Memory: It facilitates the remembrance power to computer system. It refers to the physical devices used to store programs (sequences of instructions) or data (e.g. program state information) on a temporary or permanent basis for use in a computer or other digital electronic device. The term primary memory is used for the information in physical systems which are fast (i.e. RAM), as a distinction from secondary memory, which are physical devices for program and data storage which are slow to access but offer higher memory capacity. Primary memory stored on secondary memory is called virtual memory. Primary Memory can be categorized as Volatile Memory & Non-Volatile Memory.

Memory hierarchy:

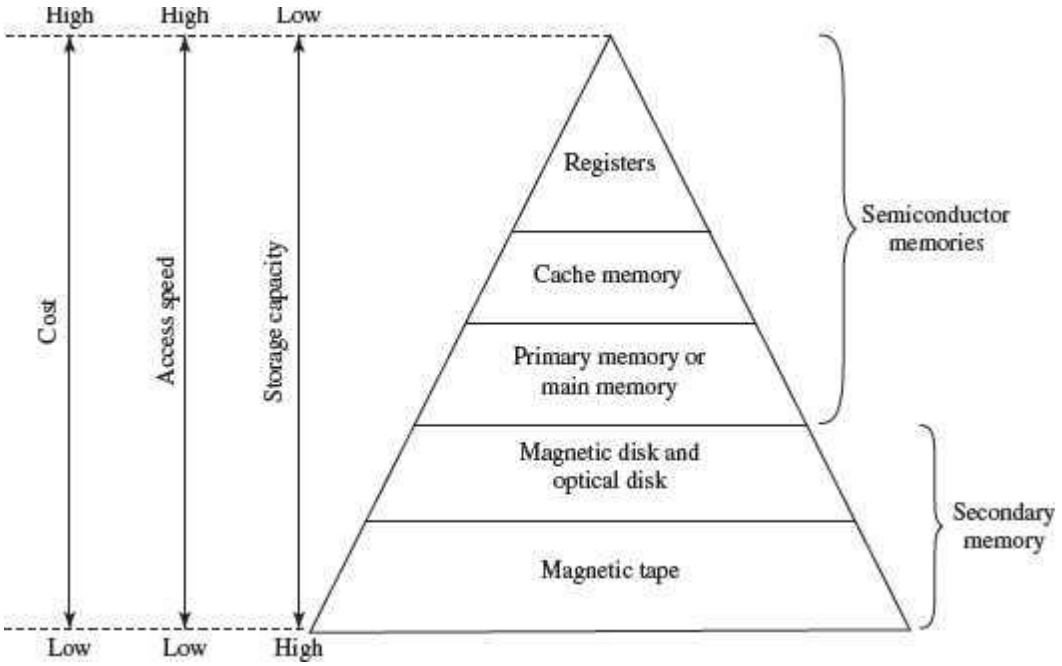


Fig: Memory hierarchy

Units of Memory:

The smallest unit is bit, which mean either 0 or 1.

1 bit = 0 or 1

1 Byte = 8 bit

1 Nibble = 4 bit

1 Kilo Byte = 1024 Byte= 210 Byte

1 Mega Byte = 1024 KB= 210 KB

1 Gega Byte = 1024 MB= 210 MB

1 Tera Byte = 1024 GB= 210 GB

1 Peta Byte =1024 TB= 210 TB

1 Exa Byte =1024 PB= 210 PB

1 Zetta Byte = 1024 EB= 210 EB

1 Yotta Byte = 1024 ZB= 210 ZB

Memory Access Method:

Sequential access: In this access, it must start with beginning and read through a specific linear sequence. This means access time of data unit depends on position of records (unit of data) and previous location.

e.g. tape

Direct Access: Individual blocks of records have unique address based on location. Access is accomplished by jumping (direct access) to general vicinity plus a sequential search to reach the final location.

e.g. disk

Random access: The time to access a given location is independent of the sequence of prior accesses and is constant. Thus any location can be selected out randomly and directly addressed and accessed.

e.g. RAM

Associative access: This is random access type of memory that enables one to make a comparison of desired bit locations within a word for a specified match, and to do this for all words simultaneously.

e.g. cache

Memory Access Time:

Disk access time: The sum of seek time, latency time and time for data transfer is the access time of the disk.

