Review Questions: The Memory Hierarchy

Principle of Locality

1. Question 1: The Principle of Locality, also known as locality of reference, reflects which observation during program execution?

   (a) A. Memory references by the processor are uniformly distributed.

   (b) B. Memory references by the processor tend to cluster.

   (c) C. Memory references are unpredictable over time.

   (d) D. Memory references are independent of past access patterns.

   Correct answer is: B. Memory references by the processor tend to cluster. Because: The Principle of Locality states that during the execution of a program, memory references by the processor tend to cluster, meaning certain memory locations are referenced repeatedly within a short period.

2. Question 2: Which of the following is NOT one of the three assertions on which the Principle of Locality is based?

   (a) A. During any interval of time, a program references memory locations non-uniformly.

   (b) B. As a function of time, the probability that a given unit of memory is referenced tends to change slowly.

   (c) C. The correlation between immediate past and immediate future memory reference patterns is low.

   (d) D. The correlation between immediate past and immediate future memory reference patterns is high and tapers off as the time interval increases.

   Correct answer is: C. The correlation between immediate past and immediate future memory reference patterns is low. Because: One of the assertions of the Principle of Locality is that the correlation between immediate past and immediate future memory reference patterns is high and tapers off as the time interval increases, not low.

3. Question 3: What are the two forms of locality mentioned in the context of memory hierarchy?

   (a) A. Read locality and Write locality

   (b) B. CPU locality and I/O locality

   (c) C. Temporal locality and Spatial locality

   (d) D. Cache locality and Main memory locality

   Correct answer is: C. Temporal locality and Spatial locality. Because: The two forms of locality are temporal locality and spatial locality.

4. Question 4: Temporal locality refers to which tendency of a program?

   (a) A. To reference units of memory whose addresses are near one another in the near future.

   (b) B. To reference in the near future those units of memory referenced in the recent past.

   (c) C. To access data locations sequentially, such as when processing a table of data.

   (d) D. To reference memory locations randomly without any pattern.

   Correct answer is: B. To reference in the near future those units of memory referenced in the recent past. Because: Temporal locality refers to the tendency of a program to reference in the near future those units of memory referenced in the recent past.

5. Question 5: What constructs can lead to the principle of temporal locality?

   (a) A. Large arrays and linked lists.

   (b) B. Constants, temporary variables, and working stacks.

   (c) C. Network protocols and device drivers.

   (d) D. Graphics processing units and sound cards.

   Correct answer is: B. Constants, temporary variables, and working stacks. Because: Constants, temporary variables, and working stacks are constructs that lead to the principle of temporal locality.

6. Question 6: Spatial locality reflects which tendency of a program?

   (a) A. To reference the same memory location repeatedly over time.

   (b) B. To reference units of memory whose addresses are near one another.

   (c) C. To access memory locations in an unpredictable order.

   (d) D. To reference memory only from external storage devices.

   Correct answer is: B. To reference units of memory whose addresses are near one another. Because: Spatial locality refers to the tendency of a program to reference units of memory whose addresses are near one another. It also reflects the tendency to access data locations sequentially.

Characteristics of Memory Systems

7. Question 7: What does "Location" refer to regarding memory systems?

   (a) A. The physical address of a memory unit.

   (b) B. Whether memory is internal or external to the computer.

(c) C. The specific room where the memory modules are installed.

(d) D. The type of semiconductor used in the memory.

Correct answer is: B. Whether memory is internal or external to the computer. Because: "Location" refers to whether memory is internal (e.g., processor registers, cache, main memory) or external (e.g., optical disks, magnetic disks, tapes) to the computer.

8. Question 8: How is memory capacity typically expressed?

   (a) A. In terms of words.

   (b) B. In terms of bits per second.

   (c) C. In terms of bytes.

   (d) D. In terms of access time.

   Correct answer is: C. In terms of bytes. Because: Memory capacity is typically expressed in terms of bytes.

9. Question 9: For internal memory, what is the unit of transfer equal to?

   (a) A. The number of bytes that can be transferred in one second.

   (b) B. The number of electrical lines into and out of the memory module.

   (c) C. The total capacity of the memory module.

   (d) D. The number of memory locations accessed in a single operation.

   Correct answer is: B. The number of electrical lines into and out of the memory module. Because: For internal memory, the unit of transfer is equal to the number of electrical lines into and out of the memory module.

10. Question 10: Which access method organizes memory into units of data called records and requires access to be made in a specific linear sequence?

    (a) A. Random access

    (b) B. Direct access

    (c) C. Associative access

    (d) D. Sequential access

    Correct answer is: D. Sequential access. Because: In sequential access, memory is organized into units of data called records, and access must be made in a specific linear sequence.

