Top computer architecture questions

1. What is computer architecture?

Answer:

Example: "Computer architecture refers to hardware instructions, software standards and technology infrastructure that define how computer platforms, systems and programs operate. This means that computer architecture outlines the system's functionality, design and compatibility."

2. What are the three categories of computer architecture?

Answer:

Example: "Computer architecture has three categories. These include:

System design

This includes all the hardware components in the system such as the data processors, direct memory access and graphics processing unit. It also includes data paths, memory controllers and miscellaneous things such as virtualization and multiprocessing.

Instruction Set Architecture

This is a part of the central processing unit that is visible to the compiler writer and programmer. It defines the CPU's capabilities and functions based on what programming it can process and perform. This includes the data formats, memory addressing modes, processor register types, word size and the instruction set those programmers use.

Microarchitecture

Also known as "computer organization," this kind of architecture defines storage elements, data processing and data paths, as well as how they should be implemented in the ISA."

3. What are some of the components of a microprocessor?

Answer:

Example: "Some of the components of a microprocessor include the arithmetic and logic unit, which performs math computations such as division, addition and subtraction and Boolean functions; registers, which act as the temporary data holding places of microprocessors; control units, which receive signals from the CPU and move data from one microprocessor to another; and memory caches, which accelerate the computing process, as the CPU doesn't have to use the slower RAM to retrieve data."

4. What is MESI?

Answer:

Example: "MESI stands for the four states of the cache blocks, which are Modified, Exclusive, Shared and Invalid. It's also known as the "Illinois protocol". It's used to maintain cache coherency in hierarchical memory systems. MESI is the most common protocol that supports write-back cache. Its use in personal computers became common with the introduction of Intel's Pentium processor."

5. What are the different hazards?

Answer:

Example: "Hazards have three classes. These include the structural hazards, which occur from resource conflicts when the hardware can't support all possible combinations of instructions in synchronized overlapped execution; data hazards, which occur when instructions that manifest data dependence change data in different stages of a pipeline; and control hazards, which occur from the pipelining of branches and other instructions that modify the PC."

6. What is pipelining?

Answer:

Example: "Pipelining, also known as "pipeline processing", is the process of collecting instruction from the processor through a pipeline. It stores and executes instructions in an orderly process."

7. What is a cache?

Answer:

Example: "A cache is a small amount of memory, which is a part of the CPU. It's placed closer to the CPU than the RAM. It temporarily holds data and instructions that the CPU is likely to reuse."

8. What is a snooping protocol?

Answer:

Example: "A snooping protocol, also referred to as a "bus-snooping protocol," maintains cache coherency in symmetric multiprocessing environments. All caches on the bus snoop or monitor the bus to determine if they have a copy of the block of data that is requested on the bus. Each cache holds a copy of the sharing status of every block of physical memory it has. Typically, several copies of a file in a multiprocessing environment can be read without any problem of coherence. However, a processor should have exclusive access to the bus to write."

9. What are the different types of interrupts in a microprocessor system?

Answer:

Example: "Interrupts can either be internal or external. Internal interrupts, which are also referred to as "software interrupts", are caused by software instruction and operate similar to a branch or jump instruction. An external interrupt, which is also referred to as a "hardware interrupt," is caused by an external hardware module."

10. What is the easiest way to determine cache locations in which to store memory blocks?

Answer:

Example: "Direct mapping is the easiest way to define cache locations in which to store memory blocks. It maps each block of the main memory into only one possible cache line. The cache in a direct-mapped cache structure is organized into several sets, with a single line per set. Based on the memory block's address, it can only use a single cache line. The cache can be framed as a column matrix."

11. What is a virtual memory on a computer?

Answer:

Example: "A virtual memory is an operating systems' memory management feature that uses software and hardware to allow computers to compensate for the shortages of physical memory by temporarily moving data from RAM to disk storage."

12. Can you state some of the common rules of assembly language?

Answer:

Example: "Some of the common rules of assembly language include the following:

In assembly language, the label field can either be empty or may define a symbolic address.

Instruction fields can specify machine pseudo instructions.

Comment fields can be commented with or left empty.

In the case of symbolic addresses, up to four characters are only allowed.

The comment field begins with a forward slash while the symbolic addresses field is terminated by a comma."

13. What is the RAID system?

Answer:

Example: "RAID, which stands for Redundant Array of Independent Disks, refers to the hard drives connected and set up in ways to help accelerate or protect the performance of a computer's disk storage. It is typically used on servers and high-performance computers."

