Assignment-2

1)

a) In next 5 years we might see processors boasting 256 cores/512 thread on single die.

b) The processor building will not change much in future. There are talks and claims of graphene as a suitable replacement but they are only in theory and not yet in any experimental stage. There are also talks about Quantum computing which can revolutionize the computation but still much progress needs to be made before they can replace the current processors. Unless we have a major technological breakthrough, there would only be some advancement in transistor technology.

c) The biggest challenge would be finding and developing new material instead of silicon for designing processor.

a) Video Editing, rendering, simulation, video games.

b) CPU is the brain of the PC. It's essential for handling multiple tasks of a PC. It handles all the OS and execution operation of the PC as whole. They have big cores which operate at high speed. Graphical Processing Unit(GPU) used for specialized tasks in combination with CPU. While GPU has many cores they are small and not as fast as CPU cores. They are used to accelerate the execution speed such as rendering, gaming, crypto mining and Artificial Intelligence.

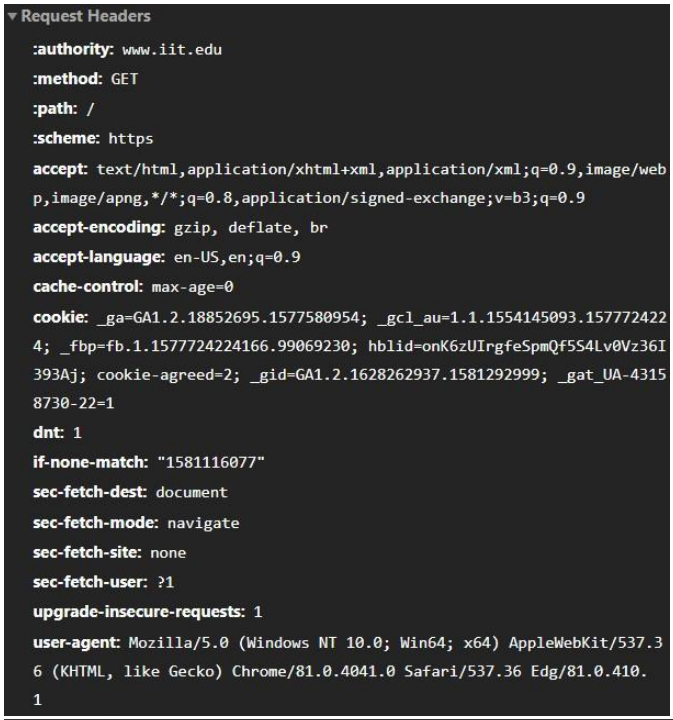
2)

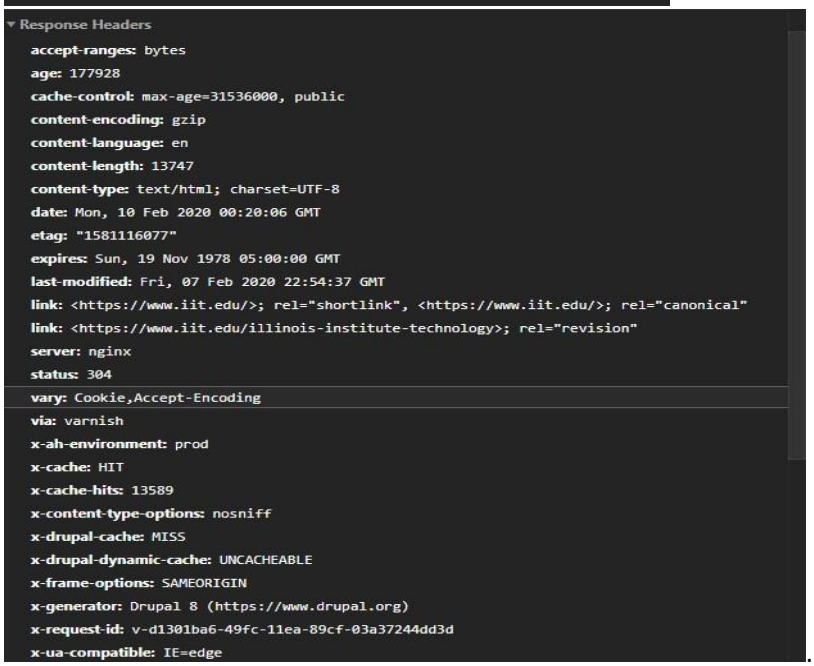
b) A thread has its own registers, program counter and a stack.

g) POSIX thread(Pthread) is useful for doing specialized process alongside the main process and it can be written only in C. OpenMP is used to speed up the process with all available resources such as cores. It can be implemented in C,C++ and FORTRAN.

