HW2 Computer System

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?

- A core is the processing unit of a CPU which is capable of executing instructions independently. The term “multi-core” refers to a CPU which has multiple processing units, each processing unit having the capability of executing instructions independently. The term “core” refers to a physical core present on the CPU.
- A hardware thread refers to a separate execution context present on a physical core. When hardware threads are enabled, the processor exposes two execution contexts per physical core.
- This implies that the operating system views a physical core as two separate logical cores and can schedule two separate software threads on these logical cores.
- Unlike multi-core technology, not all the resources of a processing unit are not replicated and the two scheduled software threads are still time-sharing the same physical core. Hardware threads allow the processor to utilize its idle resources and thus increase the throughput of the processor (core) by unto 30%.

b.How many cores do the fastest processors from each manufacturer have? Give an example (specific model, specs, and price).

- (a) Intel CPU (x86) 18 cores. Intel® Core™ i9-10980XE. It is priced at 1050$
- (b) AMD CPU (x86) The AMD Risen ThreadRipper 3990X has 64 cores and was considered the most powerful desktop processor in 2021. Its price is 8,620$.
- (c) IBMCPU(Power9) 24 cores, both the scale-out and scale-up versions of POWER9 processors come with the 24-core SMT4 variants with 96 threads supported. The IBM Power9 module 02CY296 costs 2,865$.
- (d) ThunderX CPU (ARM) 32 cores, Cavium ThunderX2 ARM processor. Its price range is between 800$ - 1795$.

c.Why do we not have processors running at 1THz today (as might have been predicted in the year 2000)?

- We do not have processors running at 1THz today because as the clock speed of the processor increases, so does its power consumption. A processor running at 1THz might be possible in theory, but it won’t be feasible practically because of its power requirements and heat dissipation. Furthermore, as the clock speed of the processor increases, its cost also increases. A processor running at 1THz would be practically impossible for a normal person to afford.

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.

- Moore’s law is an observation which states that the number of transistors on a microchip doubles every two years, whereas the price of the microchip reduces to half. The observation was state by Gordon Moore in the year 1965, and the law has been going strong for more than 50 years. However, we are very close to the physical limits of Moore’s law, meaning it won’t be forever possible for us to keep doubling the number of transistors on a microchip. The amount of energy required to cool down a transistor is more than the energy which actually passes through the transistor itself. It is predicted that somewhere in this decade, we are going to reach the physical limit of Moore’s law.

---

2. (10 points) Answer the following questions about threading:

a.Why is threading useful on a single-core processor?

- A single-core processor executes instructions sequentially. Moreover, a processor time-shares between multiple processes using the scheduler. When a processor context switches from one process to another, there is an overhead associated with saving the registers of the running process, loading the process control block of the queued process in the CPU registers and switching the address space. This overhead reduces when a processor has to switch from one thread to another (belonging to the same process) since the process threads share the same address space. Unlike process switching, the thread switching is efficient because it involves switching some attributes of the process control block like program counter, stack pointer etc instead of the entire address space.

b.Do more threads always mean better performance?

- No, increasing the number of threads does not always mean better performance.
- It is quite clear that the switching between threads is easier for a processor because they share the same address space (threads belonging to the same process), but there are a few bottlenecks associated with increasing the number of threads for a process.
    - First bottleneck is the number of physical cores available. If the number of cores are equal to the number of threads for a process, then maximum performance is achieved. However, as the number of threads become more than the physical cores, the performance increase starts to diminish. This is because as the number of threads increase, the wait time for each thread starts to increase.
    - The second limitation is the amount of memory available. Though this factor is hardly an issue in modern computing systems because of the large amount of physical memory available, spawning a large number of threads can take its toll on systems where large memory is not available like embedded systems.
    - The third issue with increase the number of threads is locking. More often than not, the threads belonging to a single thread share memory access which is locked to avoid race conditions. If the number of threads is too large, locking threads for read/write access can take impact the performance of the process and the system.

c.Is super-linear speedup possible? Explain why or why not.

- It's possible. The speedup of an algorithm is usually limited by Gustafson’s law which states that the speedup is limited by the number of processing units. However, researchers have observed and reported a speedup greater than what is predicted by Gustafson’s law. Such a speedup is known as super linear speedup. In a super linear speedup, the exception time of the algorithm is reduced more than Ts/N where Ts is the sequential time of the algorithm and N is the number of processing units used. Researchers have cited the increased cache size as the main reason behind the super linear speedup. The reasoning argues that the parallel execution uses more cache memory than the sequential execution and for a certain size of the problem, it is possible for for the entire problem to be present in the cache memory. Furthermore, some researchers have gone further in their investigation and concluded that loosely coupled processors are more likely to show super-linear speedup than the closely coupled processors due to the large amount of L3 cache available to each core. Here is the link to the papaer where these results are discussed https://annals-csis.org/Volume_8/pliks/pliks/498.pdf

d. Why are locks needed in a multi-threaded program?

- Locks are needed in a multi-threaded program to avoid race conditions. Race conditions occur when multiple threads try to access and modify the same memory space. The usage of locks guarantees that only a single thread can access the critical section (shared memory between threads) at any given time, avoiding multiple write operations on the same memory address at the same time.

e. Would it make sense to limit the number of threads in a server process?

- Yes, it does make sense to the limit the number of threads in a server process due to multiple reasons.
  - Firstly, each thread has its own stack memory which implies that the memory consumption increases as the number of threads increase.
  - Secondly, increasing the number of threads does not always mean better performance. It is usually recommended to limit the number of process threads to the number of physical cores for computation-extensive processes and equal to the number of hyper threaded/logical cores for memory-extensive processes.
- Hence, limiting the number of threads does guarantee a better performance as opposed to spawning threads without any limit.

---

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?

- In the future, there will be need of heterogenous systems where there will be cores designed for specific purposes. In the next five years, with the declining trend of Moore’s law, there are still some capabilities which they can be tried to make an improvement in power, performance and/or area. The demand currently arises for the domain specific architecture of the core.
- If the Moore’s law still behave like it currently works in the upcoming 5 years, the number of cores/threads will be quadrupled as compared to that of today.

b. How are these future processors going to look or be designed differently than today’s processors?

- Upcoming future demands heterogenous systems where different tasks get special treatment by designing the cores in such way. Also, another implementation that the architects can do is by creating chips which would have different number of cores say 2 cores on one chip and 8 cores on another chip. The chip architects must work hand in hand with systems engineers to devise a plan to figure out how they can partition functions to put together in heterogenous way. We can also look at the idea of optical computers which is currently being developed.

- On comparing the optical computers, they are much faster than electronic computers. In that future, there is chance that we will have larger CPUs but with a lower density of transistors. In terms of current market competitors, ARM is going for the more cores, more better approach, Intel is going forward with more transistors more better approach, and finally AMD is moving forward with a hybrid of both with more moderation more better approach.

c. What are the big challenges they need to overcome?

- The biggest challenges that they need to overcome are how many greater numbers of transistors that can be fitted on the silicon wafer chip, even if some solution is developed and another millions or billions of transistors are added to the current model how will they tackle the issue of heating and the need to cool down the chip.
- Another big challenge is the power consumption, with the increase in number of transistors, power consumption also increases. Also, to incorporate the architecture changes that are the need of the invention, they have to the research and development.

d. What type of workloads are hardware threads trying to improve performance for?