14. What are the two hardware methods to establish a priority? Explain each method.

Answer:

Example: "The two different ways to establish hardware priority are the parallel priority and daisy-chaining. Daisy-chaining is a method that involves connecting all the devices that can request an interrupt in a serial manner. This setting is governed by the priority of the devices, in which the device with the highest priority is placed first.

Parallel priority, on the other hand, uses a register for which bits are configured separately by the interrupt signal from each device. It may also come with a mask register, which is used to control the status of each interrupt request."

15. What are flip-flops?

Answer:

Example: "Flip-flops, also called "latches", are electronic circuits that have two stable states used to store binary data. The data stored in the states can be modified by using varying inputs. Flip-flops are fundamental components of digital electronic systems used in communications, computers and many other kinds of systems."

16. What's the difference between interrupt service routine and subroutine?

Answer:

Example: "Subroutine is a part of code within a larger program, which performs a specific task and is relatively independent of the remaining code. Interrupt service routines deal with hardware interrupts. They are not

independent threads, but more like signals. They are used if an interrupt suspends any thread. Unlike subroutine, which runs when we call it, ISR runs whenever there's a signal from either the software or hardware. The big difference is we can determine where the subroutine runs while we can't determine when the ISR will be executed."

17. What are the different types of fields that are part of instruction?

Answer:

Example: "An instruction is like a command to a computer to perform a particular operation. The instruction format is composed of various fields in them such as:

Operation code field. Also called the "op-code field", this field is used to specify the operation to be performed for the instruction.

Address field. As the term implies, this field is used to designate the various addresses, such as memory address and register address.

Mode field. This field specifies as to how an operand performs or how effective an address is."

18. What are the steps involved in an instruction cycle?

Example: "A program that resides in the memory contains a set of instructions that the computer needs to perform sequentially. The cycle for every instruction is called the instruction cycle, which consists of the following steps:

Fetch instruction. The CPU fetches the instruction from the memory. The computer gets loaded with the address of the instruction.

Decode. This allows the CPU to determine what instruction must be performed and how many operands are needed to fetch to perform an instruction.

Execute. At this step, the instruction is performed. If the instruction has logic or arithmetic, the ALU is utilized. This is the only step of the instruction cycle that's useful from the end user's perspective."

19. What are the five stages in a DLX pipeline?

Answer:

Example: "Each DLX instruction has five stages. These include:

Instruction fetch

Instruction decode and register fetch

Execution

Memory access

Writeback"

20. What are the types of micro-operations?

Answer:

Example: "Micro-operations are executed on data stored in registers. They are basic math operations performed on the information stored in one or more registers. The types of micro-operations are:

Shift micro-operations: They perform shift operations on data stored in registers.

Logic micro-operations: They execute bit manipulation operations on nonnumerical data saved in registers.

Arithmetic micro-operations: They perform arithmetic operations, such as subtractions and additions, on digital data stored in registers.

Register transfer micro-operations: They transfer binary information between registers."

21. What is the write-through method?

Answer:

Example: "Write-through is the preferred method of data storage in many applications, especially in banking and medical device control, as it's good at preventing data loss. In less critical applications, and especially when the volume of data is large, an alternative method known as "write-back" speeds up system performance because updates are typically written exclusively to the cache and are saved in the main memory only under certain conditions or at specified intervals."

22. What is associative mapping?

Answer:

Example: "The associative mapping technique uses several mapping functions to transfer data from the main memory to the cache memory. This means that any main memory is mapped into any line of the cache. As a result, the cache memory address is not in use. The associative cache controller processes and interprets the request by utilizing the main memory address format."

23. What does wait state mean?

Answer:

Example: "A wait state means that the computer processor experiences a delay when accessing a device or an external memory that is slow in its response. Wait states are considered wasteful in processor performance, which is why modern-day designs try to either minimize or eliminate wait states. These include pipelines, instruction pre-fetch and pipelines, caches, branch prediction and simultaneous multithreading. While these techniques can't eliminate wait states, they can significantly minimize the problem when they work together."

24. What is DMA?

Answer:

Example: "DMA, which stands for Direct Memory Access, is a feature of computer systems that allows an input/output device to receive or send data directly from or to the main memory, bypassing the CPU to boost memory operations. The process is performed by a chip known as the DMA controller."

