

Introduction of Computer Memory

What is the requirement of Memory:

A memory is just like a human brain. It is used to store data and instructions.

Memory can be characterized by the following points like:

- Speed
- Size
- Cost

Unit of Memory Size:

1 bit=the value of 0 or 1

8 bits = 1 byte

1024 byte =1 kilobyte

1024 kilobytes = 1 megabyte

1024 megabytes = 1 gigabyte

1024 gigabytes = 1 terabyte

1024 terabytes = 1 petabyte

Types of memory:

Memory is primarily of three types

- Cache Memory
- Primary Memory/Main Memory
- Secondary Memory

Cache Memory

Cache memory is a very high speed semiconductor memory which can speed up CPU. It acts as a buffer between the CPU and main memory. It is used to hold those parts of data and

program which are most frequently used by CPU. The parts of data and programs are transferred from disk to cache memory by operating system, from where CPU can access them.

Advantages

The advantages of cache memory are as follows:

- Cache memory is faster than main memory.
- It consumes less access time as compared to main memory.
- It stores the program that can be executed within a short period of time.
- It stores data for temporary use.

Disadvantages

The disadvantages of cache memory are as follows:

- Cache memory has limited capacity.
- It is very expensive.

Primary Memory (Main Memory)

Primary memory holds only those data and instructions on which computer is currently working. It has limited capacity and data is lost when power is switched off. It is generally made up of semiconductor device. These memories are not as fast as registers. The data and instruction required to be processed reside in main memory. It is divided into two subcategories RAM and ROM.

Types of RAM:

- Static RAM
- Dynamic RAM

Types of ROM:

- ROM
- PROM
- EPROM • EEPROM

ROM Vs RAM

There is one major difference between a ROM and a RAM chip.

- A ROM chip is non-volatile storage and does not require a constant source of power to retain information stored on it. When power is lost or turned off, a ROM chip will keep the information stored on it. So for permanent storage, ROM is used.
- In contrast, a RAM chip is volatile and requires a constant source of power to retain information. When power is lost or turned off, a RAM chip will lose the information stored on it.
- A ROM chip is used primarily in the start up process of a computer, whereas a RAM chip is used in the normal operations of a computer after starting up and loading the

operating system.

- Writing data to a ROM chip is a slow process, whereas writing data to a RAM chip is a faster process

- A RAM chip can store multiple gigabytes (GB) of data, up to 16 GB or more per chip;

A ROM chip typically stores only several megabytes (MB) of data, up to 4 MB or more per chip .

Advantages of Main Memory

These are semiconductor memories

- It is working memory of the computer.

- Faster than secondary memories.

- Cheaper than Cache memory.

Disadvantages of Main Memory:

- Volatile when power goes off.

- High cost as transistors are used.

- Usually size is less than secondary memory. Secondary Memory

This type of memory is also known as external memory or non-volatile. It is slower than main memory. These are used for storing data/Information permanently. CPU directly does not access these memories instead they are accessed via input-output routines. Contents of secondary memories are first transferred to main memory, and then CPU can access it. For example : disk, CD-ROM, DVD etc.

Advantages of Secondary Memory

- These are magnetic and optical memories
- It is known as backup memory.
- It is non-volatile memory.
- Data is permanently stored even if power is switched off.
- It is used for storage of data in a computer.

Disadvantages of Secondary Memory

- Computer doesn't deal with secondary memory at the beginning of execution.
- Slower than primary memories.

A hit is when data is found at a given memory level. –

A miss is when it is not found. –

The hit rate is the percentage of time data is found at a given memory level. –

The miss rate is the percentage of time it is not. –

Miss rate = 1 - hit rate.

The hit time is the time required to access data at a given memory level.

The miss penalty is the time required to process a miss, including the time that it takes to replace a block of memory plus the time it takes to deliver the data to the processor.

What is the principle of Locality?

An entire blocks of data is copied after a hit because the principle of locality tells us that once a byte is accessed, it is likely that a nearby data element will be needed soon.

There are three forms of locality: – Temporal locality- Recently-accessed data elements tend to be accessed again.
– Spatial locality -Accesses tend to cluster. – Sequential locality - Instructions tend to be accessed sequentially.

Q How does we Measure the performance of Memory?

The performance of hierarchical memory is measured by its effective access time (EAT). • EAT is a weighted average that takes into account the hit ratio and relative access times of successive levels of memory. • The EAT for a two-level memory is given by: $EAT = H \cdot Access_C + (1-H) \cdot Access_{MM}$. where H is the cache hit rate and AccessC and AccessMM are the access times for cache and main memory, respectively.

For example, consider a system with a main memory access time of 200ns supported by a cache having a 10ns access time and a hit rate of 99%. • Suppose access to cache and main memory occurs concurrently. (The accesses overlap.) • The EAT is: $0.99(10ns) + 0.01(200ns) = 9.9ns + 2ns = 11ns$. 6.4 Cache Memory

34

• For example, consider a system with a main memory access time of 200ns supported by a cache having a 10ns access time and a hit rate of 99%. • If the accesses do not overlap, the EAT is:

$0.99(10ns) + 0.01(10ns + 200ns) = 9.9ns + 2.01ns = 12ns$.

Q What are the dirty Blocks?

Cache replacement policies must take into account dirty blocks, those blocks that have been updated while they were in the cache. • Dirty blocks must be written back to memory. A write policy determines how this will be done. • There are two types of write policies, write through and write back.

Q what are Harvard Cache?

Many modern systems employ separate caches for data and instructions. – This is called a Harvard cache.

Q What is Victim Cache?

Cache performance can also be improved by adding a small associative cache to hold blocks that have been evicted recently. – This is called a victim cache