- The time taken to move the read/write head to the desired track is called the seek time.
- The time taken for desired sector of the track to come under read/write head is called the latency time.
- The rate at which data is written to disk or read from disk is called data transfer rate.

What are Cache Hit and Misses?

1. A cache hit refers to the situation wherein the cache is able to successfully retrieve data and content that was saved to it.
2. A cache miss occurs in the opposite situation.

Hit and Miss Ratios in Caches:

Miss ratio = # of cache misses / # of content requests

Miss ratio = 1 - Hit ratio

Hit ratio = # of cache hits / (# of cache hits + # of cache misses)

Hit ratio = 1 – miss ratio

Register, Main and Backing store:

Register

(already…..)

Main

(already…..)

Backing store (secondary memory):

In comparison to the primary memory, the secondary memory stores much larger amounts of data and information (for example, an entire software program) for extended periods of time. The data and instructions stored in secondary memory must be fetched into RAM before processing is done by CPU.

A. Hard Disk (Local Disk)

B. Optical Disks: CD-R, CD-RW, DVD-R, DVD-RW

C. Pen Drive

D. Floppy Disks

F. Memory Cards

G. External Hard Disk

Magnetic Tape:

A magnetic tape is the strip of plastic coated with a magnetic recording medium. Data can be recorded and read as a sequence of character through read / write head. It can be stopped,

started to move forward or in reverse or can be rewound. Magnetic tape is a plastic tape with magnetic coating (Figure 3.7). It is a storage medium on a large open reel or in a smaller cartridge or cassette (like a music cassette). Magnetic tapes are cheaper storage media. They are durable, can be written, erased, and re-written. Magnetic tapes are sequential access devices, which mean that the tape needs to rewind or move forward to the location where the requested data is positioned in the magnetic tape. Due to their sequential nature, magnetic tapes are not suitable for data files that need to be revised or updated often. They are generally used to store back-up data that is not frequently used or to transfer data from one system to other.

Fig: Magnetic tape

The working of magnetic tape is explained as follows:

1. Magnetic tape is divided horizontally into tracks (7 or 9) and vertically into frames (Figure below). A frame stores one byte of data, and a track in a frame stores one bit. Data is stored in successive frames as a string with one data (byte) per frame.

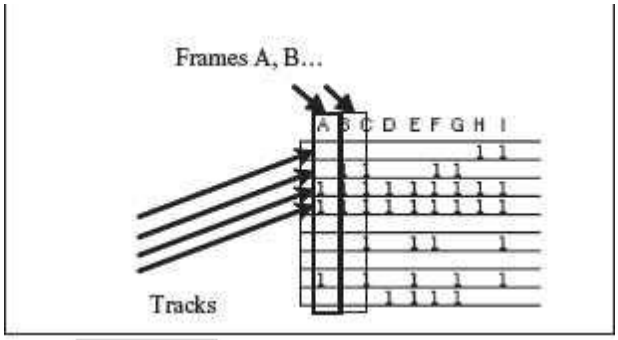


Figure: A portion of magnetic tape

2. Data is recorded on tape in the form of blocks, where a block consists of a group of data also called as records. Each block is read continually. There is an Inter-Record Gap (IRG) between two blocks that provides time for the tape to be stopped and started between records (Figure below).

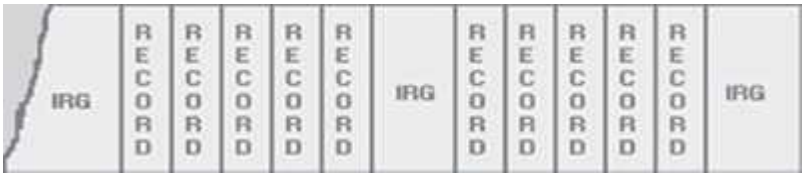


Figure: Blocking of data in a magnetic tape

3. Magnetic tape is mounted on a magnetic tape drive for access. The basic magnetic tape drive mechanism consists of the supply reel, take-up reel, and the read/write head assembly. The magnetic tape moves on tape drive from the supply reel to take-up reel, with its magnetic coated side passing over the read/write head.
4. Tapes are categorized based on their width - ¼ inch, ½ inch, etc.
5. The storage capacity of the tape varies greatly. A 10–inch diameter reel of tape which is
6. 2400 feet long can store up to 180 million characters.