25. What is a horizontal microcode?

Answer:

Example: "Horizontal microcode, which is usually contained in a fairly wide control store, comes with several discrete micro-operations that are combined into one micro-instruction for simultaneous operation."

26. Show basic components used by a Microprocessor? Explain.

Answer:

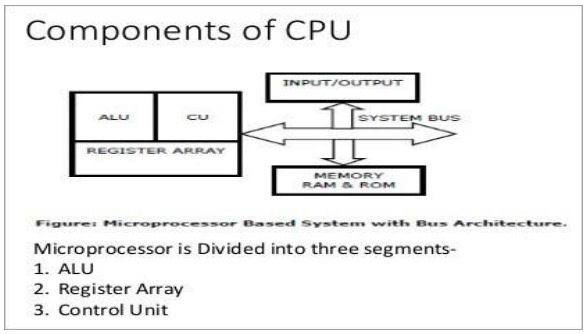
Microprocessor normally uses three basic components:

Address lines are one of the major elements of a Microprocessor as it is important for referring to the proper address of a single block.

Data lines are the elements that maintain the main criteria for transferring data for a microprocessor.

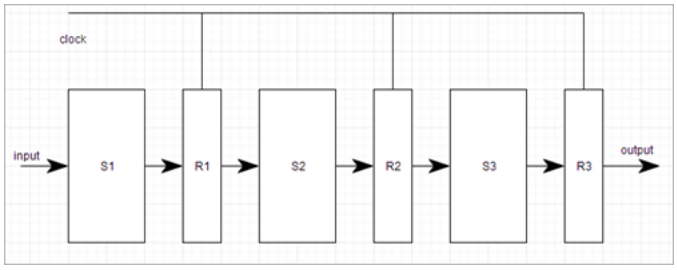
The target of the processing data comes after the completion of addressing and data transfer. IC chips are vital for data processing in a microchip.

27. What are the common Components of a Microprocessor?



Answer: Control units, I/O units, Cache, ALU, and Registers are some of the common components of a Microprocessor.

28. Are you aware of Pipelining?

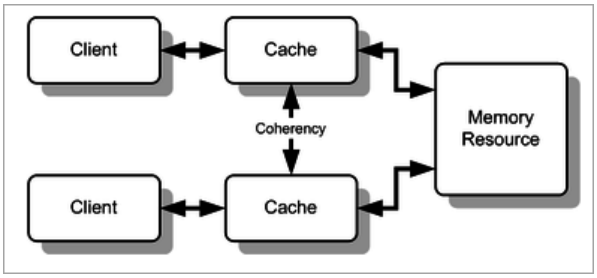


Answer: Pipelining is one of the most popular techniques used by an advanced microprocessor that is mainly used when multiple instructions enter into the system. It accumulates instruction from the processor via a pipeline and allows storing and execution of instructions in an orderly process.

The process is divided into stages and each of them is connected in a pipe-like structure. It is used where multiple instructions are overlapped during execution.

Like in a car manufacturing company, each setup of huge assembly lines, and robotic arms perform certain tasks. After one task is completed, the car moves on ahead to the next arm.

29. What is Cache Coherence?



Answer: The consistency or regularity of data stored in the cache memory is called cache coherence. It is imperative for Distributed Shared Memory (DSM) or multiprocessor systems to maintain cache and memory consistency.

Cache management is structured to see that the data is not lost or overwritten. You can use different techniques to maintain cache coherency, and that includes snarfing, snooping and directory-based coherence.

A DSM system uses a coherency protocol by imitating these techniques for maintaining consistency and it is essential to system operations. Cache coherence requires two things i.e. write propagation and transaction serialization.

In any cache, the changes to the data must be disseminated to other copies of that cache line in the peer caches. This is what the write propagation does. The job of transaction serialization is to make sure that anything which is read or written to a single memory location is seen by all the processors in the same order.

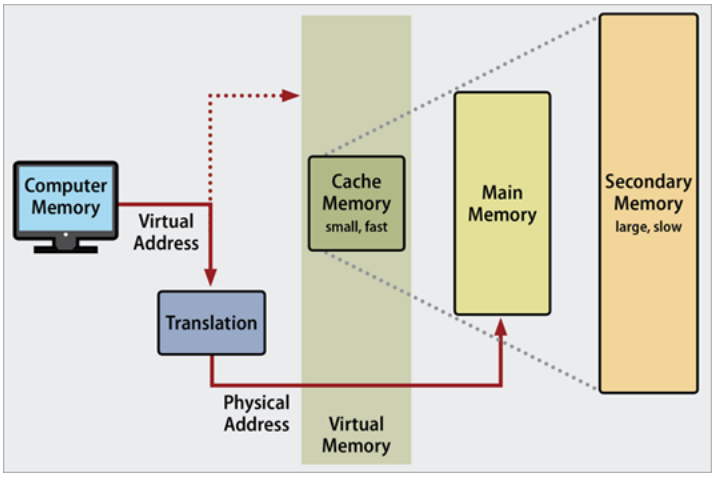
30. What is Cache miss?

