Part 1: Introduction to Memory

In modern computer systems, memory is organized in a hierarchy based on speed, size, cost, volatility, and proximity to the CPU.

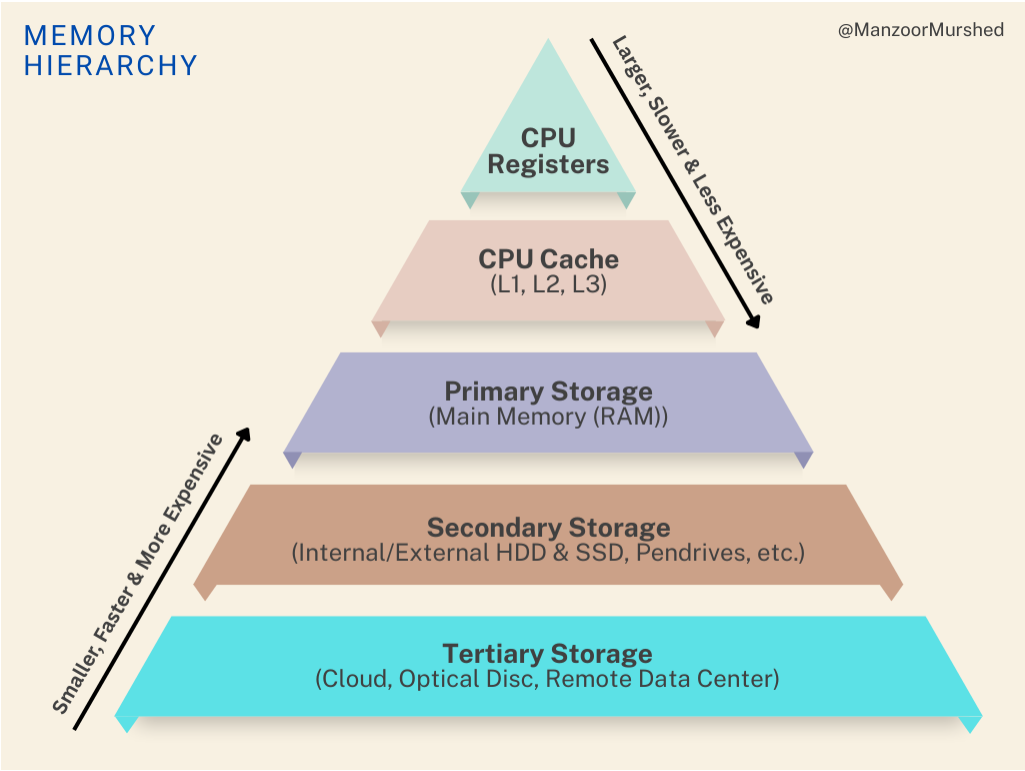

Figure 1: Computer Memory Hierarchy

The Memory Hierarchy – From Fastest to Slowest

📌 CPU Registers 📌

CPU Registers are small but very powerful memory units located inside the CPU (Central Processing Unit). Despite their small size, they play a crucial role in the performance and efficiency of a computer or any digital system.

Registers are essential temporary storage of data and instructions that the CPU is currently using. Unlike RAM or cache memory, registers are directly connected to the CPU's internal circuitry, making them extremely fast to read or write data to — typically within a single CPU clock cycle.

- Registers are directly connected to the CPU's hardware, so data does not need to be read or written from separate memory. This allows the CPU to access data very quickly, which saves time and makes processing more efficient.

- Registers are the fastest of all types of memory (such as RAM, cache, etc.). It takes only 1 CPU clock cycle to read or write data. As a result, instructions are executed very quickly.
- Registers are very small — usually just a few bits or bytes. No matter how powerful a CPU is, the number of registers is very limited. That's why compilers and hardware designers control their use very efficiently.
- When the CPU is performing a task, the required inputs (operands) or results of the task are temporarily stored in registers. For example: accumulator, program counter, instruction register, etc. — these are some of the registers that participate in the execution of the instruction.
- Registers require advanced technology to create and are directly connected to the CPU. This makes them the most expensive memory per bit. However, the speed that is available for this cost cannot be matched by any other memory.
- Registers are volatile, meaning that all data in them is lost when the computer's power is turned off. They only hold data temporarily during operation — not for permanent storage.

