



For this assignment, I requested one node for computing resource. There are two sockets on a single node and there are 10 physical cores on a single socket. Therefore, there are 40 virtual cores on a single node I requested. To achieve maximum performance, we set the largest thread size to 40. Using more threads is not productive, and does not reflect the scaling of the algorithm.

The calculation runtime decreases as I increase the thread size. The decreasing rate is very large when the thread size is small approximately from 1 to 6, and it slowly converges after that. Note that the runtime begins to increase a little bit after 38 threads. I think it is because the shortgpu node has OS and other default processes running on the 40 cores, so there is some time-sharing happened and scheduling becomes less efficient.