11. Question 11: In which access method does each addressable location in memory have a unique, physically wired-in addressing mechanism, and the time to access a given location is independent of prior accesses?

    (a) A. Sequential access

(b) B. Direct access

(c) C. Random access

(d) D. Associative access

Correct answer is: C. Random access. Because: Random access means each addressable location has a unique, physically wired-in addressing mechanism, and access time is constant regardless of prior access sequences. Main memory and some cache systems use this.

12. Question 12: Which access method involves a shared read-write mechanism and individual blocks having a unique address based on physical location, resulting in variable access time?

    (a) A. Sequential access

    (b) B. Direct access

    (c) C. Random access

    (d) D. Associative access

    Correct answer is: B. Direct access. Because: Direct access involves a shared read-write mechanism where individual blocks or records have a unique address based on physical location, and access time is variable.

13. Question 13: In which access method is a word retrieved based on a portion of its contents rather than its address, and retrieval time is constant?

    (a) A. Sequential access

    (b) B. Direct access

    (c) C. Random access

    (d) D. Associative access

    Correct answer is: D. Associative access. Because: Associative access retrieves a word based on a portion of its contents, not its address, and each location has its own addressing mechanism, leading to constant retrieval time. Cache memories may employ this method.

14. Question 14: What are the two most important characteristics of memory in terms of capacity and performance?

    (a) A. Cost and Volatility.

    (b) B. Access time and Transfer rate.

    (c) C. Capacity and Performance.

    (d) D. Organization and Physical Type.

    Correct answer is: C. Capacity and Performance. Because: Capacity and performance are identified as the two most important characteristics of memory.

15. Question 15: For random-access memory, what is "Access time (latency)" defined as?

    (a) A. The time it takes to transfer a block of data.

    (b) B. The time it takes to position the read-write mechanism at the desired location.

    (c) C. The time it takes to perform a read or write operation.

    (d) D. The time required before a second access can commence.

    Correct answer is: C. The time it takes to perform a read or write operation. Because: For random-access memory, access time (latency) is defined as the time it takes to perform a read or write operation.

16. Question 16: What is "Memory cycle time" defined as?

    (a) A. The time it takes to complete a single memory access, including any transients.

    (b) B. The rate at which data can be transferred into or out of a memory unit.

    (c) C. The time it takes to position the read-write mechanism at the desired location.

    (d) D. The time for a read or write operation plus any additional time required before a second access can commence.

    Correct answer is: D. The time for a read or write operation plus any additional time required before a second access can commence. Because: Memory cycle time is defined as access time plus any additional time required before a second access can commence, which may include time for transients to die out or to regenerate data.

17. Question 17: For random-access memory, what is the transfer rate equal to?

    (a) A. Access time.

    (b) B. Memory cycle time.

    (c) C. 1 / (cycle time).

    (d) D. The number of electrical lines into and out of the memory module.

    Correct answer is: C. 1 / (cycle time). Because: For random-access memory, the transfer rate is equal to 1/(cycle time).

18. Question 18: Which type of memory loses information naturally or when electrical power is switched off?

    (a) A. Nonvolatile memory.

    (b) B. Erasable memory.

(c) C. Volatile memory.

(d) D. Read-only memory (ROM).

Correct answer is: C. Volatile memory. Because: Volatile memory is characterized by information decaying naturally or being lost when electrical power is switched off.

19. Question 19: Which type of memory, once recorded, retains information without deterioration until deliberately changed, and needs no electrical power to retain information?

    (a) A. Volatile memory.

    (b) B. Semiconductor memory.

    (c) C. Nonvolatile memory.

    (d) D. Random-access memory.

    Correct answer is: C. Nonvolatile memory. Because: Nonvolatile memory retains information without deterioration until deliberately changed and does not require electrical power to retain information.

20. Question 20: What is semiconductor memory of the nonerasable type known as?

    (a) A. RAM.

    (b) B. Cache.

    (c) C. ROM (Read-Only Memory).

    (d) D. Magnetic disk.

    Correct answer is: C. ROM (Read-Only Memory). Because: Semiconductor memory of the nonerasable type, which cannot be altered except by destroying the storage unit, is known as read-only memory (ROM).

Memory Hierarchy

21. Question 21: What are the three design constraints on a computer's memory that are summed up by questions "How much, how fast, how expensive"?

    (a) A. Capacity, access time, and cost.

    (b) B. Speed, reliability, and security.

    (c) C. Volatility, erasability, and organization.

    (d) D. Location, unit of transfer, and physical type.

Correct answer is: A. Capacity, access time, and cost. Because: The design constraints on a computer's memory are summed up by "How much, how fast, how expensive," referring to capacity, access time, and cost, respectively.

