

Question 1:

Because CPU runs at 4 times speed of GPU and there are 10 CPU cores compared to 1000 GPUs, there should be at least 41 input data size because at 40 the GPU and CPU will have same time excluding the overhead because each core on CPU will have to run 4 times versus 1 time for GPU and then 41 would be needed then to make GPU run faster than CPU.

Question 2:

a. .

1. You would need $500 * 1024 * 2$ cycles without shared memory
2. You would need the number of cycles as well as the cycles for using shared memory which would be $30 * 1024$.
3. (a) would not benefit from using shared memory rather it would be slower.

b. .

1. You would need a lot of cycles as you would need to keep pulling from global memory twice each from different vector. You would need 1000 first then 2000 then 3000 increasing by 1000 until the longer vector ends.
2. You would need 1000 cycles for first calculation, then 1030, then 1060, increasing by 60 until the longer vector ends.
3. (b) would benefit from using shared memory because as the calculation continues, if you use shared memory the cycles increase by 30 while cycles increase by 1000 without shared memory.