Magnetic Disk:

A magnetic disk is a circular plate constructed with metal or plastic coated with magnetic material often both side of disk are used and several disk stacked on one spindle which Read/write head available on each surface. All disks rotate together at high speed. Bits are stored in magnetize surface in spots along concentric circles called tracks. The tracks are commonly divided into sections called sectors. After the read/write head are positioned in specified track the system has to wait until the rotating disk reaches the specified sector under read/write head. Information transfer is very fast once the beginning of sector has been reached.

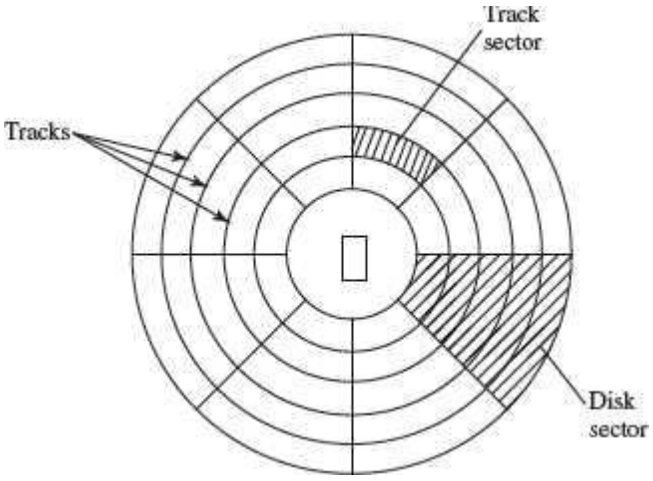


Figure: Tracks and sectors of a disk

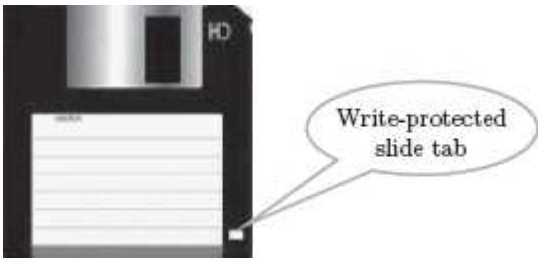
Disk access time: The sum of seek time, latency time and time for data transfer is the access time of the disk.

- The time taken to move the read/write head to the desired track is called the seek time.
- The time taken for desired sector of the track to come under read/write head is called the latency time.
- The rate at which data is written to disk or read from disk is called data transfer rate.

Floppy Disk:

Floppy disk (FD) is a flat, round, single disk made of Mylar plastic and enclosed in square plastic jacket.

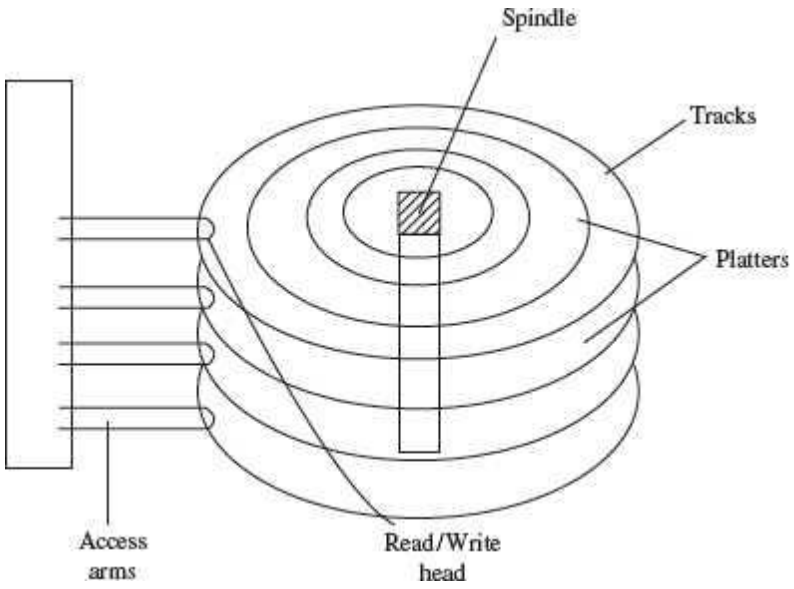
1. Floppy Disk Drive (FDD) is the disk drive for floppy disk.
2. The floppy disk is inserted into the floppy disk drive to read or write data to it.
3. Floppy disk has a write-protect slide tab that prevents a user from writing to it.
4. A floppy disk may be single-sided or double-sided disk, i.e., data can be read and written on one and both sides of floppy disk, respectively.