Answer: Sometimes, there is a failed attempt to write or read a part of the data in the cache. This miss results in longer latency in the main memory. There are three types of cache miss i.e. cold or compulsory, capacity and conflict miss.

Cold or compulsory miss starts with an empty cache and is the foremost reference to a memory block. You can refer to it as an empty hotel where the first guest hasn't arrived yet. A capacity miss occurs when the cache doesn't have enough space to hold all the blocks that you want to use. It is like a hotel where you want to stay but has no vacancy.

Conflict miss happens when the same location gets two blocks but doesn't have enough space for both of them. In an easy example, it is like you are supposed to stay on the third floor of a hotel but all the rooms on the floor are occupied and there is no room for you.

31. What is Virtual Memory?



Answer: Your computer uses memory to load the OS & run the programs and the amount of real memory, i.e. RAM, is finite. Thus, there are chances for you to run out of memory, especially when you are running too many programs at one time.

That's where virtual memory comes in handy. It increases the memory available in your computer by enlarging the "address space" i.e. places in memory where you can store data. It uses the hard disk space for allocating additional memory.

However, the hard drive is slower when compared to RAM, thus, you must map the data stored in virtual memory back to the real memory to be used. Virtual memory enables your computer to run more programs than it can.

32. What are the 5 stages of the DLX pipeline?

Answer: DLX is a RISC processor architecture. It was designed by David A. Patterson and John L. Hennessy. Its architecture was chosen based on the observations of the most frequently used primitives in programs.

Its 5 stages include:

CPU Operand Storage

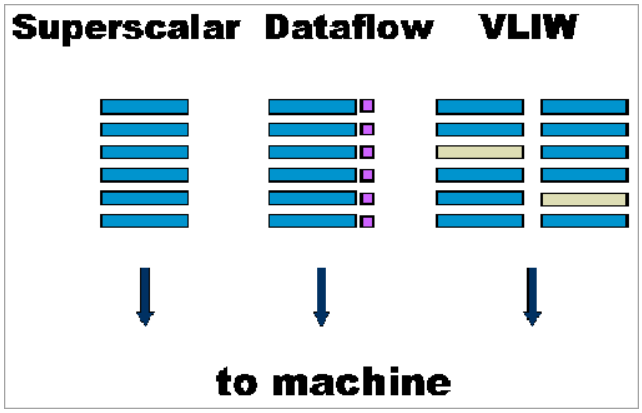
Explicit operands

Operation

Location

Type and size of operands

33. Explain Superscalar machines and VLIW machines.

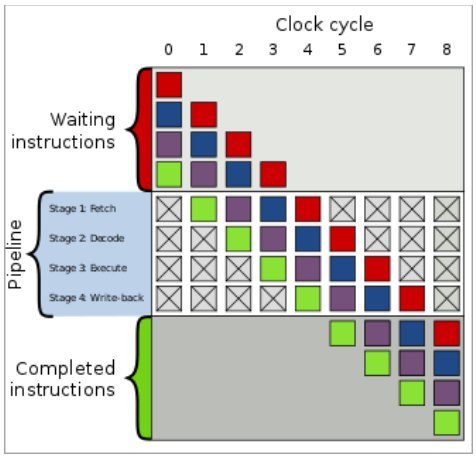


Answers: Superscalar processor is a CPU that implements instruction-level parallelism within a single processor. It can execute more than one instruction during a clock cycle. It simultaneously dispatches multiple instructions to different execution units on the processor.

Thus, it allows for more throughput when compared to others to be possible at a given clock rate.

VLIW or Very Long Instruction Word refers to a CPU architecture that is designed to take advantage of ILP or instruction-level parallelism but with minimum hardware complexities. The VLIW approach executes the operation in parallel which is based on a fixed schedule that is determined when programs are compiled.

34. What is Branch Prediction and how can it control Hazards?



Answer: In a unit of information processing that processes a pipeline, a branch prediction control device generates an address for branch prediction. This address is used to verify the instructions that are being executed speculatively.

