CS553 Homework #2

Computer Systems

Instructions:

- *Assigned date: Tuesday September 6th, 2022*
- *Due date: 11:59PM on Wednesday September 14th, 2022*
- *Maximum Points: 100%*
- *This homework can be done in teams of up to 3 students*
- *Please post your questions to BB*
- *Only a softcopy submission is required; submission is a 2-step process: 1) push changes to GIT repository, and email confirmation will be sent to your HAWK email address at the deadline; a confirmation document with all team member names and A# must be submitted through BlackBoard for your submission to be graded; only 1 student must submit the assignment, and only the submitting student will receive the confirmation email*
- *Late submission will be penalized at 10% per day*

Your Assignment

1. (15 points) Answer the following questions about computer processors:
    a. Describe what a core and hardware thread is on a modern processor, and the difference between them?
    b. How many cores do the fastest processors from each manufacturer have? Give an example (specific model, specs, and price).
        (a) Intel CPU (x86)
        (b) AMD CPU (x86)
        (c) IBM CPU (Power9)
        (d) ThunderX CPU (ARM)
    c. Why do we not have processors running at 1THz today (as might have been predicted in the year 2000)?
    d. Describe Moore's Law. Is it going to go on forever? If not, when will it end? Justify your answer to why it will end and when.
2. (10 points) Answer the following questions about threading:
    a. Why is threading useful on a single-core processor?
    b. Do more threads always mean better performance?
    c. Is super-linear speedup possible? Explain why or why not.
    d. Why are locks needed in a multi-threaded program?
    e. Would it make sense to limit the number of threads in a server process?
3. Processors (14 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. What type of workloads are hardware threads trying to improve performance for?
    e. Compare GPU and CPU chips in terms of their strength and weakness. In particular, discuss the tradeoffs between power efficiency, programmability and performance.
    f. Identify what a thread has of its own (not shared with other threads):
    g. What is the advantage of OpenMP over PThreads?
4. Network (10 points):

a. A user types in a browser 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.

5. Power (16 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?

6. 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?

7. 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 1PB (petabyte) of data. Their choices for software stack are: MySQL (https://en.wikipedia.org/wiki/MySQL) and Spark (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 128-cores of computing running at 2.0GHz of faster. There are no requirements on the processors used (as long as they are x86 compatible). There should be enough memory to store 0.4% of the dataset in memory, and there should be enough storage to reliably store 1PB of storage. If a multi-node approach is taken, the network should be as fast as possible (e.g. 200GbE) to ensure good scalability. Assume administration cost is 20% of a full-time system administrator (at a salary of $150,000/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).

Where you will submit

You will have to submit your solution to a private git repository created for you at git@github.com:datasys-classrooms/cs553-fall2022-hw2-<team name>.git. The repository is create through GitHub Classroom and you will need to accept the assignment by accessing this link https://classroom.github.com/a/0RWf2cgT. You will also need to create a new team or join an existing team before you can clone this repository. **Your submission** will not be graded unless you submit a confirmation document through BlackBoard (BB) that clearly shows the pushing of your final homework to your GIT repository. This confirmation document can simply be a screen shot of your final commands to push your repository to GIT. The timestamp on the BB submission will be used to determine if the submission is on-time. You can verify the validity of your submission by running the checks and tests from https://github.com/datasys-classrooms/cs553-fall2022-hw2-testing. If you cannot access your repository contact the TAs. You can find a git cheat sheet here: https://www.git-tower.com/blog/git-cheat-sheet/. **You must also include the names and A# of all your team members in this confirmation document.**

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.