5. They are portable. They can be removed from the disk drive, carried or stored separately.
6. They are small and inexpensive.
7. Floppy disks are slower to access than hard disk. They have less storage capacity and are less expensive than hard disk.
8. They come in two basic sizes—5-¼ inch and 3-½ inch.
9. The 5-¼ inch disk came around 1987. It can store 360 KB to 1.2 MB of data.
10. The 3-½ inch disk has capacity of 400 KB to 1.44 MB. It usually contains 40 tracks and 18 sectors per track and can store 512 bytes per sector.

Hard Disk:

1. A hard disk (HD) consists of one or more platters divided into concentric tracks and sectors. It is mounted on a central spindle, like a stack. It can be read by a read/write head that pivots across the rotating disks. The data is stored on the platters covered with magnetic coating (Figure below).



2. Hard disk is a fixed disk. The disk is not removable from the drive, unlike floppy disk.
3. The hard disk and Hard Disk Drive (HDD) is a single unit.
4. Hard disk can store much more data than floppy disk. The data in hard disk are packed more closely (because fast spinning uses smaller magnetic charges) and they have multiple platters, with data being stored on both sides of each platter. Large capacity hard disks may have 12 or more platters.
5. Unlike floppy disk, the read/write head of hard disk does not touch the disk during accessing.
6. Hard disk can spin at the speed of up to 10,000 revolutions per minute and have an access time of 9—14 ms. It stores 512 bytes per sector but the number of sectors are more per track (54 or more) than floppy disk.
7. Nowadays, hard disks are available that can store up to 500 GB of data. Generally, PCs come with 160 GB hard disk.
8. Hard disk is the key secondary storage device of computer. The operating system is stored on the hard disk. The performance of computer like speed of computer boot up, loading of programs to primary memory, loading of large files like images, video, audio etc., is also dependent on the hard disk.
9. Nowadays, portable external hard disk drive is available which can be attached to the USB drive of the computer. They come in the storage capacities of 80 GB to 500 GB.

CD-ROM:

1. Originally, Compact Disk (CD) was a popular medium for storing music. Now, it is used in computers to store data and is called Compact Disk-Read Only Memory (CD-ROM). As the name suggests, CD-ROM (Figure 3.16) is an optical disk that can only be read and not written on. CD-ROM is written on by the manufacturer of the CD-ROM using the laser light.
2. As CD-ROM is read only, no changes can be made into the data contained in it.

3. Since there is no head touching the disk, but a laser light, CD-ROM does not get worn out easily.

DVD:

1. Digital Video Disk-Read Only Memory (DVD-ROM) is an optical storage device used to store digital video or computer data (Figure below).
2. DVDs look like CDs, in shape and physical size.
3. It is a high-density medium with increased track and bit density.
4. DVD-ROM uses both sides of the disk and special data compression technologies. The tracks for storing data are extremely small.
5. New DVD-ROMs use layers of data track, to double its capacity. Such dual layer disks can store 17 GB of data.

CD-R:

A CD R also known as compact disc recordable is a disc that can be written to once and then can be read repeatedly. CD-R disc cannot be formatted after it has been created, and data cannot be removed from it.

CD-RW:

CD-RW (Compact Disc-Rewritable) is a digital optical disc storage format introduced in 1997. A CD-RW compact disc (CD-RWs) can be written, read, erased, and re-written. CD-RWs must be erased or blanked before reuse. Erasure methods include full blanking where

the entire surface of the disc is erased. According to what the manufacturers of these disks say, it is possible to record and erase such a disk up to 1000 times.