# PERFORMANCE EVALUATION

Hadoop:

The Single node hadoop cluster as mentioned runs on a single node. And because of that takes a lot of time when compared with multiple nodes. We can see that the time taken to run 10GB dataset is very high. If we run 100GB in single node it would take more time than 16 node cluster would take. The 16 nodes cluster the data is accessed by the master as well the slaves which process the data and sort the data that it takes and provides a faster result than single node.

|	Time in Seconds	Throughput(MB/Sec)
Single Node 10GB	1117	9.167
16 Node 100GB	10955	9.347

## DataSet execution time

Throughput
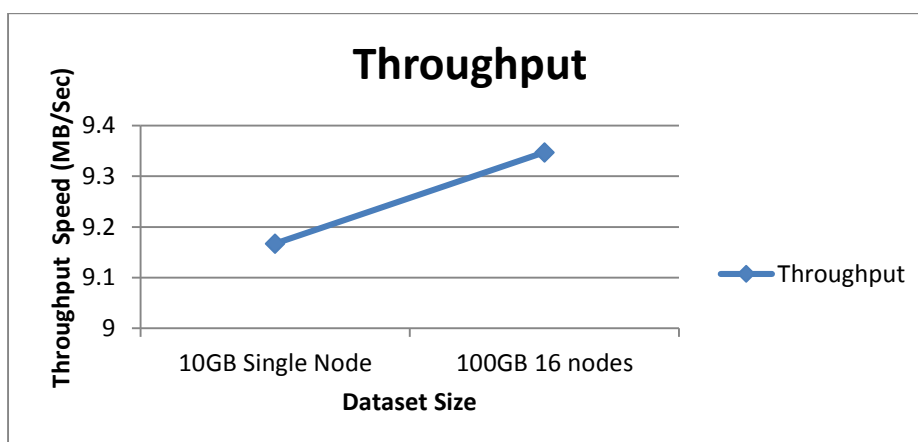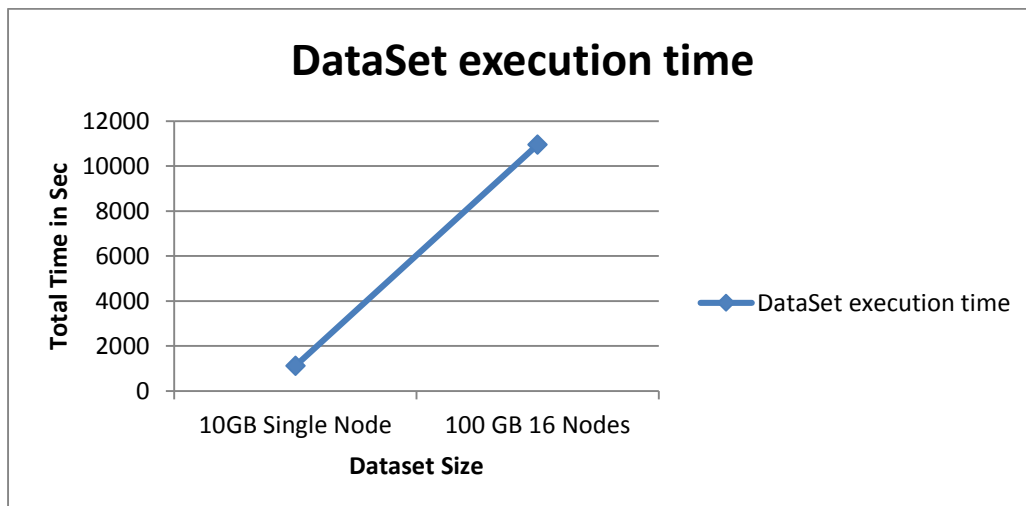
Spark:

In the spark framework the single node just like the hadoop framework takes more time than the 16 nodes cluster. The single node only has a one node which processes the data and sorts the dataset. The 16 nodes has master and slaves nodes which access chunks of the data and process those chunks of data ultimately sorting them and producing a sorted output. When compared to Hadoop spark is a little faster in execution and sorts the output at a quicker rate

The sort time obtained is 176.314 seconds and throughput is 58.078. The total time taken is 996.845 seconds and throughput is

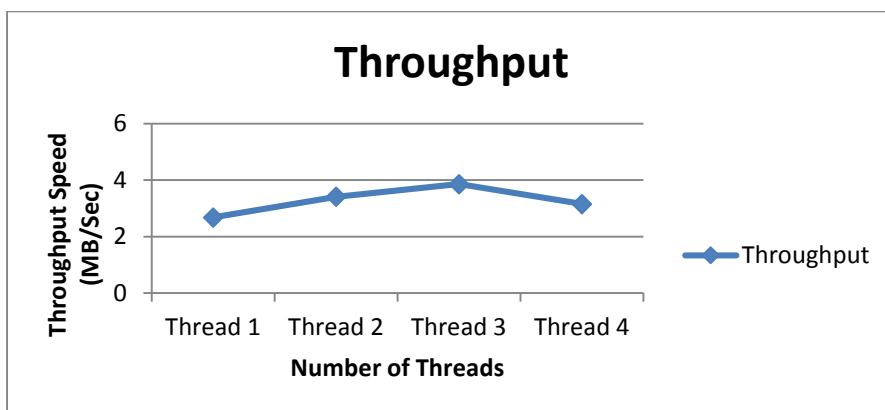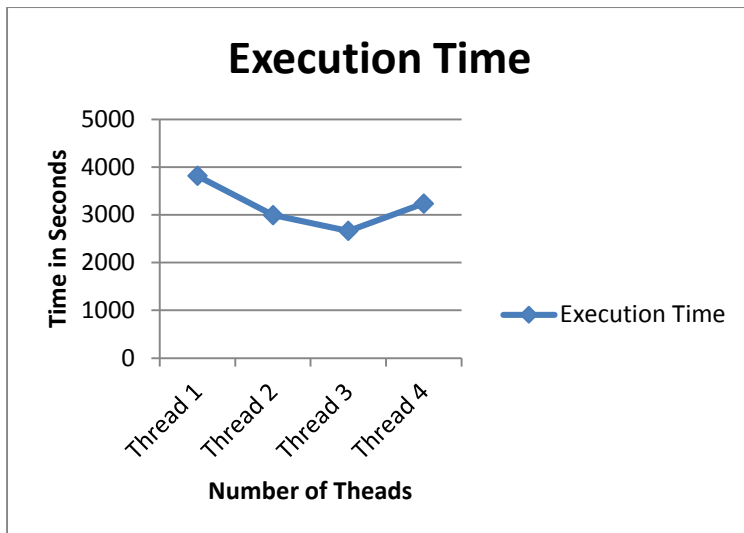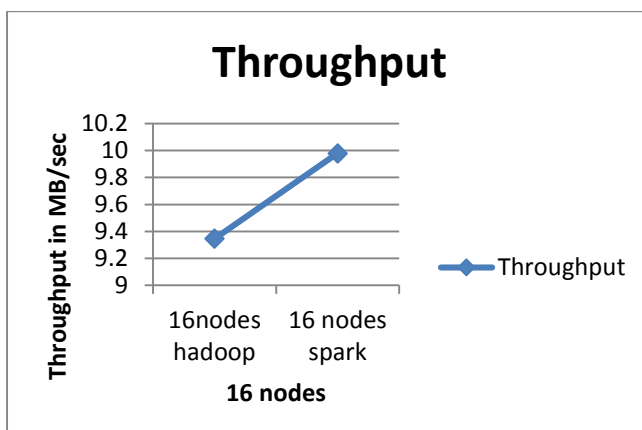|	Time in Seconds	Throughput(MB/Sec)
Single Node 10GB	1401	7.31
16 Node 100GB	10260	9.98

Throughput
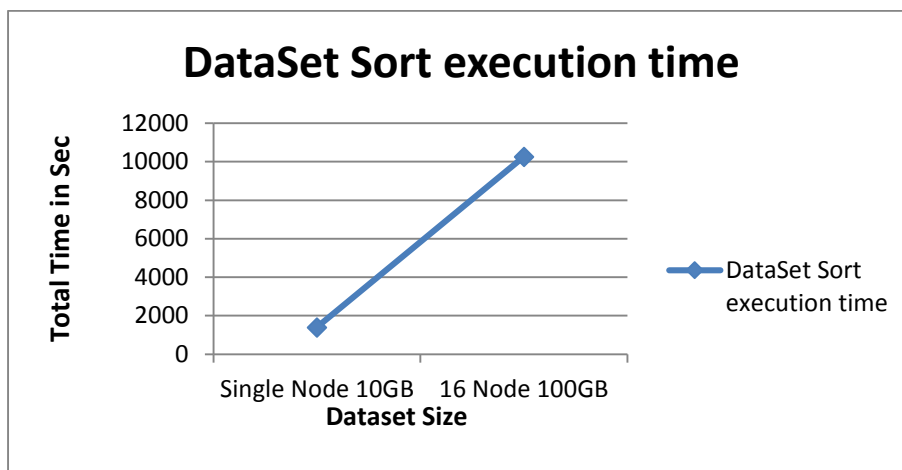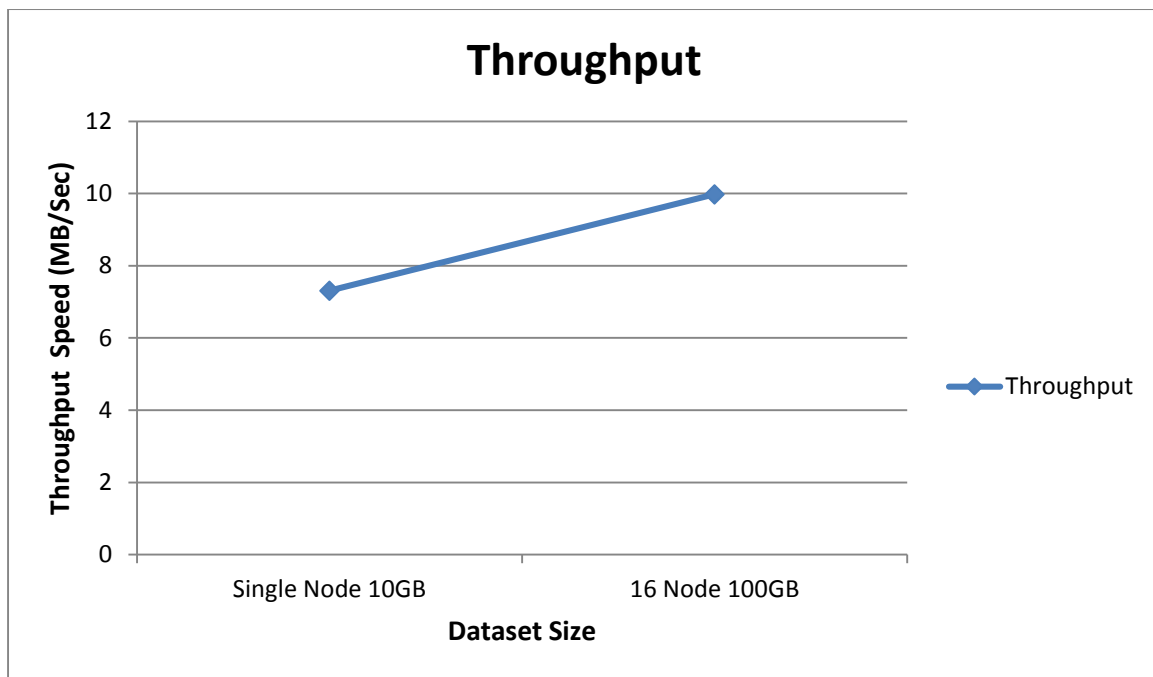
Shared Memory Java:

