①:

NAME IKRAMULLAH

ID # 15072

DEPARTMEN BS-CS

SUBJECT C-ARCHITECTURE

ANSWER#1

As access time becomes faster the cost per bit inches. As memory size increase the cost per bit is smaller. Also with greater capacity, the access time becomes slower.

ANSWER#2

Another distinction among memory types is the method of accessing units of data.

Sequential access

Memory is organized into units of data called records. Access must be made in a specific linear sequence. Stored addressing information is used to separate records and assist in the retrievel process. A shared red-write mechanism is used and this must be moved from its current location to the desired location.

Direct access

Direct access involves a shared read write mechanism. Access is accomplished

Scanned with CamScanner

by direct access to reach a general vicinity plus sequential searching, counting or waiting to reach the final location.

Random access.

Each addressable location in memory has a unique, physically wired-in addressing mechanism. Any location can be accessed at random and directly addressed and accessed. Main memory and some cache systems are random access.

Associative

This is a random access type of memory each location has own addressing mechanism and retrievel time is constant independent of location or prior access patterns.

ANSWER#3

Direct maping

The simplest technique

Maps each book of main memory into only possible cache line

Associative maping

To determine wether a block is inthe cache. The cache control logic must simutaneously examine every line's Tag for match.

Scanned with CamScanner

Memory unit of transfer

it is the maximum number of bits that can be read or written into the memory at a time in case of main memory it is mostly equal to word size in case of external memory unit of transfer is not limited to the word size.

Memory performance parameters

The two most important characteristics of memory are capacity and performance Three performance parameters are used

Access time

For random access memory, this is the time it takes to perform a read or write operation that is the time from the instant that an address is presented to the me mory to the instant data have been stored.

Memory cycle time

This concept is primarily applied to random-access memory and consists of the access time plus any additional time required before a second access can commence This additional time may be required for transients to die out on signal lines or to regenerat

data if they are read destructively

Transfer rate

This is the rate at which data can be transferred into or out of a memory unit. For random-access memory it is equal to 1/(cycle time).

Disk Cache

A portion of main memory can be used as a buffer to hold data temporarily that is to be read out to disk. A few large transform of data can be used instead of many small transfers of data. Can be used instead of many small transfers of data

Principle of locality

The principle of locality states that data in the vicinity of a referenced word are likely to be referenced in the near future

Replacement algorithms

Once the cache has been filled when a new block is brought into the cache one of the existing blocks must be replaced. For direct maping there is

There is only one possible line for any particular blocks and no choice is possible.

Possible approaches to Cache Coherency.

Possible approaches to cache coherency include the following.

Bus watching with write through

Each cache Controller monitors the address lines to detect write operations to memory by others bus masters. if another master writes to a location in shared memory that also resides inthe cache... memory the Cache Controller invalidses that cache entry

Hardware transparency.

Additional hardware is used to ensure that all updates to main memory in addition any matching words inthe others caches are similarly updated.

Non-Capable memory.

only a portion of main memory is shared by more than one processor and this is designated as non-cheable in such a system all accesses to shared

Sequential

Memory is organized into units of data called records. Access must be made in a specific linear sequence. Stored addressing information is used to seperate records and assist inthe retrievel proces.

Direct

As with sequential access direct access involves a shared read-write mechanism individual blocks or records have a unique address based on physical location

Random access

Each addressable location in memory has a unique phsically wired in addressing mechanism The time to access a given location is independent of the sequence of prior accesses and is constant.

Direct Maping

The direct maping technique is simple and inexpensive to implement its main disadvantages is that three is a fixed cache location for any given block if a program happens to reference words repeatedly from two different blocks.

Scanned with CamScanner

Associative maping

With associative maping there is flexibility as to which block to replace when a new black is read into the cache. Replacement algorithms discussed later in this section are designed to maximize the hit ratio.

Set associative maping

Set associative maping is a compromise that exhibits the strength of both the direct and associative approaches while reducing their disadvantages in this case the cache consist of a number sets each of which consist of a number of lines

Write through

Simplest technique
All write operations are made to main memory as well as to the cache
The main disadvantages of this technique is that generates substantial memory traffic and may creat a bottleneck

Write Back

Minimize memory writes. update are made only inthe cache. This make for