- Hardware threads use hyperthreading technology. In hyperthreading, with a single CPU core and an extra CPU state and interrupt logic unit the OS is made to think that it has two separate cores. In an ideal situation, the performance boost should be twice the amount of performance power with a normal single CPU. However, this is not the case, there is at maximum a performance boost of 15% to 20%. This is also possible in the workloads where the instruction pipeline can be utilized. In all the other types of workloads, the hardware threads do not perform to that extent and needs improvement. These workloads can be the software which does not support hardware threads and software implementation is needed to run, intensive applications which require lot of processing in those applications, hardware performance is degraded due to overheating, and some application-dependent workloads also need performance enhancements.

e. Compare GPU and CPU chips in terms of their strength and weakness. In particular, discuss the tradeoffs between power efficiency, programmability and performance.

- CPU needs more memory than GPU, while GPU requires less memory than a CPU.
- In terms of speed, CPU has lower speeds than a GPU.
- Since CPUs have powerful cores, they are more powerful than GPU which have weak cores.
- CPU are suitable for serial instruction processing while GPU are suitable for parallel instruction processing.
- The emphasis of CPU is low latency, while the GPU emphasis more on high throughput.
- In terms of power efficiency, programmability and performance, GPU can outperform CPU in single and multithreaded operations. Since it has more ALU than CPU, the execution finishes faster in a GPU, so the total power consumption is less than CPU. GPU have higher execution rate and execute upto 800 instructions more per clock than a CPU making them more efficient. Also, GPUs today are more programmable which gives them the flexibility to accept and accelerate a wide range of applications other than just graphic rendering.

f. Identify what a thread has of its own (not shared with other threads):

- For threading implementation, resources are shared between threads such as script to execute, data to load, file to use, etc. However, certain resources like program counter, stack, stack pointer, and registers are not shared by threads and each thread has its own stack, stack pointers, and registers.

g. What is the advantage of OpenMP over PThreads?

- PThreads or POSIX threads is a low-level API for working with threads and it gives a lot of flexibility to manage threads. OpenMP or Open Multi-Processing is a high-level API for working with threads and is more portable than PThreads. Advantages of OpenMP over PThreads:
  - It is simple and portable than PThreads.
  - It has incremental parallelism that means it can work on one part of program at one time without any code changes.
  - It uses the fork-join method to achieve parallelism which may result in super linear speed-up.
  - It has the advantage of working on cross platform and can handle parallelization of loops as well.

---

4. Network (10 points):

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.

- Without the internet, we would be in an existential crisis now. By utilizing networks, wireless technology, and communications protocols, technology allows us to relate to any data, obtain any information, or establish connections. Going back to our original question, when we visit a website, behind the scenes we connect to web hosting servers that are housed in enormous data centers all over the world to access the information that is hosted there.

- The internet is readily available to us. Today, everyone of any age group who wants knowledge will Google it as soon as possible because of technology's ease of accessibility to the internet. When we say that someone "Googles it," we mean that they type their search term into the address www.google.com and push enter to access potentially millions of webpages containing the information they were looking for.

- When a user generally types in a URL like www.iit.edu and presses enter, a specific process occurs in such a tiny amount of time that most of the time the user has no idea what has happened. These seven steps can be used to explain the entire process.

- 1. User enters in the URL in a browser application:
  - The 4 components of a URL generally consist of protocol, hostname, port, and path to resource file. It has the following format “protocol://hostname:port/path_to_resource_file”. Using www.iit.edu as an example we see that it does not contain any specific protocol, port, or path to any resource file. As a result, the browser will use HTTP as the default protocol and port 80 as a default port and the root directory as the default file path.
  - Any website, including the one in our query, has a variety of domain levels. With www.iit.edu, iit is the primary domain, www is the subdomain, and edu is the top-level domain. Consider an analogy, to locate someone's home, one must know where they live, or their address. Like that, the browser needs these domains in order to identify the IP address of the server hosting the

website. The web browser then uses this IP address to connect to the servers via the TCP/IP (Transmission Control Protocol/Internet Protocol) protocol. The IP address is regarded as the address where the website resides in the world of the internet.

- 2. DNS lookup:
    - Continuing with our earlier illustration, we may checkup someone's address in the phone book to determine where they are in the actual world. In a similar fashion, the browser uses the Domain Name System, or DNS, to find up the IP address. An IP address is entirely composed of numbers, unlike a real-world address, which is described using both words and numbers. IPv4 and IPv6 are the two different types of IP addresses. Both contain only numbers despite having different formats. The user must discover the IP address of www.iit.edu in order to access it.
    - The DNS lookup translates the words “www”, “iit”, “edu”, and “.” and tries to find www.iit.edu using the domains that user typed in the world of web starting with browser cache, if not found, then looks into operating system cache, if not found, uses the Internet service providers cache and tries to find it all over the world, if not found, it reports an error that no such webpage found.
- 3. TCP connection is formed:
    - The whole TCP connection formation happens in 3 steps. Once the IP address is found, the connection to the server associated with this IP is formed. For this connection, communication and exchange of data is done via TCP protocol. TCP client starts the communication by sending a synchronization packet to website server on the communication port, in our case, port 80. If the port 80 is open for communication, it will reply to the client via a synchronization/acknowledgement packet. To this reply, the TCP client sends its own acknowledgement packet, and the transfer of data starts with this process.
- 4. HTTP request sent to server:
    - In the protocol, a translated version of the URL is sent by the browser which is a HTTP GET request which comprises of various headers and inform how the information and connections should be handled. As soon as the server receives this request, it starts to map the request to a file or program and once done, sends back a response. Even if something goes wrong, an error message will be the response.
- 5. Server’s role in response to HTTP request:
    - The web server assumes that firewall is allowing incoming and outgoing connections and it has the responsibility to serve static web content like HTML and CSS scripts. In case of a dynamic web content requested by HTTP request, the dynamic web content goes to application server which in turn connects to a database to get the information.
- 6. Server sends its HTTP response:
    - The HTTP response just like the previous one, it contains status code, one or more headers that has information about the content that should returned in the body/content.
- 7. Webpage display:
    - After these all steps the final step/ goal of this process is to see the webpage. HTTP and HTML contents work in this step. The browser will try to send multiple requests to HTTP GET to host all the files. After the code content is full read, a complete webpage will be displayed.

---

5. Power(16 Points)

a. Why power consumption is critical to datacenter operations?

- A data center is a dedicated space within a building used to house computer systems and related components. It needs to provide basic capacity, must include uninterruptible power supplies, and ensure high availability and avoid single points of failure, so redundant capacity and backup power are essential. Based on this, the store's consumption is enormous.
- And from the business point of view, the data center stores a lot of core data, such as customer's account password identity information, transaction information, etc. If at the time when the service needs to access the data to verify the customer's identity information to fulfill the demand of placing orders.
- If the power supply to these systems fluctuates, then the service cannot access the data, and eventually the terminal cannot realize the order or match the user with the error (data confusion), which will cause serious consequences such as business loss or user information leakage, so the power consumption is critical to the data center.

b. What is dynamic voltage frequency scaling (DVFS) technique?

- Dynamic voltage and frequency scaling (DVFS) is the adjustment of the power and speed settings of the various processors, control chips, and peripherals of a computing device to optimize resource allocation for tasks and to maximize power savings when those resources are not needed.
- Dynamic voltage scaling and dynamic frequency scaling are subsets of DVFS that dynamically reduce voltage (only) based on performance requirements. Because higher frequencies require higher supply voltages for the digital circuit to yield results.

