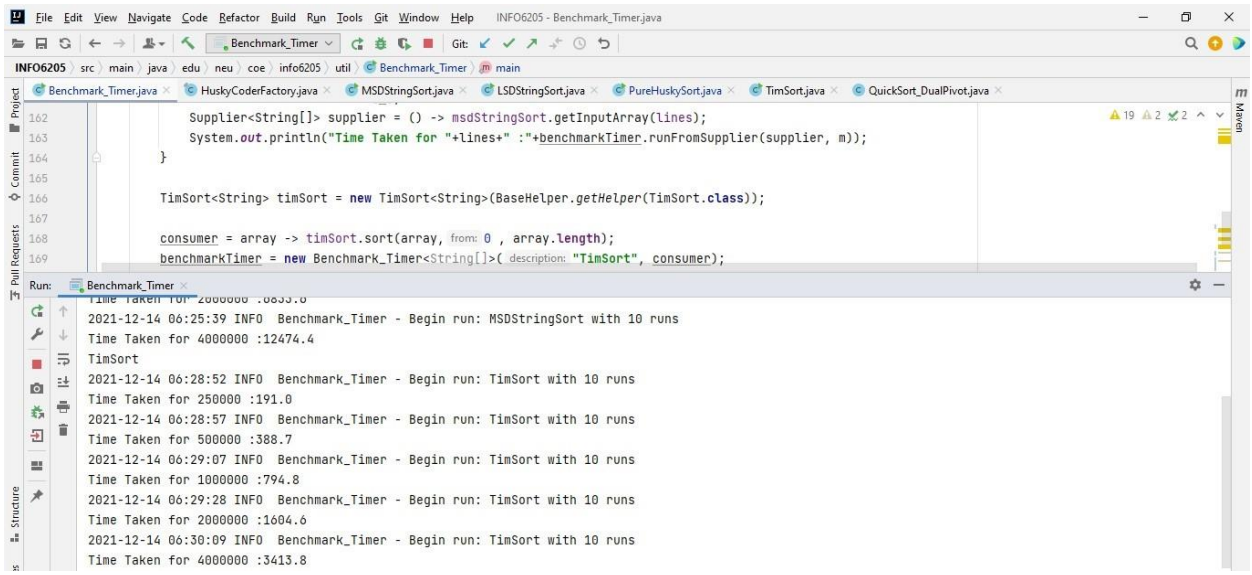


Algorithm	Number of words	Time(ms)
MSDStringSort	250000	1000.1
MSDStringSort	500000	2184
MSDStringSort	1000000	3945.8
MSDStringSort	2000000	6833.6
MSDStringSort	4000000	12474.4

2. Timsort:

https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/linearithmic/TimSort.java

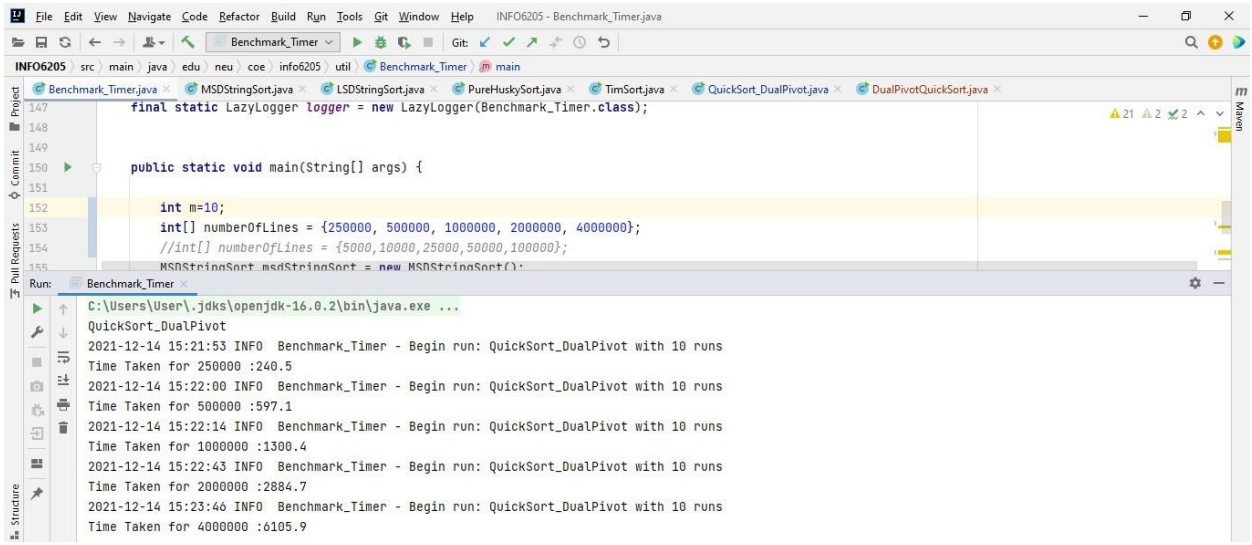
The benchmarking results for Timsort are as follows:



