

CS553 Homework #2

Computer Systems

Instructions:

- Assigned date: Friday January 31st, 2020
- Due date: 11:59PM on Monday February 10th, 2020
- Maximum Points: 100% (79 points)
- This homework can be done in groups up to 3 students
- Please post your questions to the Piazza forum
- Only a softcopy submission is required; it will automatically be collected through GIT after the deadline; email confirmation will be sent to your HAWK email address
- Late submission will be penalized at 10% per day; an email to the TA with the subject "CS553: late homework submission" must be sent

Your Assignment

1. Processors (15 points):
 - a. Today's commodity processors have 1 to 64 cores, with some more exotic processors boasting 72-cores, and specialized GPUs having 5000+ CUDA-cores. About how many cores/threads are expected to be in future commodity processors in the next five years?
 - b. How are these future processors going to look or be designed differently than today's processors?
 - c. What are the big challenges they need to overcome?
 - d. ~~Describe what a core and hardware thread is on a modern processor, and the difference between them?~~
 - e. What type of workloads are hardware threads trying to improve performance for?
 - f. Compare GPU and CPU chips in terms of their strength and weakness. In particular, discuss the tradeoffs between power efficiency, programmability and performance.
 - g. ~~Why do we not have processors running at 100GHz today (as might have been predicted in 2000)?~~
2. Threading (6 points):
 - a. ~~Why is threading useful on a single-core processor?~~
 - b. Identify what a thread has of its own (not shared with other threads):
 - c. ~~Do more threads always mean better performance?~~
 - d. ~~Is super-linear speedup possible? Explain why or why not.~~
 - e. ~~Why are locks needed in a multi-threaded program?~~
 - f. ~~Would it make sense to limit the number of threads in a server process?~~
 - g. What is the advantage of OpenMP over PThreads?
3. Network (11 points):
 - a. A user is in front of a browser and types in www.iit.edu, and hits the enter key. Think of all the protocols that are used in retrieving and rendering the main webpage from IIT. Describe the entire sequence of operations, commands, and protocols that are utilized to enable the above operation.
4. Power (12 points):
 - a. Why power consumption is critical to datacenter operations?
 - b. What is *dynamic voltage frequency scaling* (DVFS) technique?

- c. If you were to build a large \$100 million data center, which would require \$5M/year in power costs to run the data center and \$5M/year in power costs to cool the data center with traditional A/C and fans. Name 2 things that the data center designer could do to significantly reduce the cost of cooling the data center?
 - d. Is there any way to reduce the cost of cooling in (C)? If yes, how low could the costs go? Explain why or why not?
5. Storage (15 points):
- a. If a manufacturer claims that their HDD can deliver sub-millisecond latency on average, can this be true? Justify your answer?
 - b. Explain why flash memory SSD can deliver better performance for some applications than HDD.
 - c. What types of workloads benefit the most from SSD storage?
 - d. If a manufacturer claims they have built a storage system that can deliver 1 Terabit/second of persistent storage per node, would you believe them? Justify your answer to why this is possible, or not. Make sure to use specific examples of types of hardware and expected performance.
 - e. In this problem you are to compare reading a file using a single-threaded file server with a multi-threaded file server. It takes 8 msec to get a request for work, dispatch it, and do the rest of the necessary processing, assuming the data are in the block cache. If a disk operation is needed (assume a spinning disk drive with 1 head), as is the case one-fourth of the time, an additional 16 msec is required. What is the throughput (requests/sec) if a multi-threaded server is required with 4-cores and 4-threads, rounded to the nearest whole number?
6. SQL vs Spark (20 points):
- a. You hired by a company to help them decide what software stack and hardware they should adopt to store, process, and analyze 500TB (terabyte) of data. Their choices for software stack are: MySQL (<https://en.wikipedia.org/wiki/MySQL>) and Spark ([https://en.wikipedia.org/wiki/Spark \(software\)](https://en.wikipedia.org/wiki/Spark_(software))). It has been determined that most queries will only touch 1% of the data using primarily a random-access pattern. The computation to be done seems to be scalable, and that the more computing resources, the faster the computation will run, as long as it can be maintained in memory. The requirement is that there should be at least 224-cores of computing running at 2.7GHz or faster. There are no requirements on the processors used (as long as they are x86 compatible). There should be enough memory to store 1% of the dataset in memory, and there should be enough storage to reliably store 500TB of storage. If a multi-node approach is taken, the network should be as fast as possible (e.g. 100GbE) to ensure good scalability. Assume administration cost is 20% of a full-time system administrator (at a salary of \$100,000/year). Assume power costs \$0.15 per KWH, and that cooling costs are in-line with the power costs of powering the hardware. Use the ThinkMate website (<https://www.thinkmate.com>) to come up with the a solution for MySQL and one for Spark in terms of costs over a 5 year period, including hardware, power, cooling, and administration. Note that your solution has to be rack mountable (you cannot use desktops or laptops).

What you will submit

When you have finished your written responses, you should hand in:

1. **Report:** A written document (typed, named hw2-report.pdf) describing your answers to the above questions.

Submit report through GIT.

Grades for late programs will be lowered 10% per day late.