22. Question 22: What is the trade-off relationship between faster access time and cost per bit in a memory hierarchy?

    (a) A. Faster access time, smaller cost per bit.

    (b) B. Faster access time, greater cost per bit.

    (c) C. Faster access time, no change in cost per bit.

    (d) D. Faster access time, irrelevant to cost per bit.

    Correct answer is: B. Faster access time, greater cost per bit. Because: In a memory hierarchy, there is a trade-off where faster access time correlates with a greater cost per bit.

23. Question 23: How does greater capacity relate to cost per bit and access time in a memory hierarchy?

    (a) A. Greater capacity, greater cost per bit, faster access time.

    (b) B. Greater capacity, smaller cost per bit, faster access time.

    (c) C. Greater capacity, greater cost per bit, slower access time.

    (d) D. Greater capacity, smaller cost per bit, slower access time.

    Correct answer is: D. Greater capacity, smaller cost per bit, slower access time. Because: A greater capacity in a memory hierarchy typically means a smaller cost per bit and slower access time.

24. Question 24: What manages the unit of transfer for the "Registers" level in a memory architecture?

    (a) A. Operating system (OS)

    (b) B. Processor hardware

    (c) C. Compiler

    (d) D. OS/user

    Correct answer is: C. Compiler. Because: For the Registers memory level, the unit of transfer (Word - 32 bits) is managed by the Compiler.

25. Question 25: What manages the unit of transfer for the "Cache" level in a memory architecture?

    (a) A. Compiler

    (b) B. Processor hardware

    (c) C. Operating system (OS)

(d) D. OS/user

Correct answer is: B. Processor hardware. Because: For the Cache memory level, the unit of transfer (Cache block - 32 bytes) is managed by the Processor hardware.

26. Question 26: What is "secondary memory" or "auxiliary memory" also referred to as?

    (a) A. Internal volatile memory.

    (b) B. External nonvolatile memory.

    (c) C. Primary memory.

    (d) D. Processor registers.

    Correct answer is: B. External nonvolatile memory. Because: External, nonvolatile memory is also referred to as secondary memory or auxiliary memory.

27. Question 27: Which design principle for a memory hierarchy dictates that all information items are originally stored in the level most remote from the processor?

    (a) A. Locality.

    (b) B. Coherence.

    (c) C. Inclusion.

    (d) D. Performance.

    Correct answer is: C. Inclusion. Because: The Inclusion principle dictates that all information items are originally stored in level Mn, where n is the level most remote from the processor.

28. Question 28: Which design principle states that copies of the same data unit in adjacent memory levels must be consistent, and if a word is modified in the cache, its copies must be updated at higher levels?

    (a) A. Locality.

    (b) B. Inclusion.

    (c) C. Coherence.

    (d) D. Hierarchy. Correct answer is: C. Coherence. Because: Coherence states that copies of the same data unit in adjacent memory levels must be consistent, requiring updates at higher levels if a word is modified in the cache.

    (e) Question 29: What is the principle that makes effective use of a memory hierarchy possible?

        i. A. Coherence.

ii. B. Inclusion.

iii. C. Locality.

iv. D. Random access.

Correct answer is: C. Locality. Because: Locality is the principle that makes effective use of a memory hierarchy possible.

(f) Question 30: A cache acts as a buffer between which two components, creating a two-level internal memory?

i. A. CPU and hard disk.

ii. B. Main memory and processor.

iii. C. I/O controller and external memory.

iv. D. Registers and main memory.

Correct answer is: B. Main memory and processor. Because: A cache acts as a buffer between main memory and the processor, creating a two-level internal memory.

(g) Question 31: Which of the following are other instances of a two-level memory approach that exploit locality and are at least partially implemented in the operating system?

i. A. Processor registers and cache.

ii. B. Main memory and ROM.

iii. C. Virtual memory and disk cache.

iv. D. Magnetic disks and optical disks.

Correct answer is: C. Virtual memory and disk cache. Because: Virtual memory and disk cache are two other instances of a two-level memory approach that exploit locality and are at least partially implemented in the operating system.

(h) Question 32: In the operation of a two-level memory, what happens if an attempt to access an item in the upper-level memory (M1) fails?

i. A. The system reports an error.

ii. B. The item is directly accessed from M2 without copying.

iii. C. A block of memory locations is copied from M2 to M1, and the access then takes place via M1.

iv. D. The processor waits until M1 is available.

Correct answer is: C. A block of memory locations is copied from M2 to M1, and the access then takes place via M1. Because: If an access attempt in M1 fails, a block of memory locations is copied from M2 to M1, and the access then occurs via M1.