- Link:
  - https://www.techtarget.com/whatis/definition/dynamic-voltage-and-frequency-scaling-DVFS
  - https://en.wikipedia.org/wiki/Dynamic_frequency_scaling
  - https://en.wikipedia.org/wiki/Dynamic_voltage_scaling

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?

- The most common energy efficiency metric for data center energy efficiency is Power Usage Effectiveness (PUE), PUE = Total Facility Power / IT Equipment Power, so there are two ways to reduce this metric, which are to reduce redundancy consumption and to reduce the consumption of devices such as CPU/GPU.

- 1. Using multiple cooling technologies with siting geography of the location：
  - Such as liquid cooling，Air cooling，Raise the temperature, Waste heat reuse and Free cooling.
  - If the data center is built in the Arctic Circle, it is more suitable for air cooling, if it is built on the west coast of the United States, it is more suitable for liquid cooling technology, if it is built at the equator, it may be a good way to place solar panels and use solar power.
  - For example, Google uses the indoor temperature to 80 °F and uses outdoor air for cooling, and Microsoft uses boiling liquid technology to reduce costs and improve cooling efficiency.
- 2. Different data center architectures or power management techniques can reduce power consumption.

  - According to Moore's law, the number of transistors will double every two years, so naturally the power consumption will increase, so we have to be able to use effective power management and system management to reduce power consumption.
  - For example, we can use the dynamic voltage frequency scaling (DVFS) technique to give the required power resources on demand by controlling the CPU frequency.
  - Changing data center network architectures, according to the paper "Power Analysis of Data Center Architectures", a three-tier architecture shows the best performance. When used as proposed - consumes a lot of power due to the resilient paths to the servers. When not used networking components are switched off, the power consumption can be reduced by about 60% as shown with the elastic-tree architecture.

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?

- Using AI models, the Data Center Infrastructure Management (DCIM) approach was optimized to determine cooling performance based on scenarios, with Thermostat Control to reduce the frequency of cooling.
- For example, Google Data center consuming significantly less energy than a typical data center. They raise indoor temperatures to 80°F, use outdoor air for cooling, and build custom servers.
- The artificial intelligence approach involves the use of neural networks, a method of demonstrating cognitive behavior to recognize patterns between complex input and output parameters. it can reduce the error to 0.004 Power Utilization Effectiveness (PUE) or 0.34-0.37 percent of the Google's best-performing data center can have a PUE even lower than 1.06.

- Link:
  - https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8581422
  - https://www.bmc.com/blogs/dcim-data-center-infrastructure-management/

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?

- HDD can't deliver sub-millisecond latency on average.
  - A sub-millisecond is 0.001s.
  - The disk latency is the time delay between a request for data and the return of the data. Three main factors are determined the disk latency: rotational latency, seek time and transfer time.
  - We only consider rotational latency, today's most common RPM rates, in both laptop and desktop PCs are between 5400 and 7200RPM. The current best performing HDD is 15000RPM.
- We can calculate the latency considering only RPM.
  - 5400RPM: 1/(5400/60) = 11ms
  - 7200RPM: 1/(7200/60) = 8ms
  - 15000RPM: 1/(15000/60) = 4ms
- In real world, we also need to consider seek time and transfer time after finding the data, so it's not just rational latency, so even the best HDDs today can't reach sub-millisecond latency on average.

b. Explain why flash memory SSD can deliver better performance for some applications than HDD.

- SSD(Solid-state Drive) is a solid-state storage device that using flash memory to store data. Compared with HDD, it lack the physical spinning disks and movable read-write heads. But SSDs are typically more resistant to physical shock, run silently, and have higher input/output rates and lower latency. HDDs are made of magnetic tape and have mechanical components inside that are larger and slower to read and write than SSDs.
    - HDD read and write speeds are limited by mechanical mechanisms, i.e. rotational speed, and the current maximum speed of HDDs is 15,000RPM. Although parallel disks, caching, and more RAM can be used to increase some speeds, eventually performance will bottleneck.
    - SSDs use electrical circuits and have no physical moving parts. This reduces latency. At the lowest level, floating gate transistors register a charge (or lack thereof) to store data. The gates are organized into a gate pattern, which is itself organized into a block. The size of the blocks can vary, but each row that makes up the gate is called a page.

c. What types of workloads benefit the most from SSD storage?

- Performance: SSD offer faster load times for games, applications and movies. As an example, sometimes we need situations where the operating system requires a shorter boot time and the application data needs to be accessed frequently, SSDs are superior to HDDs.
- Reliability: SSDs have no moving parts, so they have better reliability, That's because without movement, SSDs aren't affected by vibration or related thermal issues.
- Less power consumption: Because of the technology they use, SSDs are lighter and more resistant to movement and drops. In addition, SSDs consume less power, allowing computers to run cooler. It is used for computationally intensive tasks.

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.

- It's impossible to build a storage system that can deliver 1 Terabit/second of persistent storage per node. The performance characteristics of any storage system are evaluated by 2 main metrics: input/output operations per second (IOPS), and data throughput.
    - The IOPS metric shows how many read and/or write operations per second a storage device can perform.
    - The block size means that the largest amount of bits/bytes that can be allocated to a single I/O operation.
    - Data throughput measures the amount of data transferred to and from a storage device per second.
    - `Throughput = IOPS * block size`
- Business Data:
    - Generally a HDD will have an IOPS range of 55-180, while a SSD will have an IOPS from 3,000 – 40,000.
    - SQL server logs in 64 kb blocks, while Windows Server might use 4 kb blocks, and the underlying vSphere hypervisor uses 1 MB blocks.
    - Most SSDs have blocks of 128 or 256 pages, which means that the size of a block can vary between 256 KB and 4 MB.

-
    - We can use maximum value to calculate throughput: 4MB * 40,000 = 160,000 MBit/s = 160 Gbit/s much less than 1Tbit/s.
    - PCIe 5.0 is the fastest high-speed serial computer expansion bus standard that has been hardware implemented to date.It has a maximum speed of **128GB/sec**(Theoretical)
- A terabit is a measurement for 1 trillion(1*10^12) bits or pieces of binary data. (1Tbit/s = 1000Gbit/s = 1,000,000Mbit/s).
- As we all know, the performance and throughput rate of SSD is better than HDD, so let's use SSD (best SSD in 2022) as an example:
    - Silicon Motion's MonTitan PCIe 5.0 SSD: **14GB/s**, **3M IOPS**
    - SK hynix Platinum P41 1TB PCIe NVMe Gen4 M.2 2280 Internal SSD: **Sequential read throughput: 7,000 MB/s**, **IOPS: 1,400,000**

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?

- Single-threaded file server(single core):
    - **Request per second from block cache**: 1000ms / 8msec = **125req/s**
    - **Request per second from disk**: 1000ms / (8msec + 16msec) = **42req/s**
    - Due to one-fourth(**1/4**) of the time the data can getting from disk, so the throughput is:
        - 1/4 * **Request per second from disk** + 3/4 * **Request per second from block cache**
        - = (1/4) * (1000/(8+16)) + (3/4) * (1000/8) = **104 req/s**
- Multi-threaded file server(exp: 4-cores and 4-threads):
    - We assume a spinning disk drive with 1 head. So if a request want to get a data from disk, multi-thread need wait until last thread finish reading. If a request get a data from block cache, they can be read simultaneously and executed in parallel.
    - 1/4 * Request per second from disk + 3/4 * Request per second from block cache multiply 4.
    - = (1/4) * (1000/(8+16)) + (3/4) * (1000/8) * **(4)** = **385req/s**

7.SQL vs Spark (20 points):

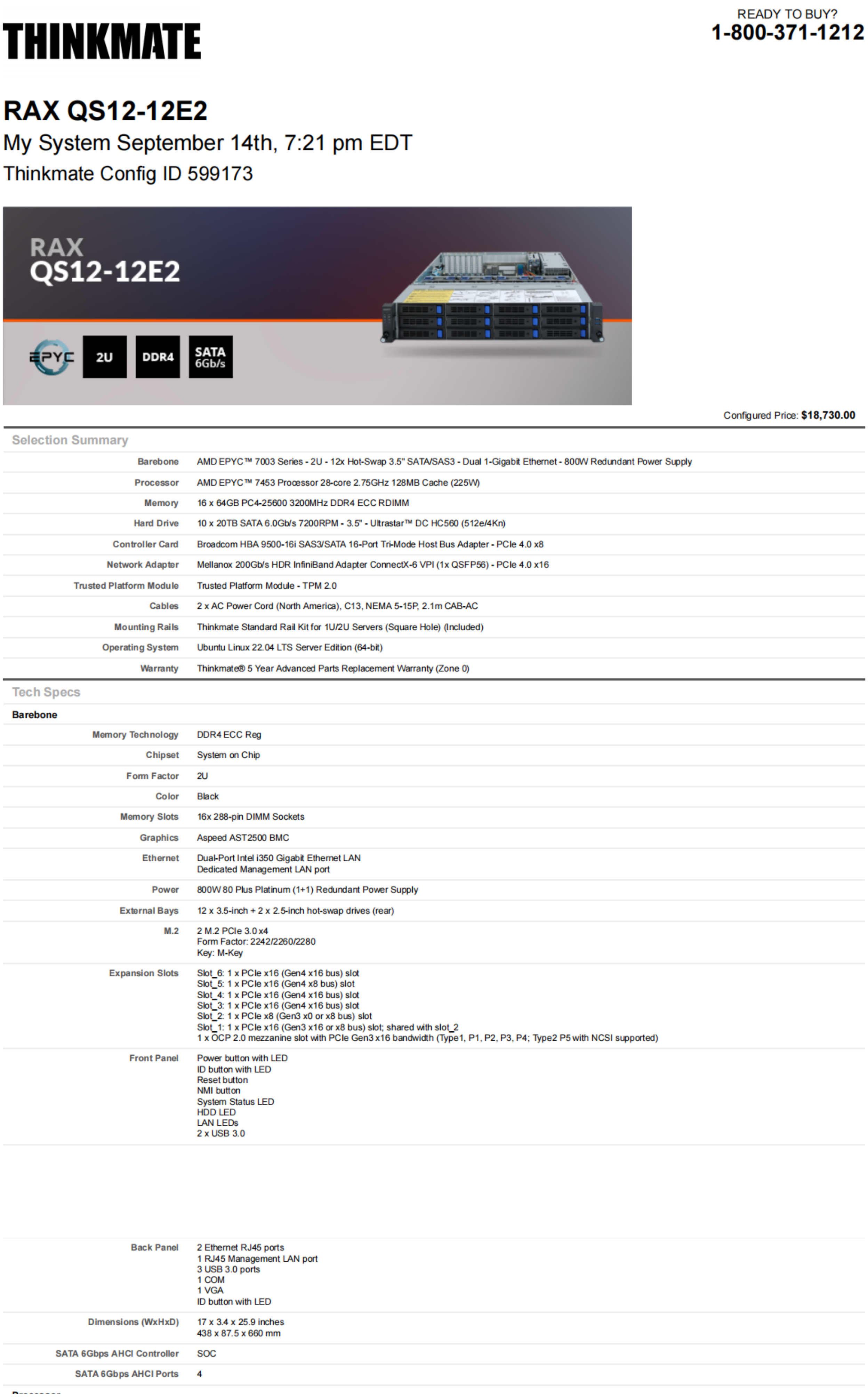
a. You hired by a company to help them decide what software stack and hardware they should adopt to store, process, and analyze 1PB (petabyte) of data. Their choices for software stack are: MySQL (https://en.wikipedia.org/wiki/MySQL) and Spark (https://en.wikipedia.org/wiki/Spark_(software)). It has 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)

- MySql only use on CPU core per query, whereas Spark can use all cores on all cluster nodes. So in MySql solution, we only use single node, but the Spark we have to use multiple hardware and deploy them as cluster.
- Overall hardware config:
    - Processor:
        - **128-cores** (The sum of all the node)
        - Computing running speed: **2.0GHz** (Single node)
        - x86 compatible => **Intel x86(Core, Pentium, Xeon) + AMD**
    - Storage: **1PB** (The sum of all the node)
    - Memory: 1PB * 0.4% =1024 * 1024GB * 0.4 = 4096GB = **4TB** (The sum of all the node)
    - Network: **200GbE** (Single node) **(Only Spark server consider)**
- Spark Solution: Assume **5 nodes** hardware and deploy them as a cluster.
- Hardware **Per 1 node**: 25.6core 2.0GHz, 204.8TB, 0.8TB, 200GbE
    - **BAREBONE**: AMD EPYC™ 7003 Series - 2U - 12x Hot-Swap 3.5" SATA/SAS3 - Dual 1-Gigabit Ethernet - 800W Redundant Power Supply. (https://www.thinkmate.com/system/rax-qs12-12e2)
    - **PROCESSOR**: AMD EPYC™ 7453 Processor 28-core 2.75GHz 128MB Cache (225W)
    - **MEMORY**: 16 * 64GB PC4-25600 3200MHz DDR4 ECC RDIMM
    - **Storage**: 10 * 20TB SATA 6.0Gb/s 7200RPM - 3.5" - Ultrastar™ DC HC560 (512e/4Kn)
    - **Controller Card**: Broadcom HBA 9500-16i SAS3/SATA 16-Port Tri-Mode Host Bus Adapter - PCIe 4.0 x8
    - **NETWORK ADAPTER**: Mellanox 200Gb/s HDR InfiniBand Adapter ConnectX-6 VPI (1x QSFP56) - PCIe 4.0 x16
    - **TRUSTED PLATFORM MODULE**: Trusted Platform Module - TPM 2.0
    - **2CABLES**: AC Power Cord (North America), C13, NEMA 5-15P, 2.1m CAB-AC
    - **MOUNTING RAILS**: Thinkmate Standard Rail Kit for 1U/2U Servers (Square Hole) (Included)
    - **OPERATING SYSTEM**: Ubuntu Linux 22.04 LTS Server Edition (64-bit)
    - **WARRANTY**: Thinkmate® 5 Year Advanced Parts Replacement Warranty (Zone 0)
    - **Switch**: Mellanox Spectrum-2 SN3700 32-Port 200GbE Open Ethernet Switch with ONIE - Part ID: MSN3700-VS2RO
- Cost:
    - Hardware: $18,730 * 5 = 93650
        - Switch: $27,460
    - 5 year's Power cost: $0.15 * 24 * 365 * 5 * 0.4695 * 5 = 15423.075
    - 5 year's Cooling cost: Same as power cost = 15423.075
    - Administration cost: 20% * $150,000 * 5 = $150,000
    - **Total: $274496.15(Not include switches)** (93650 + 15423.075 + 15423.075 + 150,000)

  - **Overall Total: $301956.15** (93650 + 15423.075 + 15423.075 + 150,000 + 27460)

- Mysql Solution: Only 1 nodes

- Hardware **Per 1 node**:
  - **BAREBONE**: AMD EPYC™ 7003 Series - 1U - 4x 3.5" SATA/SAS3 - 1x M.2 NVMe - Dual Intel 1-Gigabit Ethernet (RJ45) - 1200W Redundant (https://www.thinkmate.com/system/rax-qs4-21e2/599148)
  - **PROCESSOR**: 2 * AMD EPYC™ 7713 Processor 64-core 2.00GHz 256MB Cache (225W)
  - **MEMORY**: 32 * 128GB PC4-25600 3200MHz DDR4 ECC RDIMM
  - **Storage**:
    - 57 * 18TB SATA 6.0Gb/s 7200RPM - 3.5" - Ultrastar™ DC HC550 (512e/4Kn)
    - 1TB SATA 6.0Gb/s 7200RPM - 3.5" - Ultrastar™ DC HA210 (512n)
  - **Controller Card**: Broadcom HBA 9500-8i SAS3/SATA 8-Port Tri-Mode Host Bus Adapter - PCIe 4.0 x8
  - **NETWORK ADAPTER**: Mellanox 200Gb/s HDR InfiniBand Adapter ConnectX-6 VPI (1x QSFP56) - PCIe 4.0 x16
  - **TRUSTED PLATFORM MODULE**: Trusted Platform Module - TPM 2.0
  - **CABLES**:
    - 2 * AC Power Cord (North America), C13, NEMA 5-15P, 2.1m CAB-AC
    - 0.5-Meter External SAS Cable - 12Gb/s to 6Gb/s SAS - SFF-8644 to SFF-8088
    - AC Power Cord (North America), C13, NEMA 5-15P, 2.1m CAB-AC
  - **MOUNTING RAILS**: Thinkmate Standard Rail Kit for 1U/2U Servers (Square Hole) (Included)
  - **OPERATING SYSTEM**: Ubuntu Linux 22.04 LTS Server Edition (64-bit)
  - **WARRANTY**:
    - 2 * Thinkmate® 5 Year Advanced Parts Replacement Warranty (Zone 0)
- Cost:
  - Hardware: $30,369.71 + $56,861.00 = $87230.71
  - 5 year's Power cost: $0.15 * 24 * 365 * 5 * (0.7455 + 0.3705) = 7332.12
  - 5 year's Cooling cost: Same as power cost = 7332.12
  - Administration cost: 20% * $150,000 * 5 = $150,000
  - **Total: $251894.95** (87230.71 + 7332.12 + 7332.12 + 150,000)



THINKMATE

READY TO BUY?
1-800-371-1212

RAX QS12-12E2

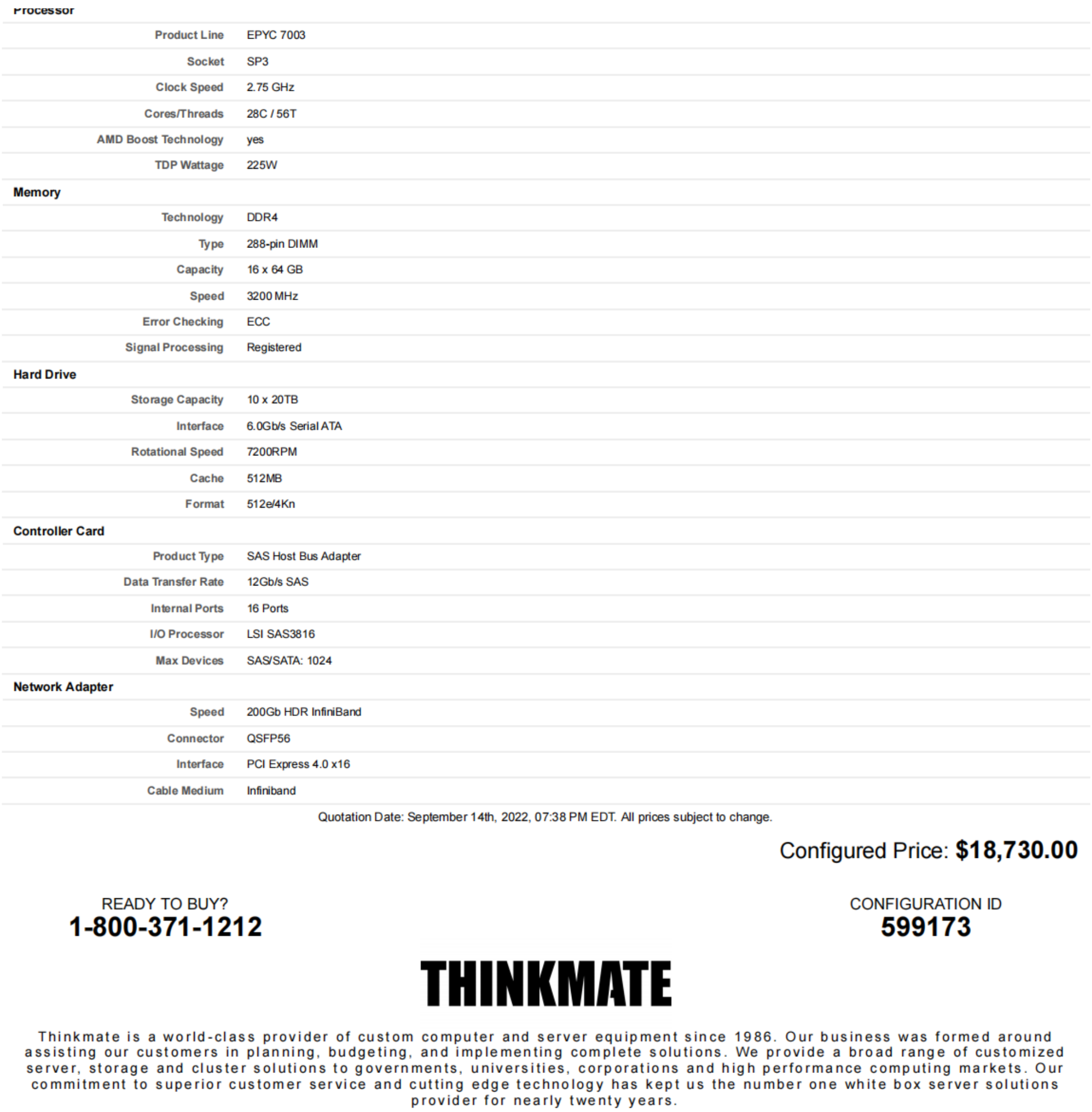
My System September 14th, 7:21 pm EDT

Thinkmate Config ID 599173

Configured Price: **$18,730.00**

Selection Summary	
Barebone	AMD EPYC™ 7003 Series - 2U - 12x Hot-Swap 3.5" SATA/SAS3 - Dual 1-Gigabit Ethernet - 800W Redundant Power Supply
Processor	AMD EPYC™ 7453 Processor 28-core 2.75GHz 128MB Cache (225W)
Memory	16 x 64GB PC4-25600 3200MHz DDR4 ECC RDIMM
Hard Drive	10 x 20TB SATA 6.0Gb/s 7200RPM - 3.5" - Ultrastar™ DC HC560 (512e/4Kn)
Controller Card	Broadcom HBA 9500-16i SAS3/SATA 16-Port Tri-Mode Host Bus Adapter - PCIe 4.0 x8
Network Adapter	Mellanox 200Gb/s HDR InfiniBand Adapter ConnectX-6 VPI (1x QSFP56) - PCIe 4.0 x16
Trusted Platform Module	Trusted Platform Module - TPM 2.0
Cables	2 x AC Power Cord (North America), C13, NEMA 5-15P, 2.1m CAB-AC
Mounting Rails	Thinkmate Standard Rail Kit for 1U/2U Servers (Square Hole) (Included)
Operating System	Ubuntu Linux 22.04 LTS Server Edition (64-bit)
Warranty	Thinkmate® 5 Year Advanced Parts Replacement Warranty (Zone 0)

Tech Specs	
Barebone	
Memory Technology	DDR4 ECC Reg
Chipset	System on Chip
Form Factor	2U
Color	Black
Memory Slots	16x 288-pin DIMM Sockets
Graphics	Aspeed AST2500 BMC
Ethernet	Dual-Port Intel i350 Gigabit Ethernet LAN Dedicated Management LAN port
Power	800W 80 Plus Platinum (1+1) Redundant Power Supply
External Bays	12 x 3.5-inch + 2 x 2.5-inch hot-swap drives (rear)
M.2	2 M.2 PCIe 3.0 x4 Form Factor: 2242/2260/2280 Key: M-Key
Expansion Slots	Slot_6: 1 x PCIe x16 (Gen4 x16 bus) slot Slot_5: 1 x PCIe x16 (Gen4 x8 bus) slot Slot_4: 1 x PCIe x16 (Gen4 x16 bus) slot Slot_3: 1 x PCIe x16 (Gen4 x16 bus) slot Slot_2: 1 x PCIe x8 (Gen3 x0 or x8 bus) slot Slot_1: 1 x PCIe x16 (Gen3 x16 or x8 bus) slot; shared with slot_2 1 x OCP 2.0 mezzanine slot with PCIe Gen3 x16 bandwidth (Type1, P1, P2, P3, P4; Type2 P5 with NCSI supported)
Front Panel	Power button with LED ID button with LED Reset button NMI button System Status LED HDD LED LAN LEDs 2 x USB 3.0
Back Panel	2 Ethernet RJ45 ports 1 RJ45 Management LAN port 3 USB 3.0 ports 1 COM 1 VGA ID button with LED
Dimensions (WxHxD)	17 x 3.4 x 25.9 inches 438 x 87.5 x 660 mm
SATA 6Gbps AHCI Controller	SOC
SATA 6Gbps AHCI Ports	4



Processor	
Product Line	EPYC 7003
Socket	SP3
Clock Speed	2.75 GHz
Cores/Threads	28C / 56T
AMD Boost Technology	yes
TDP Wattage	225W
Memory	
Technology	DDR4
Type	288-pin DIMM
Capacity	16 x 64 GB
Speed	3200 MHz
Error Checking	ECC
Signal Processing	Registered
Hard Drive	
Storage Capacity	10 x 20TB
Interface	6.0Gb/s Serial ATA
Rotational Speed	7200RPM
Cache	512MB
Format	512e/4Kn
Controller Card	
Product Type	SAS Host Bus Adapter
Data Transfer Rate	12Gb/s SAS
Internal Ports	16 Ports
I/O Processor	LSI SAS3816
Max Devices	SAS/SATA: 1024
Network Adapter	
Speed	200Gb HDR InfiniBand
Connector	QSFP56
Interface	PCI Express 4.0 x16
Cable Medium	Infiniband

Quotation Date: September 14th, 2022, 07:38 PM EDT. All prices subject to change.

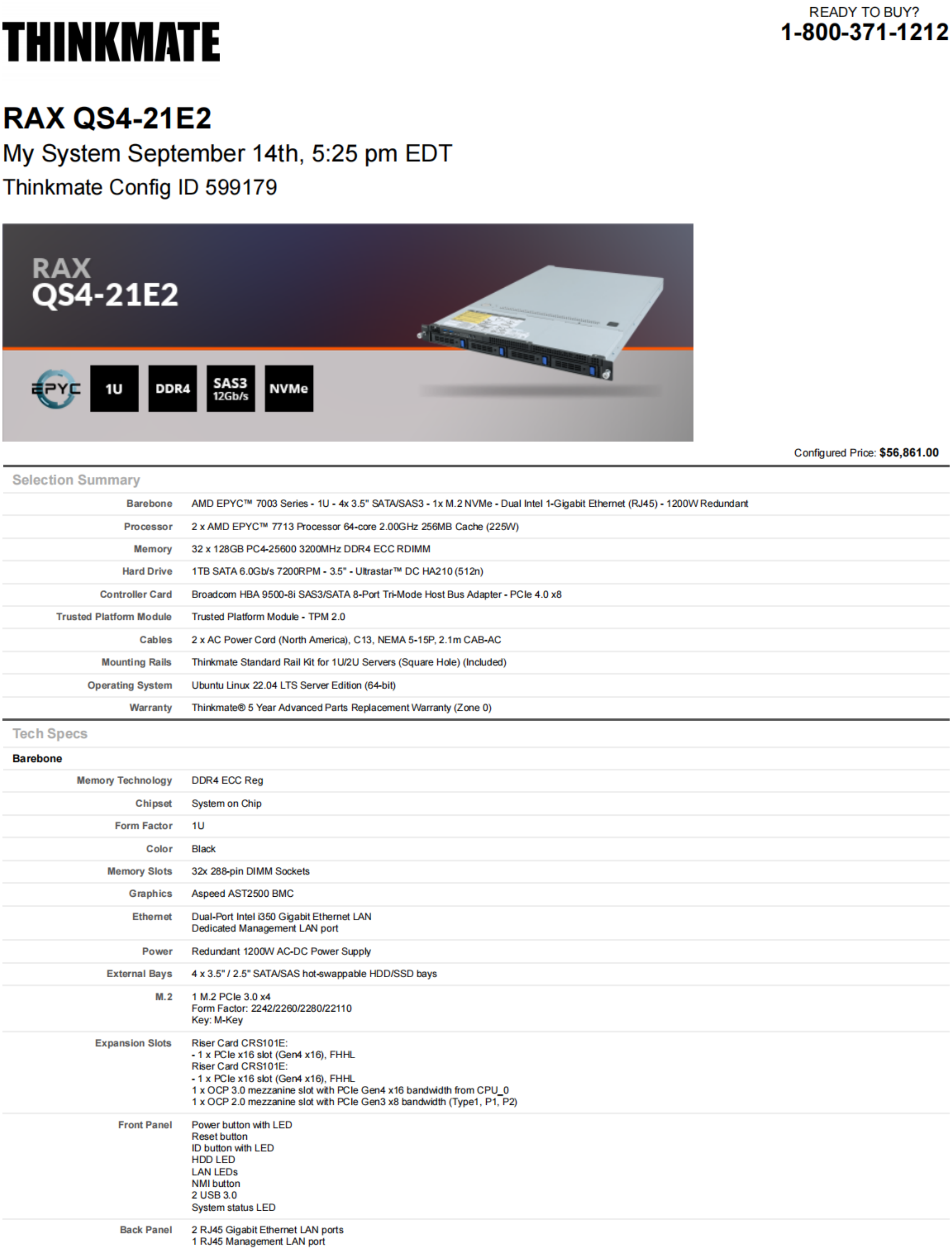
Configured Price: **$18,730.00**

READY TO BUY?
1-800-371-1212

CONFIGURATION ID
599173

THINKMATE

Thinkmate is a world-class provider of custom computer and server equipment since 1986. Our business was formed around assisting our customers in planning, budgeting, and implementing complete solutions. We provide a broad range of customized server, storage and cluster solutions to governments, universities, corporations and high performance computing markets. Our commitment to superior customer service and cutting edge technology has kept us the number one white box server solutions provider for nearly twenty years.


RAX
QS4-21E2
EPYC
1U
DDR4
SAS3 12Gb/s
NVMe

THINKMATE

READY TO BUY?
1-800-371-1212

RAX QS4-21E2

My System September 14th, 5:25 pm EDT

Thinkmate Config ID 599179

Configured Price: **$56,861.00**

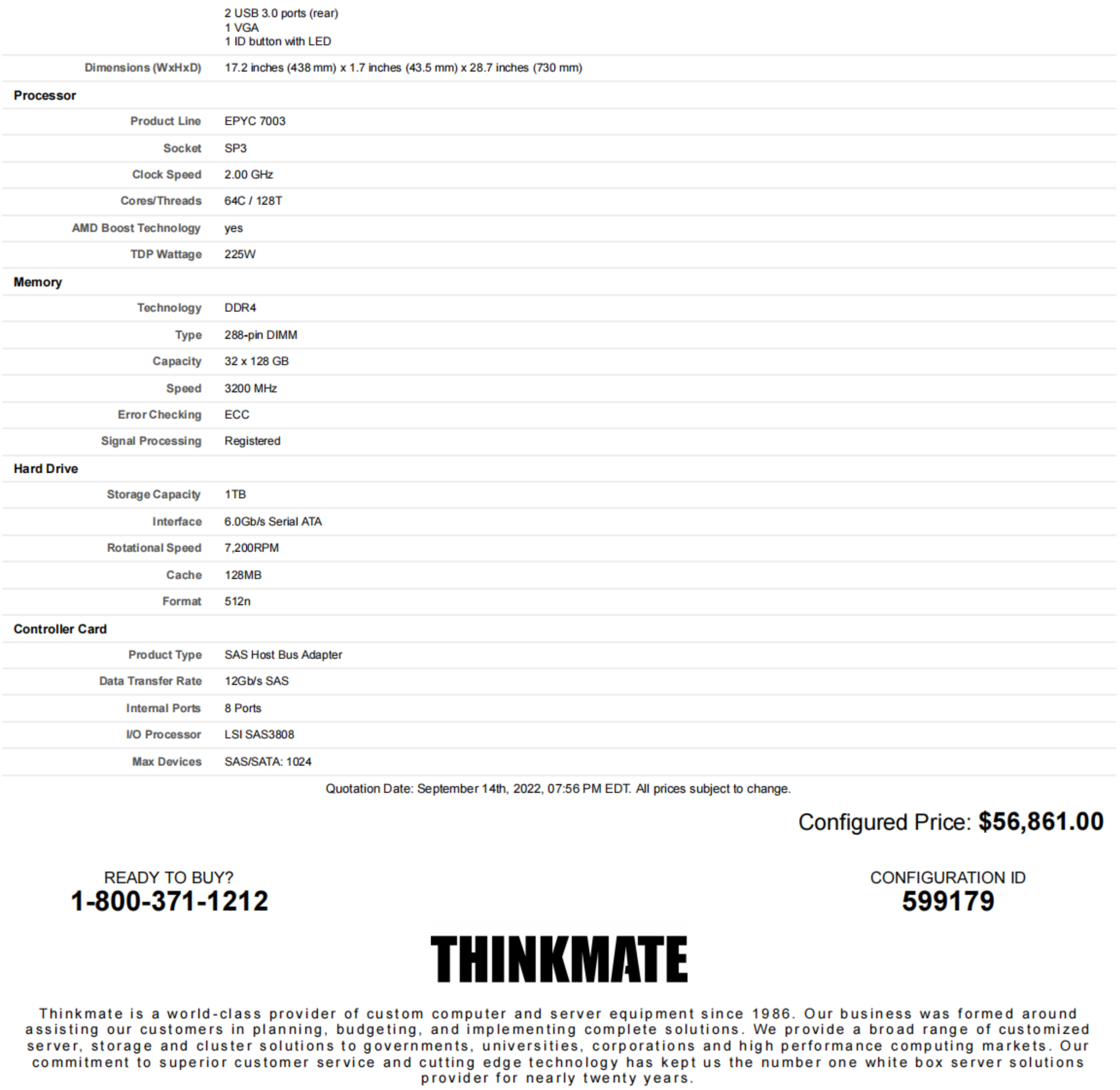
Selection Summary

Barebone	AMD EPYC™ 7003 Series - 1U - 4x 3.5" SATA/SAS3 - 1x M.2 NVMe - Dual Intel 1-Gigabit Ethernet (RJ45) - 1200W Redundant
Processor	2 x AMD EPYC™ 7713 Processor 64-core 2.00GHz 256MB Cache (225W)
Memory	32 x 128GB PC4-25600 3200MHz DDR4 ECC RDIMM
Hard Drive	1TB SATA 6.0Gb/s 7200RPM - 3.5" - Ultrastar™ DC HA210 (512n)
Controller Card	Broadcom HBA 9500-8i SAS3/SATA 8-Port Tri-Mode Host Bus Adapter - PCIe 4.0 x8
Trusted Platform Module	Trusted Platform Module - TPM 2.0
Cables	2 x AC Power Cord (North America), C13, NEMA 5-15P, 2.1m CAB-AC
Mounting Rails	Thinkmate Standard Rail Kit for 1U/2U Servers (Square Hole) (Included)
Operating System	Ubuntu Linux 22.04 LTS Server Edition (64-bit)
Warranty	Thinkmate® 5 Year Advanced Parts Replacement Warranty (Zone 0)

Tech Specs

Barebone

Memory Technology	DDR4 ECC Reg
Chipset	System on Chip
Form Factor	1U
Color	Black
Memory Slots	32x 288-pin DIMM Sockets
Graphics	Aspeed AST2500 BMC
Ethernet	Dual-Port Intel i350 Gigabit Ethernet LAN Dedicated Management LAN port
Power	Redundant 1200W AC-DC Power Supply
External Bays	4 x 3.5" / 2.5" SATA/SAS hot-swappable HDD/SSD bays
M.2	1 M.2 PCIe 3.0 x4 Form Factor: 2242/2260/2280/22110 Key: M-Key
Expansion Slots	Riser Card CRS101E: - 1 x PCIe x16 slot (Gen4 x16), FHHL Riser Card CRS101E: - 1 x PCIe x16 slot (Gen4 x16), FHHL 1 x OCP 3.0 mezzanine slot with PCIe Gen4 x16 bandwidth from CPU_0 1 x OCP 2.0 mezzanine slot with PCIe Gen3 x8 bandwidth (Type1, P1, P2)
Front Panel	Power button with LED Reset button ID button with LED HDD LED LAN LEDs NMI button 2 USB 3.0 System status LED
Back Panel	2 RJ45 Gigabit Ethernet LAN ports 1 RJ45 Management LAN port



	2 USB 3.0 ports (rear) 1 VGA 1 ID button with LED
Dimensions (WxHxD)	17.2 inches (438 mm) x 1.7 inches (43.5 mm) x 28.7 inches (730 mm)
Processor	
Product Line	EPYC 7003
Socket	SP3
Clock Speed	2.00 GHz
Cores/Threads	64C / 128T
AMD Boost Technology	yes
TDP Wattage	225W
Memory	
Technology	DDR4
Type	288-pin DIMM
Capacity	32 x 128 GB
Speed	3200 MHz
Error Checking	ECC
Signal Processing	Registered
Hard Drive	
Storage Capacity	1TB
Interface	6.0Gb/s Serial ATA
Rotational Speed	7,200RPM
Cache	128MB
Format	512n
Controller Card	
Product Type	SAS Host Bus Adapter
Data Transfer Rate	12Gb/s SAS
Internal Ports	8 Ports
I/O Processor	LSI SAS3808
Max Devices	SAS/SATA: 1024

Quotation Date: September 14th, 2022, 07:56 PM EDT. All prices subject to change.