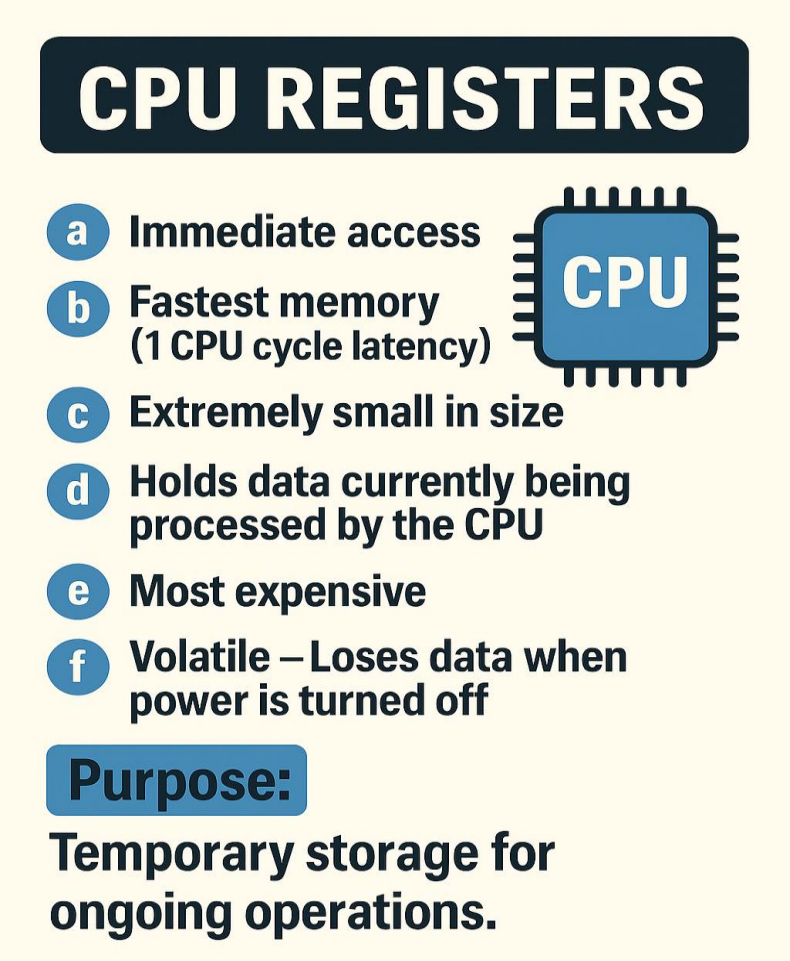

Figure 2: CPU Registers

📌 Cache Memory 📌

Cache Memory is a type of very fast working memory which is very close to the CPU and basically stores data and instructions that are needed repeatedly. This allows the CPU to take data directly from the Cache without going to the Main Memory repeatedly, resulting in a significant speed increase.