Algorithm	Number of words	Time(ms)
TimSort	250000	191
TimSort	500000	388.7
TimSort	1000000	794.8
TimSort	2000000	1604.6
TimSort	4000000	3413.8

3. **Dual-pivot Quicksort:**
   https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/DualPivotQuickSort.java

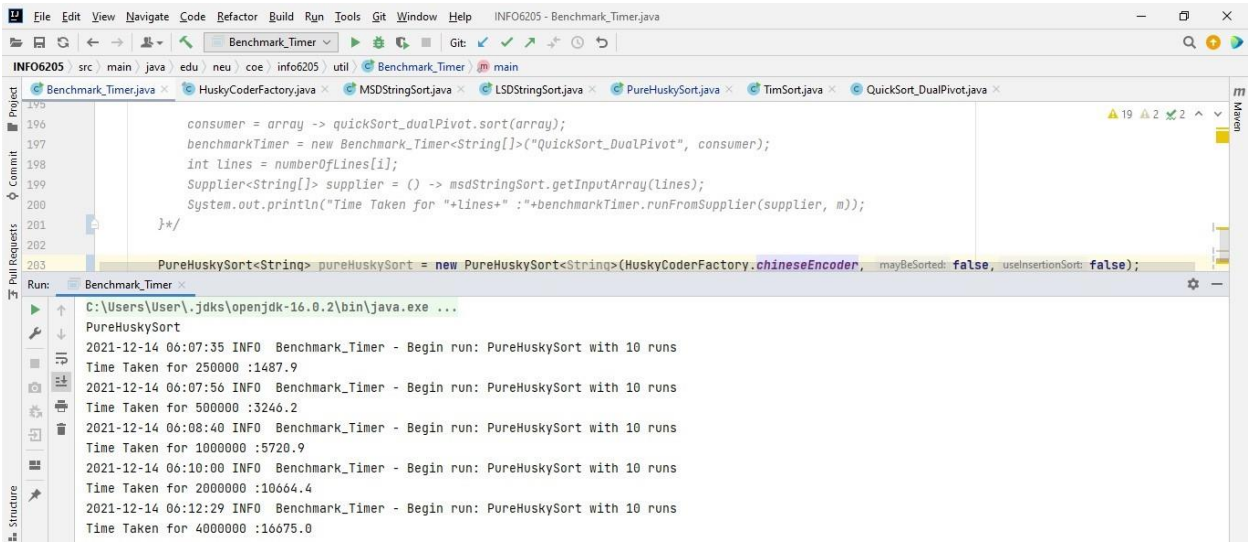
The benchmarking results for Dual-pivot Quicksort are as follows:



Algorithm	Number of words	Time(ms)
QuickSort_DualPivot	250000	176.6
QuickSort_DualPivot	500000	446
QuickSort_DualPivot	1000000	1069.7
QuickSort_DualPivot	2000000	2302.9
QuickSort_DualPivot	4000000	4971.9

4. **Huskysort:**
   https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/huskySort/PureHuskySort.java

The benchmarking results for Huskysort are as follows:

