Introduction:

This study evaluates the performance of four distinct sorting algorithms -- Insertion Sort, Heap Sort, Quick Sort, and Merge Sort -- across seven different input array sizes. The objective is to empirically assess and compare the efficiency of each algorithm when handling large datasets of random integers.

Methods:

*Array Sizes:*

Seven different array sizes were selected for testing:

|  |
| --- |
| n = 300,000 |
| n = 400,000 |
| n = 500,000 |
| n = 600,000 |
| n = 750,000 |
| n = 900,000 |

(Minimum array sizes were chosen for runtimes >= 10ms)

*Data Population:*

For each specified array size, an empty array variable is created and every element was assigned a random integer value between X and Y using Java's Random class, which produces values with a uniform distribution across the entire array. The array generation process is as follows:

A computer screen with text

Description automatically generated

To ensure a consistent comparison, a copy of the original input array was created for each algorithm, guaranteeing that each sorting algorithm operated on identical and unsorted data.

A screen shot of a computer code

Description automatically generated

*Runtime Measurement:*

The measureRuntime(); method captured the start and stop times at the sorting function's invocation using Java's System.currentTimeMillis(); function. After sorting, each algorithm's output array was verified for correct ordering using the isSorted(); method, which simply iterates trough the array to ensure that it is arranged in increasing order.

A screen shot of a computer code

Description automatically generated

Validation function:

A computer screen with colorful text

Description automatically generated

*Execution and Output Logging:*

The runtime and validation results were printed to the console on each function invocation. For each array size, the program executed each sorting algorithm once.

A screen shot of a computer code

Description automatically generated

Results:

The main program ran 10 times to gather 10 data points for each algorithm in order to minimize unrelated artifacts.

All operations were performed on the same machine hardware and in the same execution environment to minimize external variables.

Discussion: