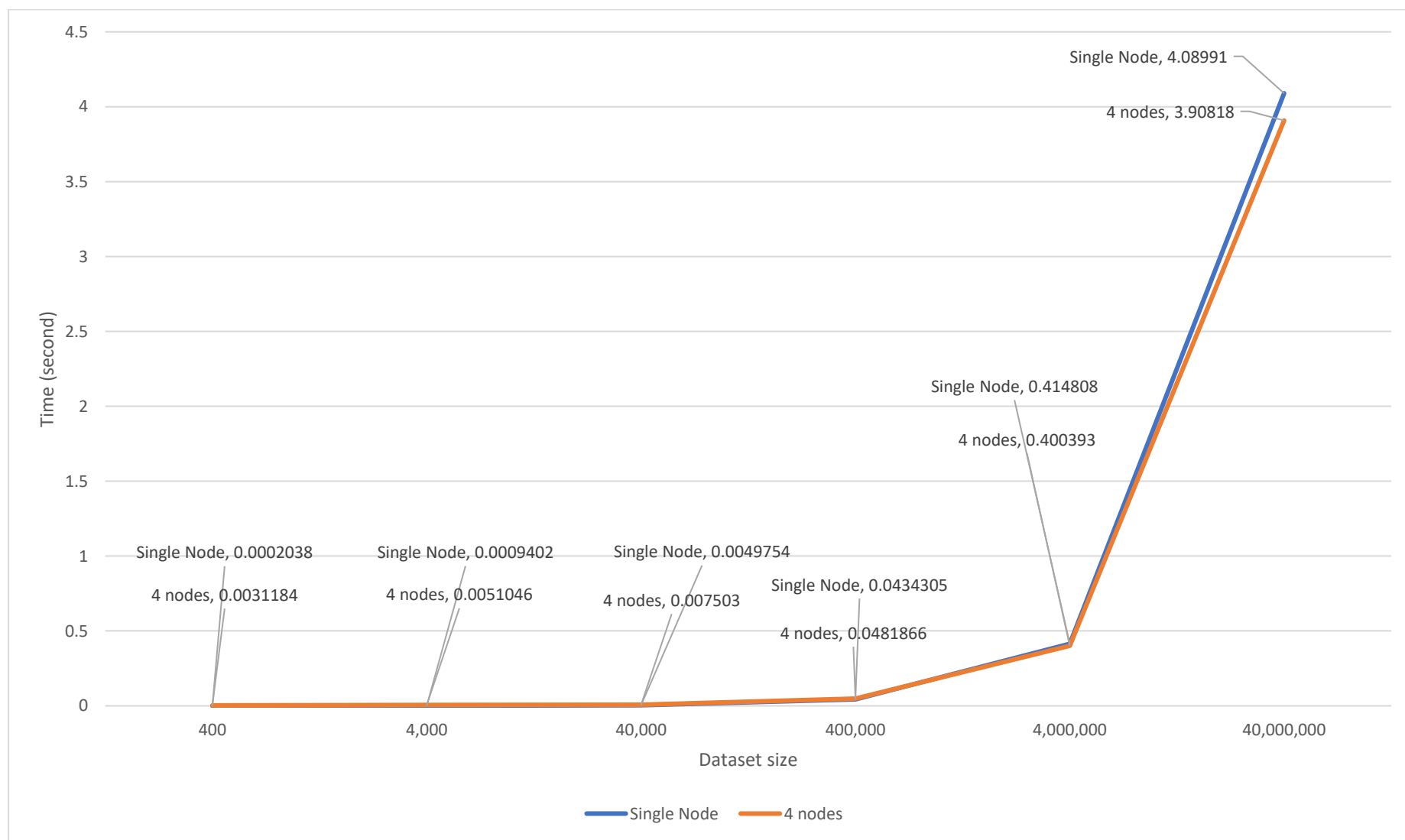


# **Report of assignment 1 part 2**

**Student Name : Yi How Tan**

**Student Number : 3011712**

**Course code : BSCH-DS**



Dataset size	400	4,000	40,000	400,000	4,000,000	40,000,000
Single node	0.000204s	0.00094s	0.004975s	0.043431s	0.414808s	4.08991s
4 nodes	0.003118s	0.005105s	0.007503s	0.048187s	0.400393s	3.90818s

From the graph above, compute standard deviation of dataset 400, 4000, 40000, 400000 by using single node is faster than 4 nodes, but from 4,000,000 onward, 4 nodes perform better than single node. In the conclusion, single node is doing well in small task but multiple nodes will perform well in big task.