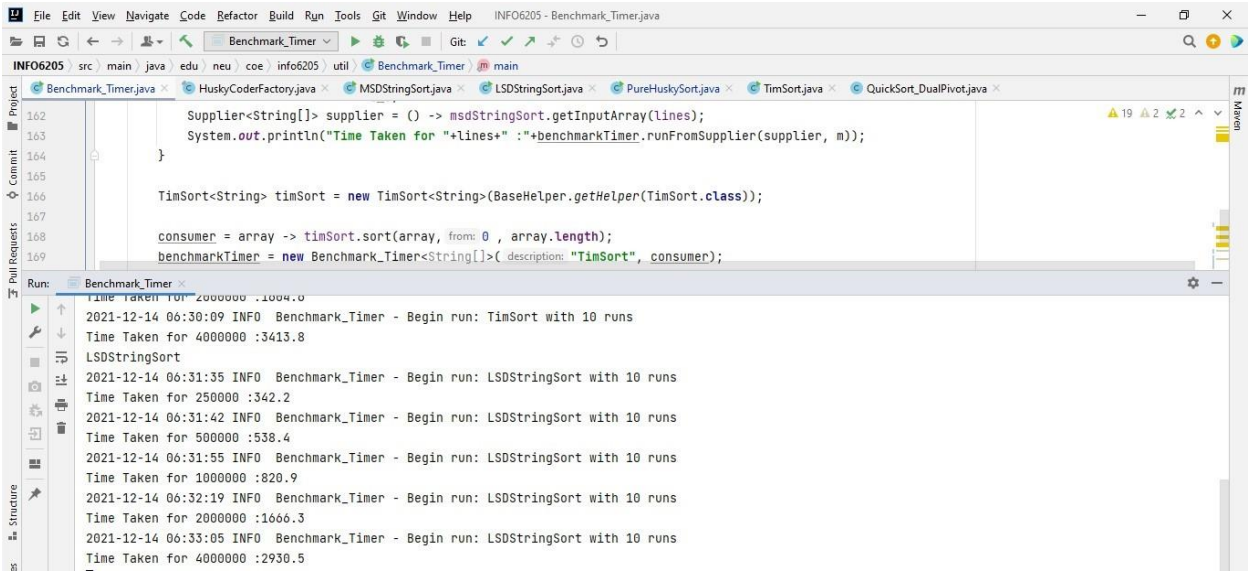


Algorithm	Number of words	Time(ms)
PureHuskySort	250000	623.2
PureHuskySort	500000	1285
PureHuskySort	1000000	2589.8
PureHuskySort	2000000	5223.2
PureHuskySort	4000000	10506.6

5. LSD radix sort:

https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/counting/LSDStringSort.java

The benchmarking results of LSD radix sort are as follows:



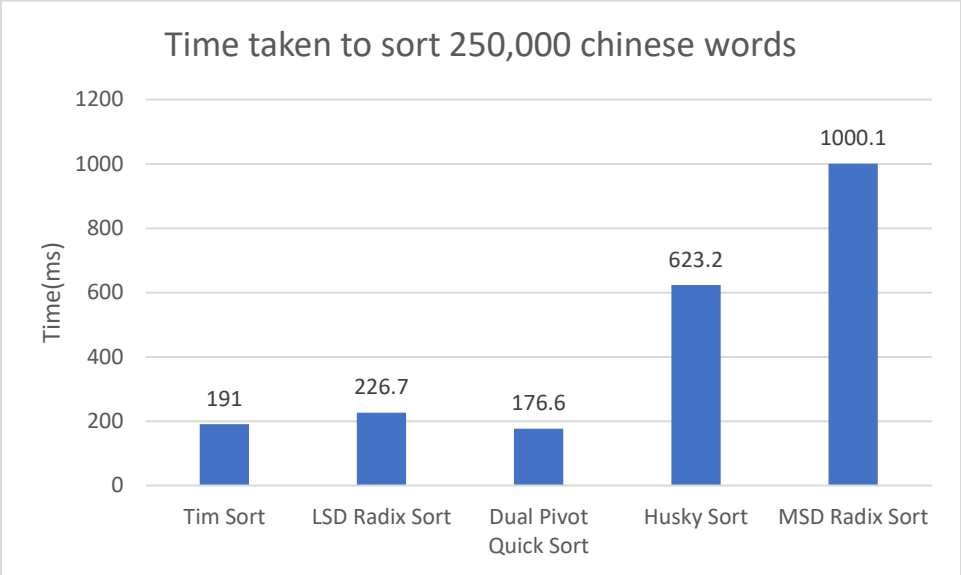
Algorithm	Number of words	Time(ms)
LSDStringSort	250000	226.7
LSDStringSort	500000	539.8
LSDStringSort	1000000	848.5
LSDStringSort	2000000	1503.7
LSDStringSort	4000000	2977.3