The device has a first return address storage unit that stores the return address for prediction. Then, there is a storage unit for the second return address that stores a return address that is generated based on the result of an execution of the call instruction.

There is also a storage unit for a branch prediction address that sends a stored prediction return address as a branch prediction address and stores the branch prediction addresses that are sent.

When the return address is generated after the execution of a branch instruction that differs from the branch prediction address, then the contents that are stored in the storage unit for the second return address are duplicated to the storage unit for the first return address.

35. Can you calculate the number of sets given with its size and way in a cache?

Answers: In the hierarchy of primary storage, a cache carries cache lines collected into sets. The cache can be called k-way associative if each set holds k lines. A data request possesses an address that specifies the position of the data requested.

You can place only one cache-line data of the size of the chunk from the lower level into one set. Its address decides the set in which it can be placed. The mapping between the sets and addresses must have a fast and easy implementation. For fast implementation, only a part of the address chooses the set.

After that, a request address is separated into three fragments as shown below:

A specific position within a cache line is identified by an offset part.

The set that has the requested data is identified by a set part.

There must be a saved tag part along with its data in each cache line to distinguish the different addresses that could be put in the set.

36. How do you find a Block in a Cache?

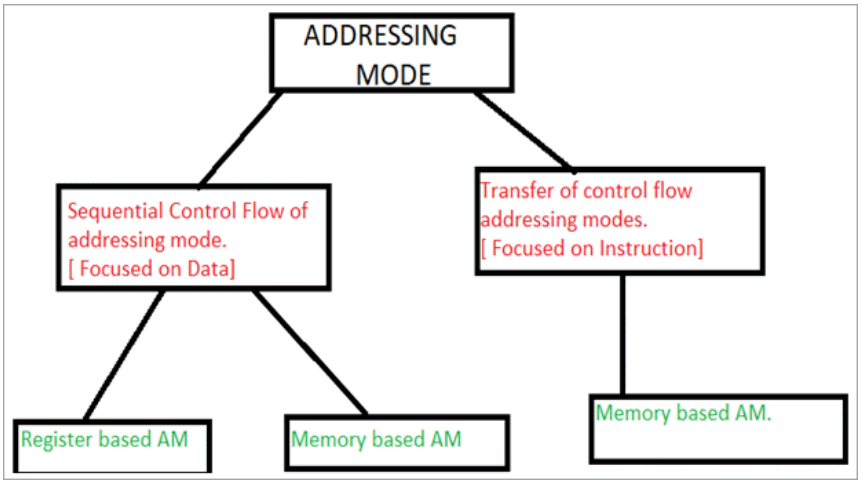
Answer: Block's tag is recorded by each place in the cache along with its data. The place in the cache might be unoccupied, so it usually maintains a valid bit.

Thus, to find the block in cache:

Determine the place or set of places used in the index of block address.

Check if a valid bit is set for each place and compare the tag with that address block parallelly for all places in a set.

37. What is an Addressing Mode?



Answer: In the most central processing unit designs, there is a characteristic of the instruction set architecture called addressing modes.

The diverse addressing modes are explained in a given instruction set architecture and these modes define how ML instructions in the given architecture recognize the operands of every instruction.

Addressing modes specify the way to compute the effectual memory address of an operand with the use of the information kept in registers or/and constants held within an ML instruction or somewhere else.

38. What is Aliasing.

Answer: Aliasing, in the world of computing, describes a circumstance where you can access the location of data in memory through separate symbolic names in the program. Thus, by changing the data through one name, you can implicitly modify the values correlated with every aliased name.

That's something the programmer might not have anticipated. Hence, the programs become difficult to optimize, understand and analyze.

39. What is the difference between Software and Hardware interrupts?

Answer:

Software Interrupts	Hardware Interrupts
These can be invoked with the help of INT instruction.	These are caused by external devices, especially hardware failure.
It is synchronous.	It is asynchronous.
It is caused by any internal system of the computer.	It happens when the signal for the processor is from an external device or hardware.
This is often the result of either an exceptional condition in the processor or special instruction in the instruction set.	It is the result of outside interference, be it from peripherals, users, through a network, or other hardware devices.
PC incremented.	PC isn't incremented.
It has the highest priority.	It has the lowest priority.

40. You want to do other tasks, but the CPU is busy. Suggest a solution.

Answer: I will create an interrupt that is non-maskable and then give the jump instruction to the essential subroutine.