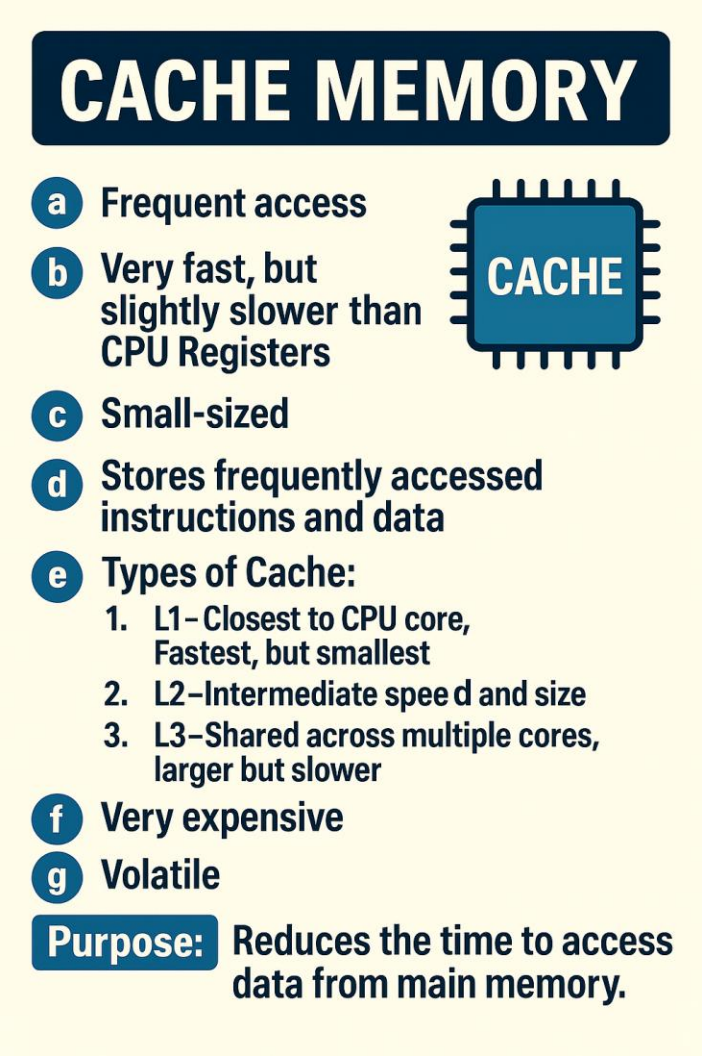

Figure 3: CPU Cache Memory

- Cache Memory stores data and instructions that the CPU uses very frequently. It does not have to bring data from the Main Memory repeatedly, so it works quickly.
- Cache Memory works very fast, but is a little slower than CPU Registers. Although CPU Registers are very immediate, Cache is a little slower with a little more space, but it is very effective.

- The cache space is not very large. Because if it is too large, its speed will decrease. So even though it is limited in size, its work is very important.
- Instructions or data that the CPU uses repeatedly are kept in the cache on a priority basis. This saves time and the computer responds quickly.
- Types of Cache:
  - ✓ L1 Cache: It is very close to the CPU core. Its speed is the highest, but its size is very small. Sometimes it is 32KB or 64 KB. The most important data is kept here.
  - ✓ L2 Cache: It is a little larger than L1, and the speed is medium. It is outside or next to the CPU, but it still works very fast. Sometimes it can be from 256KB to 1 MB.
  - ✓ L3 Cache: It usually uses multiple CPU cores together. The speed is relatively low, but the size is very large. It can be 4MB, 8MB, or even larger.
- Cache Memory is very expensive to make. Because it is made with very high-speed technology. So, its price is much higher than normal memory.
- Cache Memory is Volatile, meaning that all its data is deleted when the computer's power is turned off. It is only used during work.

📌 Primary Storage

Primary Storage is a place where the computer keeps the things it needs during its work, so that the CPU can get to them very quickly. It is a lot like a table — we keep the papers or things we are working on on the table. We can get them quickly if we need them. But after work, we put them in a cupboard (a place that can hold thousands of things) — this cupboard is the Secondary Storage.

- Primary Storage is a type of memory where the currently running (active) programs on the computer are stored. For example, if we run Microsoft Word and Chrome at the same time, the information of these two programs is stored in Primary Storage.
- Although its speed is much faster, it is a little slower than Cache Memory. This is because the physical distance and technical structure between RAM and Cache are different.
- The size of Primary Storage is usually much larger than Cache, but smaller than a hard drive or SSD. It is considered a medium-sized memory.
- The software or files we are running, or the data we need for our work, are also stored in RAM. This is called storage for ‘currently active data’.
- Moderate cost: The price of this type of memory is lower than Cache Memory, but higher than a hard drive or SSD. That is, it is a medium-priced memory.

- Primary Storage is Volatile Memory, meaning that when the computer is turned off, all the data in it is erased. For this reason, nothing can be stored in it permanently.
- Primary Storage is commonly called RAM (Random Access Memory). It is the main and most used memory unit of the computer.