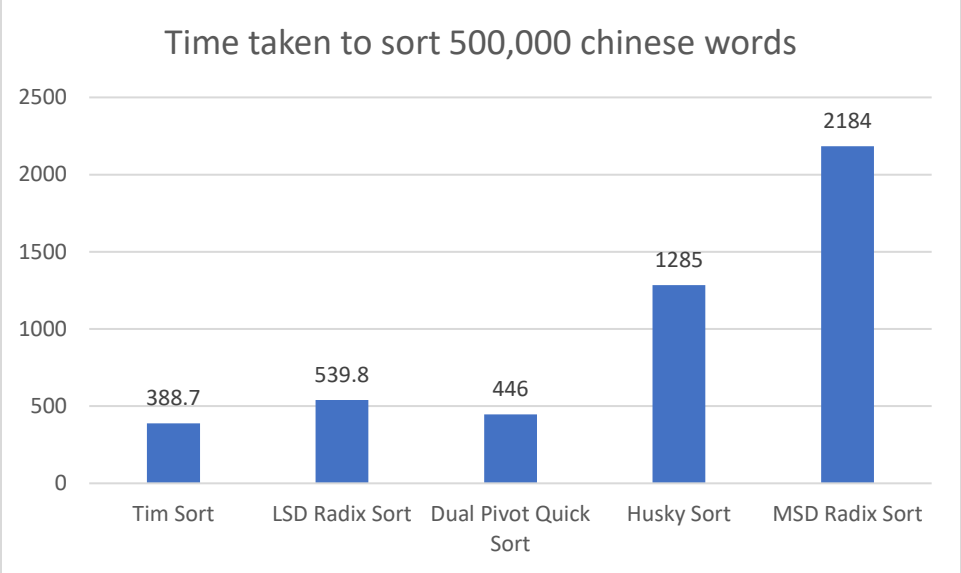
III. Conclusion:

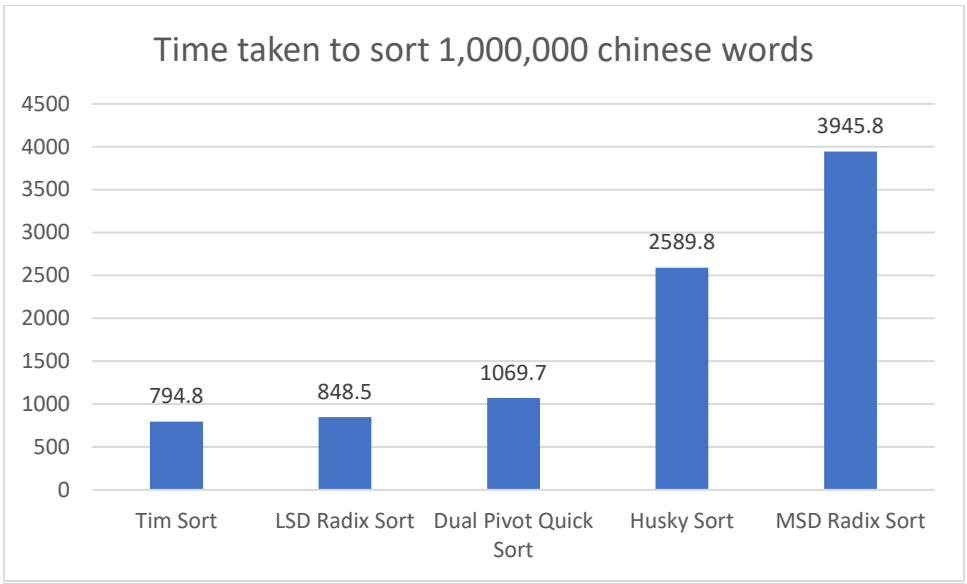
After implementing all the five given algorithms we can say that Timsort takes the least time to sort, followed by LSD radix sort, which is followed by Dual-Pivot Quicksort, Huskysort and MSD radix sort respectively.

