



For the above graph:

M1 = 14000000

M2 = 13000000

M3 = 100000000

1. For the GPU version to be better than the sequential version, the size of the data set must be large enough.
2. M2 is lower than M1 because the branch divergence helps thread actions to be carried out faster with less amount of thread context switching happening
3. M3 is higher than M1, as it requires data set to be large enough to compensate for the cost of sharing the data set amongst all threads.
4. No optimizations were used besides branch divergence and shared memory usage, as having more optimizations may also lead to higher overhead costs, leading to slower performance time.