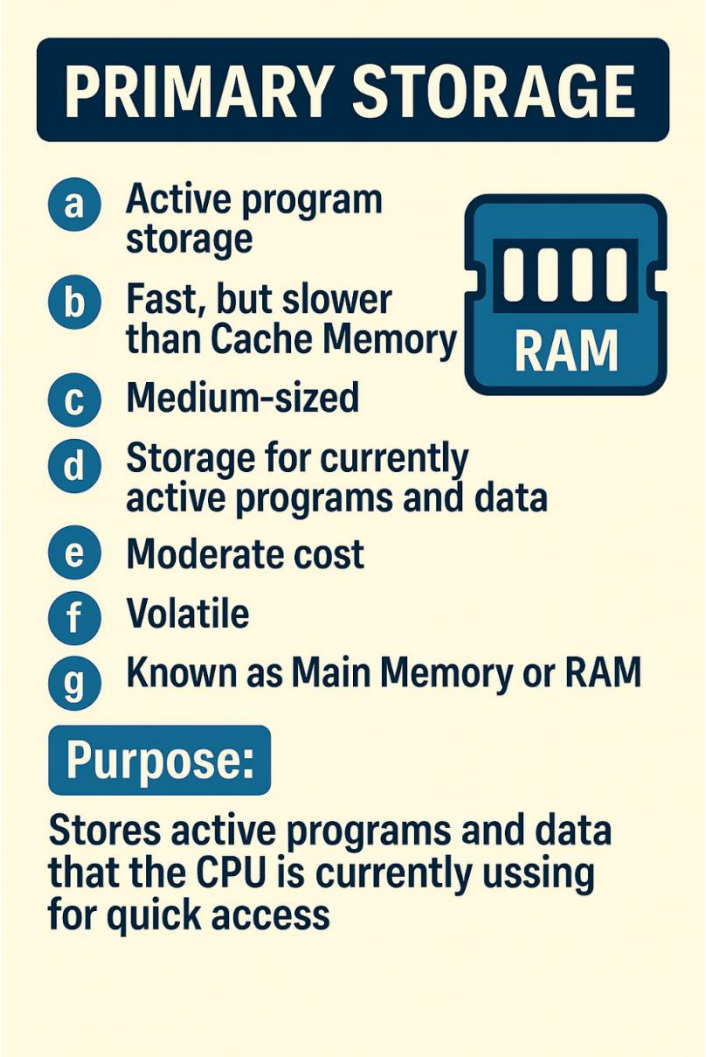

Figure 4: Primary Memory (RAM)

📌 Secondary Storage 📌

Secondary Storage is a type of computer memory where we store files, software, images, videos, documents, etc., for a long time.

- Files that need to be kept for a long time, such as images, videos, and software, are stored in this memory.

- Its speed is slower than Primary Memory, but it makes it suitable for storing data permanently.
- This memory has a much higher capacity, easily storing data up to many GB or TB.
- The operating system (OS), applications, office files, songs, or movies that we keep on the computer—all are stored in this storage.
- Its price is relatively low, so users can easily buy a large amount of memory.
- Even if the power is turned off, the data stored in it is not erased, but remains safe.
- Examples: HDDs, SSDs, external drives, and USB pen drives.