3) When we enter "iit.edu" and press enter following process happens:
The DNS finds and resolves IP address of website. The browser cache is checked to see if there is any record of website for faster lookup. Then the DNS query runs into the OS cache If there's no record of it in cache, then the requests goes to ISP and the ISP resolves the query. If the ISP can't resolve it then the request is forwarded to the domain the site belongs under. In this our website ends with ".edu" so it belongs to Top Level Domain(TLD). ISP asks TLD to give IP address of the website. TLD forwards the resolve to Authoritative Name Server. ANS holds the DNS records for a particular address. ISP gets the IP address from ANS and then sends it back to the browser.
When the IP address is located, a TCP/IP three-way handshake connection is initiated:

- Client sends a SYN packet to server asking if it is open for new connection
- If there are open ports then server will send ACK packet to client
- When client receives the ACK packet, it will send an ACK packet informing that it has received it. Hence, a TCP connection is established
  The browser then sends HTTP/HTTPS request and server sends back a response





- The IP protocol at Network layer responsible for host-to-host connection works and makes a connection between the end systems.
- Then we have ARP protocol which is used to resolve the MAC address from the IP address to recognize the host.
- Then the data ink layer which is responsible for hop to hop delivery is used.
- Further this the data travels in the physical layer using an transmission medium
- If everything is fine, then the browser receives and goes through HTML code and renders HTML along with other scripts such as CSS and JS. In the end, website is displayed.

4)

a) Power consumption is critical to data center as it can cost anywhere between 30-40% of the total cost of data center operation. This is a huge portion and if reduced can result in huge savings and more competitive prices which will give the organization an edge over their competitors.
b) Dynamic Voltage Frequency Scaling(DVFS) technique is used to increase or decrease the voltage or frequency or both in a microprocessor based on the requirement at hand. This can be used to perform overclocking in case there is requirement for high processing power than normal or to reduce the power consumption to save on energy or in case there is low battery in case of laptops
c) Two things can that be used to reduce the cooling costs are:
    i) Use multiple sensors to monitor the temperature of various areas in a data center. Based on the measurements, it can be decided how much cooling should each area receive instead of having a uniform cooling all across the center. This is due to the fact that the center doesn't have uniform hating and some areas are more hot that others.
    ii) Use power maintenance tools for servers – they can help in making sure that the server is consuming is lesser energy in case of low loads or inactivity which will result in less heating
d) Install in rack or in-row cooling – these systems bring cold air closer to the servers and can result in 3 times less energy consumption than conventional cooling systems

5)

a) Sub-millisecond latency on HDD is not possible. This is because of the design and technical limitations of HDD. HDD can perform will only up to a certain point. HDD has moving mechanical parts and disks which increases the chances of failure, the faster It gets. For example – one of the biggest issue is that above 15k rpm speeds, the platter starts to stretch which will result in disk failure and we only get 2ms latency at 15k speeds. To even reach 1ms, a speed of 30k rpm will be required.
b) SSD doesn't have moving mechanical parts like HDD. All the information is stored in chips. SSD use NAND flash memory and is non-volatile which means the stored data is retained even when a system is turned off. Instead of a mechanical head, SSD has a "controller" which is responsible for performing SSD operations. This all results in very low latency and in applications such as bank transaction systems or short trading, we need as low latency as possible.
c) SSD are useful for performing operations and running programs which will constantly access the memory. SSD have a superior read and write speeds compared to HDD which means faster I/O operations. Enable faster compression and decompression of data. Faster copying and backup capability to RAID drives.
d) This doesn't seem likely possible as the best PCI currently available is PCI Express 5.0 which has a max capacity of 63GB/s and this is still not yet in production and will still not reach 1Terabit or 125GB/sec speed. Thus, the claim is not true

e) Single Threading:

Cache hit is 8 sec and Cache miss takes 24 msec. The mean request takes (¾ *8) + (¼ *24) = 12 msec.

Therefore, server request takes 4 msec.

Multithreading:

Using round-robin scheduling, since there are 4 threads the requests per second is $(1-(4/12)^4)*1000/8 = 123.45$ req/sec

6)

a) MYSQL Solution

For SQl – power consumption = 6115 = $0.91/hr

Cooling will be same = $0.91/hr

Total power cost = $1.82/hr = $79,716 for 5 years

Hardware cost = $2,96,000

Administration cost = $20,000/year = 1,00,000 for 5 years

Total cost for 5 years = $4,75,716

For Spark – power consumption = 2527 = $0.38/hr

Cooling will be same = $0.38/hr

Total power cost = $0.76/hr = $33,288 for 5 years

Hardware cost = $2,22,000

Administration cost = $20,000/year = 1,00,000 for 5 years

Total cost for 5 years = $3,56,014