41. What do you know about Latches? What are the various types of Latches?

Answer: Latch, also known as a bistable-multivibrator due to its two stable states of active high and active low, is a type of logic circuit. Through a feedback lane, it holds the data, thereby acting as a storage device.

As long as the apparatus stays active, the latch can store a 1-bit of data. The latch can instantly change the data stored once enable is declared.

Types of Latches:

SR or set/reset latch, the asynchronous apparatus, works independently to control signals. It is done depending on the set-state and reset-input.

Gates SR Latch is the latch that carries the third input. This input must be active for set/reset inputs to work.

D latch or the data latch removes the chance of undesirable conditions of input.

Gated D latch is designed by making some changes in the gated SR latch. The change made is that the reset input must be changed to the inverter set.

JK latch is similar to RS latch. It comprises two inputs i.e. J and K. When the inputs of JK latch are high, the output is bound to toggle.

T latch is formed when the inputs of JK latch are shorted. T latch toggles the output when the latch's input is high.

42. explain something about Flip Flops.

Answer: Just like latch, a flip flop is an electronic circuit. It carries two stable states that can store binary data. By applying various inputs, you can change the stored data. Like latches, it is the building block of electronic and digital systems of computers, in communication and many other systems.

43. Explain the differences between Latches and Flip-flops.

Answer:

Latches	Flip-Flops
These building blocks can be built from logic gates.	While latches are used to build these building blocks.
It checks the inputs continuously and changes the output accordingly.	Flip-flop does the same thing but only at the time set by the clocking signal.
Latches are sensitive to the pulse duration and when the switch is turned on, it can receive and send the data.	It is sensitive to the change in the signal. The transfer of data can take place only at a single instant. You can't change the data until the signal changes next. These are used as registers.
Enable function input is what it works on.	It works on clock pulses.

44. What do you know about the Real-time Operating System?

Answer: Also known as a data processing system, the real-time operating system requires an extremely small-time interval for processing and responding to the inputs. The time it takes to respond and display the required updated information is called response time.

We use real-time when the time requirements for operating a processor or for the flow of data are rigid. In a dedicated application, we can use the real-time system as a control device. This system must have definitive and fixed time constraints, else it will feel.

45. Difference between Write-back and Write-through Cache.

Answer:

Write Back Cache	Write Through Cache
Write back cache differs the write until that cache line has been used for read. This, in turn, puts a question mark on its integrity, especially when many processors access the same data employing its internal cache.	The write through caches flushes for each writes hence is considered better in integrity.
It saves many write or memory write cycle, hence giving a good performance.	Compared to write back cache, it doesn't give such a good performance.

46. Define Multiprocessing.

47. What is meant by instruction?

48. What is Bus? Draw the single bus structure.

49. Define Pipeline processing.

50.Draw the basic functional units of a computer.

51. Briefly explain Primary storage and secondary storage.

52. What is register?

53. Define RAM.

54. Give short notes on system software.

55. Write down the operation of control unit?

56. Define Memory address register.

57. What is stack & queue?

58. Define Addressing modes.

59. Write the basic performance equation?

60. Define clock rate.

61. List out the various addressing techniques.

62. Draw the flow of Instruction cycle.

63. Suggest about Program counter.

64. List out the types in displacement addressing.

65. What is meant by stack addressing?

66. Define carry propagation delay.

67. Draw a diagram to implement manual multiplication algorithm.

68. Perform the 2's complement subtraction of smaller number (101011) from larger number (111001).

69. What are the basic operations performed by the processor?

70. Define Data path.

71. Define Processor clock.

72. Define Latency and throughput.

73. Discuss the principal operation of micro programmed control unit.

74. What are the differences between hardwired and micro programmed control units?

75. Define nano programming.

76. What is control store?

77. What are the advantages of multiple bus organization over a single bus organization?

78. Write control sequencing for the executing the instruction. Add R4,R5,R6.

79. What is nano control memory?

80. What is meant by hardwired control?

81. What are the types of micro instruction?

82. Name the methods for generating the control signals.

83. Explain different types of hazards that occur in a pipeline.

84. Explain various approaches used to deal with conditional branching.

85. Explain the basic concepts of pipelining and compare it with sequence processing with a neat diagram.

86. Explain instruction pipelining.

87. What is branch hazard? Describe the method for dealing with the branch hazard?

88. What is data hazard? Explain the methods for dealing with data hazard?