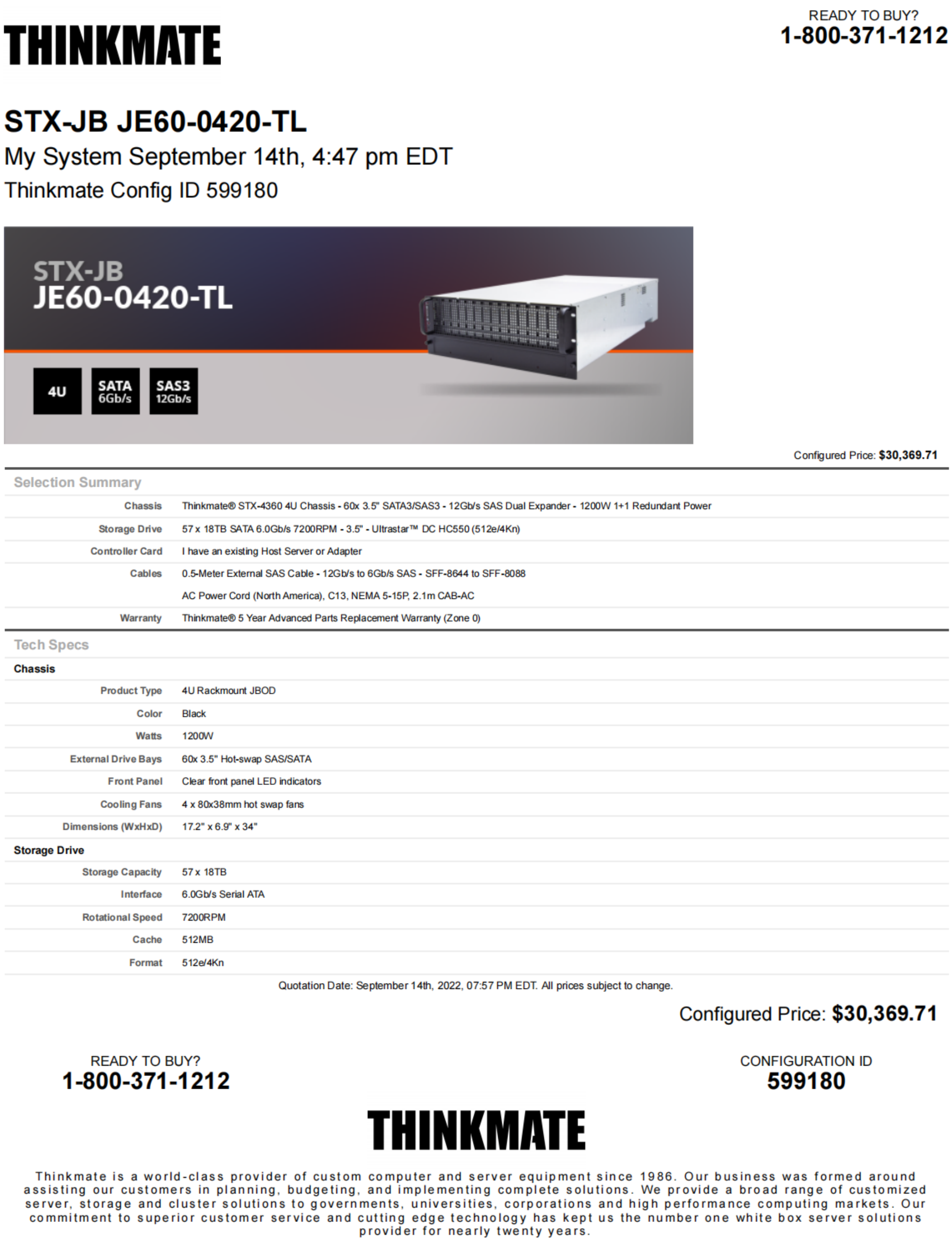
Configured Price: **$56,861.00**

READY TO BUY?
1-800-371-1212

CONFIGURATION ID
599179

THINKMATE

Thinkmate is a world-class provider of custom computer and server equipment since 1986. Our business was formed around assisting our customers in planning, budgeting, and implementing complete solutions. We provide a broad range of customized server, storage and cluster solutions to governments, universities, corporations and high performance computing markets. Our commitment to superior customer service and cutting edge technology has kept us the number one white box server solutions provider for nearly twenty years.



THINKMATE

READY TO BUY?
1-800-371-1212

STX-JB JE60-0420-TL

My System September 14th, 4:47 pm EDT

Thinkmate Config ID 599180

STX-JB
JE60-0420-TL
4U
SATA 6Gb/s
SAS3 12Gb/s

Configured Price: **$30,369.71**

Selection Summary	
Chassis	Thinkmate® STX-4360 4U Chassis - 60x 3.5" SATA3/SAS3 - 12Gb/s SAS Dual Expander - 1200W 1+1 Redundant Power
Storage Drive	57 x 18TB SATA 6.0Gb/s 7200RPM - 3.5" - Ultrastar™ DC HC550 (512e/4Kn)
Controller Card	I have an existing Host Server or Adapter
Cables	0.5-Meter External SAS Cable - 12Gb/s to 6Gb/s SAS - SFF-8644 to SFF-8088
	AC Power Cord (North America), C13, NEMA 5-15P, 2.1m CAB-AC
Warranty	Thinkmate® 5 Year Advanced Parts Replacement Warranty (Zone 0)

Tech Specs	
Chassis	
Product Type	4U Rackmount JBOD
Color	Black
Watts	1200W
External Drive Bays	60x 3.5" Hot-swap SAS/SATA
Front Panel	Clear front panel LED indicators
Cooling Fans	4 x 80x38mm hot swap fans
Dimensions (WxHxD)	17.2" x 6.9" x 34"
Storage Drive	
Storage Capacity	57 x 18TB
Interface	6.0Gb/s Serial ATA
Rotational Speed	7200RPM
Cache	512MB
Format	512e/4Kn

Quotation Date: September 14th, 2022, 07:57 PM EDT. All prices subject to change.

Configured Price: **$30,369.71**

READY TO BUY?
1-800-371-1212

CONFIGURATION ID
599180

THINKMATE

Thinkmate is a world-class provider of custom computer and server equipment since 1986. Our business was formed around assisting our customers in planning, budgeting, and implementing complete solutions. We provide a broad range of customized server, storage and cluster solutions to governments, universities, corporations and high performance computing markets. Our commitment to superior customer service and cutting edge technology has kept us the number one white box server solutions provider for nearly twenty years.