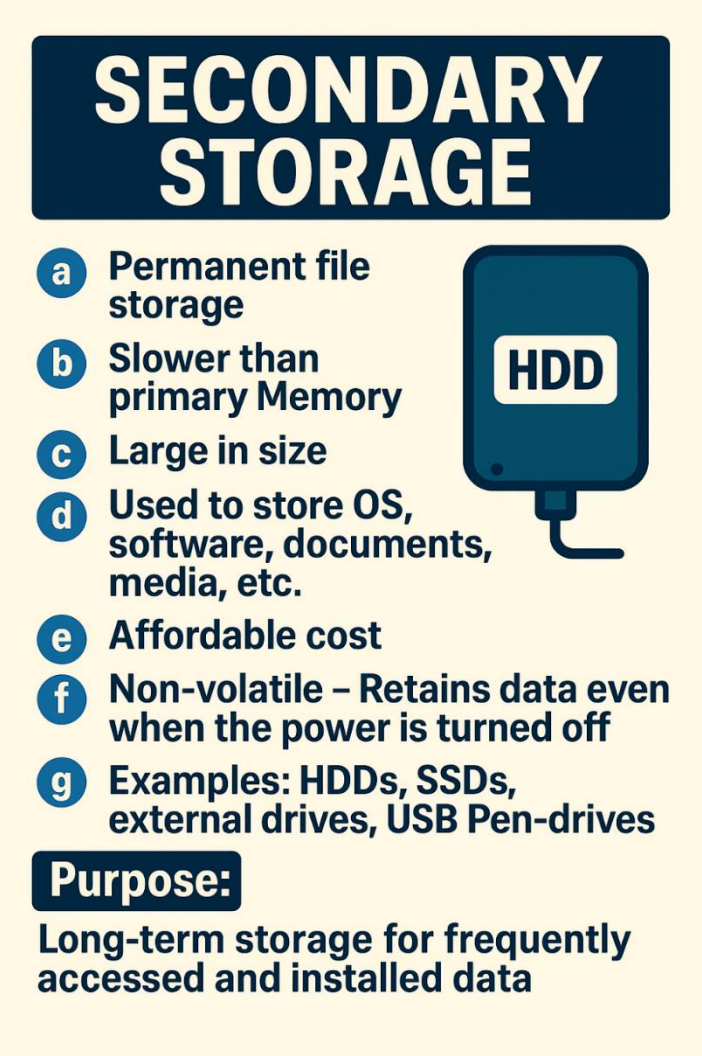

Figure 5: Secondary Storage

📌 Tertiary Storage 📌

Tertiary Storage is a type of memory that is usually used for backups, storing old data, or storing data on remote servers.

- This memory is used to store old, less-needed data or secure backups.
- The slowest, because these are mainly for data that is rarely used or kept for archiving purposes.
- Used for file backups, storing old data, and storing remote data.
- The cost of this type of storage is very low, so it is possible to store a lot of data.
- Once the data is stored, it remains even if there is no power, as is the case with Secondary Storage.
- It is highly effective and economical for storing large volumes of data that are not frequently accessed but still need to be preserved.

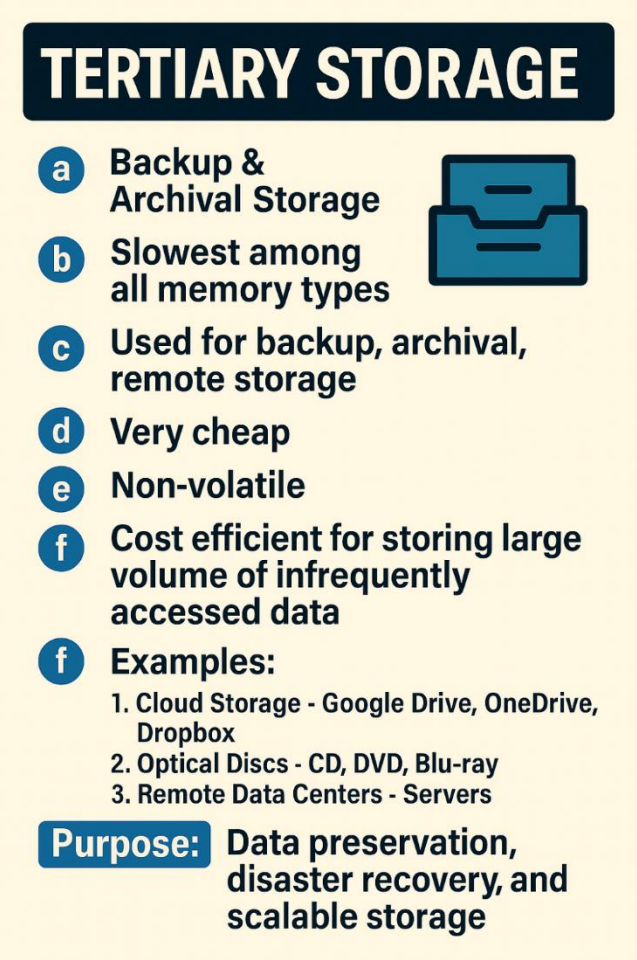

Figure 6: Tertiary Memory

- Examples:
  - ✓ Cloud Storage: e.g., Google Drive, OneDrive, Dropbox
  - ✓ Optical Discs: e.g., CD, DVD, Blu-ray
  - ✓ Remote Data Centers: Large servers where data is stored remotely.

Mnemonic for Easy Recall

"Really Cool Programmers Store Tapes"

- R - Registers
- C - Cache
- P - Primary (RAM)
- S - Secondary (HDD/SSD)
- T - Tertiary (Backups/Cloud)