The shared memory java access the data in terms of 1,2,4,8 threads. It handles the data in such a way that for a single thread just processes the data by the single thread. For 2 threads the data is split into chunks which are handled by the two threads separately which sorts and then combines the data to produces the output. Like 2 threads, 4 and 8 threads follow a similar process.
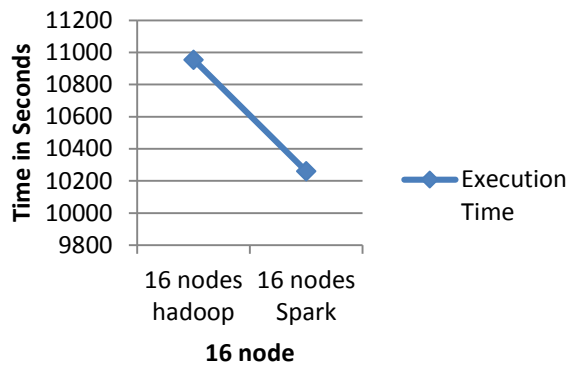
Note: This takes more time to sort than Hadoop and Spark for more number of nodes.
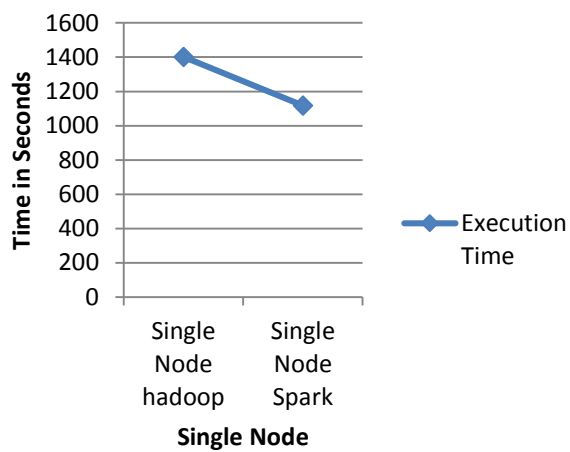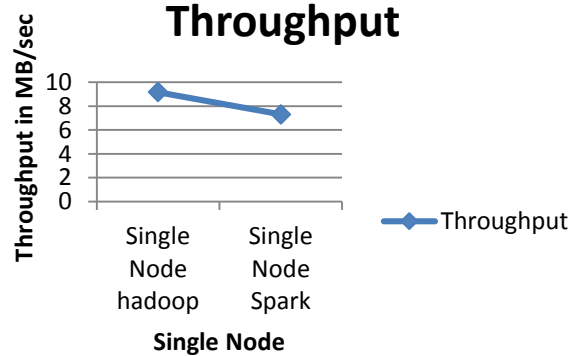
Execution Time

Throughput

Comparison:

Throughput

Execution Time

16 node chart: Time in Seconds, comparing 16 nodes hadoop (~10950) and 16 nodes Spark (~10260). Legend: Execution Time

Execution Time

Single Node chart: Time in Seconds, comparing Single Node hadoop (~1400) and Single Node Spark (~1120). Legend: Execution Time

Throughput

Single Node chart: Throughput in MB/sec, comparing Single Node hadoop (~9) and Single Node Spark (~7). Legend: Throughput

Here we can observe that for 16 nodes Spark performs faster sorting operation than Hadoop and as a result of that the throughput is higher than that hadoop and hadoop takes more time in sorting . As far as single node is concerned quicker performance is obtained in Hadoop than in spark hence getting a higher throughput.