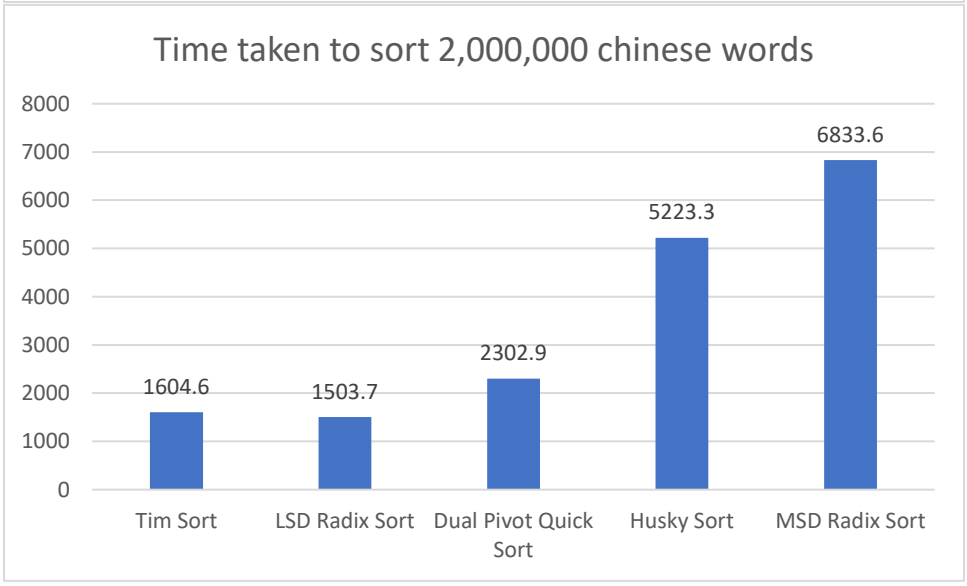
Evidence for the same can be seen in the following time comparisons of each algorithm.

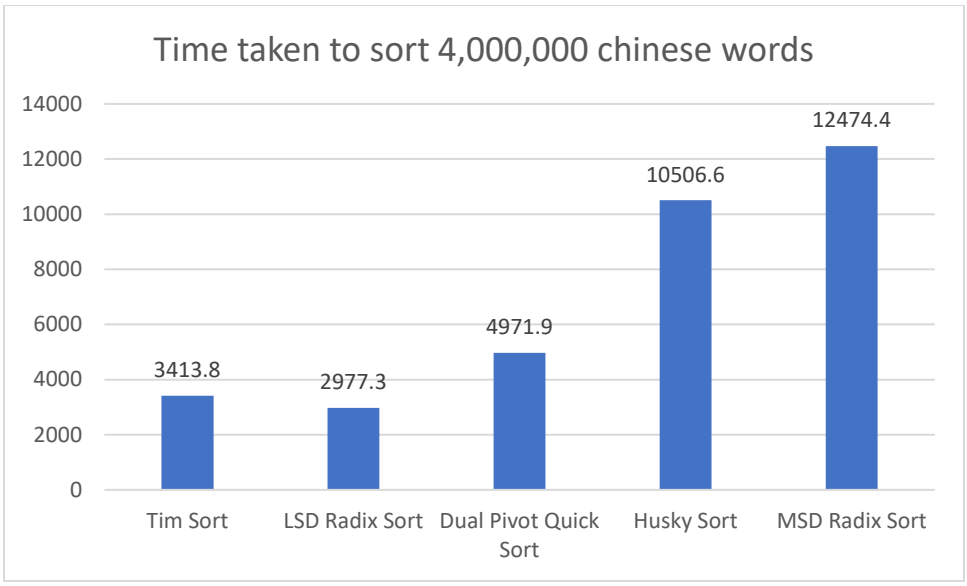
IV. Evidence:


Time taken to sort 250,000 chinese words
Time(ms)
1200
1000
800
600
400
200
0
191
226.7
176.6
623.2
1000.1
Tim Sort
LSD Radix Sort
Dual Pivot Quick Sort
Husky Sort
MSD Radix Sort


Time taken to sort 500,000 chinese words
2500
2000
1500
1000
500
0
388.7
539.8
446
1285
2184
Tim Sort
LSD Radix Sort
Dual Pivot Quick Sort
Husky Sort
MSD Radix Sort


Time taken to sort 1,000,000 chinese words
4500
4000
3500
3000
2500
2000
1500
1000
500
0
794.8
848.5
1069.7
2589.8
3945.8
Tim Sort
LSD Radix Sort
Dual Pivot Quick Sort
Husky Sort
MSD Radix Sort


Time taken to sort 2,000,000 chinese words
8000
7000
6000
5000
4000
3000
2000
1000
0
1604.6
1503.7
2302.9
5223.3
6833.6
Tim Sort
LSD Radix Sort
Dual Pivot Quick Sort
Husky Sort
MSD Radix Sort


Time taken to sort 4,000,000 chinese words
14000
12000
10000
8000
6000
4000
2000
0
3413.8
2977.3
4971.9
10506.6
12474.4
Tim Sort
LSD Radix Sort
Dual Pivot Quick Sort
Husky Sort
MSD Radix Sort