CS 350 – Fall 2020, Assignment 1

Answer 1

a) **The CPU has an utilization of 28.6%. The SSD also has an utilization of 28.6%. Finally the GPU has an utilization of 42.8%**.

b) Since the system processes a request every 7ms, and has a Multiprogramming level of 1, then in a second, it can process 142 complete requests (**Throughput of 142/per second**).

c) **The bottleneck for the system is the GPU**, as every request takes 2ms of CPU time, 2ms of SSD time, and 3ms of GPU time, which might not seem like a big difference but it will eventually lead to a bottleneck of processes waiting for GPU time.

d) Based on a small simulation that lasted 10ms, **the CPU utilization is 50%, the SSD utilization is 60%, and the GPU has an utilization of 60%**.

e) The throughput for a 20ms run is 4, so, extrapullating, we have that **the throughput is around 200/per second (Given a little error margin as this isn't exactly linear)**.

f) After a quick simulation, I managed a throughput of 5 with ML=3, however I'm already seeing a significant bottleneck in GPU use, and any attempts with higher MLPs will result in a reduction in throughput, so the clear **choice for M level for maximum throughput is 3**.

Answer 2

a) After some simulations I reached that, on average, both the CPU SSD idle for 1ms every time the GPU is running a process, which means, this 1ms would be parallelizable. meaning about 1/3rd of the code is parallelizable. So doing the math for **N=2 f=1/3**, we get that the speedup is of $\frac{1}{1-\frac{1}{3}(1-\frac{1}{2})} = 1.2$, so **the new system would be 20% faster than the older one.**

b) After running another simulation, in the same period (20ms), **we had an increase in capacity of 50%**, as this time the system made it through 6 requests in 20ms instead of 4 requests without the improvements.

c) **After the upgrade the bottleneck is (for higher MLP levels), in the SSD**, as, despite the process taking the same percentage of SSD and CPU usage, the split in CPU utilization eases the use of the resource, on the other hand, the SSD is finding it hard to cope with the requests for higher MLP levels.

d) Considering that the request must wait for the CPU to be consumed, and he waited 10ms to be used, then he has been waiting for what is likely 5 other requests that are delayed, so the total time would be: **10ms(wait) + 1ms(CPU) + 20ms(SSD wait) + 3ms(GPU) + 10ms(CPU wait 2) + 1ms(CPU)=45ms**.

Answer 3

a) If we do some math, we can see that accelerating the CPU by 20% means that the average time spent in the CPU goes from 2.2mins to 1.76mins, so, applying the basic math from Amdahl's Law, **we get a speedup of 9%**, ie: the system is 0.09x faster than before.

b) On the other hand, if we do the math for the I/O upgrade, we get that the average time spent on I/O goes from 1.375mins to 1.1mins, so **we get a speedup of 5%**, ie: the system is 0.05x faster than before.

c) If we divide the speedup per dollar spent, **we get that the first option costs 1111.11USD per 1% of speedup, while the second option costs 7000USD per 1% of speedup**. So clearly the first one is more efficient

d) If we decide to spend the money on both options we are looking at a speedup of 14.9%.

e) Considering that an infinitely fast CPU would make the calculations immediate, then the time spent in the CPU would go from 2.2mins to 0mins, **giving us a speedup of 67.6%**.

f) On the other hand, if we had infinitely fast I/0 devices, we'd encounted **a speedup of 33.3%**.