2-Memory unit in a computer

In this chapter, we will gain an understanding of the memory units in a computer and how data is internally stored and processed. When discussing memory in a computer, several questions arise: What are the different types of memory? How are they organized? How do they work internally?

The main component of a computer is the processor, also known as the CPU, which we discussed in the previous chapter. The CPU is often referred to as the brain of the computer and is capable of executing any task or instruction given to it, being made up of semiconductor technology.

While the CPU can execute instructions and provide the desired output, these instructions, or programs, need a storage medium. In the 1950s, many multinational companies like IBM spent years researching to find a solution for data storage. They finally developed a device called the Hard Disk Drive (HDD) as a storage medium.

The Role of HDD and BUS in a Computer

When instructions are given and the desired output is to be displayed, we use the HDD to store these instructions. The path used for storing and executing these instructions is called a BUS.

Fig. 1: **HDD drive**

BUS:

A set of wires that carry information from one unit to another in a computer is called a BUS.

It transfers data between various components inside a computer or between computers.

The BUS is a critical component that facilitates communication between different parts of the computer, including the memory unit, CPU, and peripheral devices.

The instructions are passed to the processor slowly. Once they reach the processor, it immediately provides the output within a fraction of a second. However, because HDDs are made with electromechanical and electromagnetic technology, they are slower compared to semiconductor technology. This speed mismatch reduces the overall efficiency and performance of the computer.

Disadvantages of HDD:

- HDDs are bulky in size.
- They are slow in execution due to electromechanical and electromagnetic technology.
- They are noisy while working.
- HDDs consume a lot of power due to many mechanical parts.
- They are more prone to mechanical failure and damage, making them less durable.

Transition to RAM

After years of research, a new storage device called RAM (Random Access Memory) was invented. Data or instructions stored on an HDD for execution were shifted to RAM, which is connected to the processor using the same semiconductor technology to match the speed.

Advantages of Semiconductor Technology:

- It has a very fast execution speed.
- It is small and compact in size.
- It is less noisy compared to other technologies.
- This advancement allowed programs to run and execute successfully, providing the desired output without the speed mismatch issues.

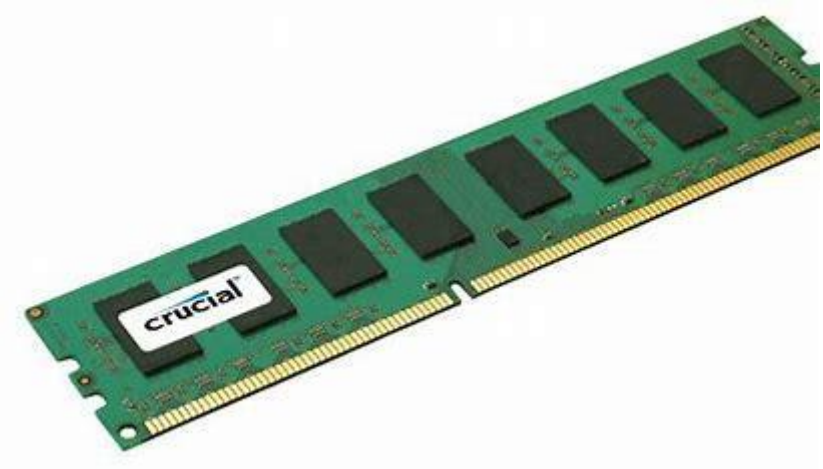
Disadvantages of RAM:

- RAM is volatile, meaning it loses data when power is lost.
- It requires a continuous power supply to function; even a microsecond power outage can result in data loss.
- RAM is expensive compared to other storage options.

Comparing HDD and RAM

While HDDs are less volatile and do not need a continuous power supply to function, they are slower. To maintain data safety and ensure fast program execution, data is stored from RAM to HDD (saving) and loaded from HDD to RAM (loading).

fig.2:**RAM memory**

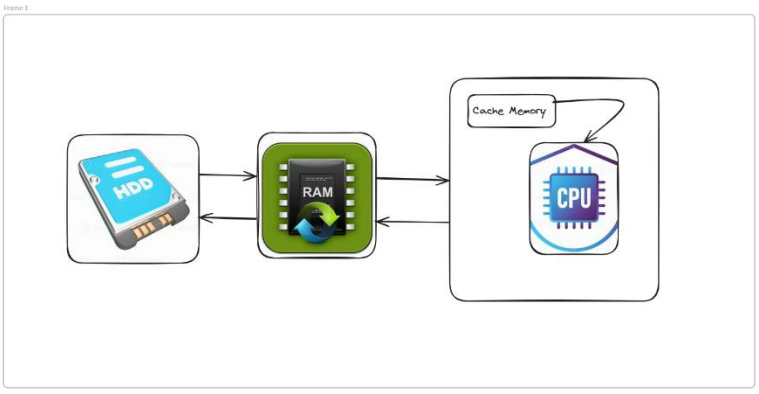


Advantages of HDD over RAM:

- HDDs are less volatile compared to RAM.
- Data on HDDs is safe and retrievable even after years, without needing continuous power.
- HDDs are less expensive compared to RAM.

When executing a program, it is first loaded from the HDD to RAM and then processed. For example, playing a multimedia file like a video takes time to load initially but plays smoothly once fully loaded.

Memory hierarchy and cache memory



The storage area on an HDD is called a **file,** while the storage area in RAM is called a **byte**. The smallest unit for performing tasks based on instructions is called a **register**. The cost hierarchy in terms of speed is: CPU > RAM > HDD.

RAM is called **main memory** because it is connected closer to the processor, while HDD is called **secondary memory**. When running multiple tasks beyond the capacity of RAM, we experience lag or hanging situations. Thus, having more RAM allows the CPU to perform tasks faster and more smoothly without lag.

A device faster than RAM, called **cache memory**, is placed between the RAM and the processor. Cache memory stores tasks accessed multiple times, enhancing performance. Cache memory is measured in kilobytes (KB) to megabytes (MB), while RAM is measured in gigabytes (GB) and HDDs in terabytes (TB) or more.

Conclusion

Now we have a complete understanding of how data is stored and processed behind the scenes. In the next chapter, we will learn about what is meant by a platform in terms of programming.

Stay curious and happy learning!