# Storage: Background

In this section we will discuss theoretical concepts related to the storage of data, the three major types of cloud storage, and their characteristics. A concise list of the topics to be discussed is given below:

- Memory, Storage, and how do they differ from one another.
- On-prem storage, its advantages and disadvantages.
- Cloud storage, and its three different types.
- Object Storage and its use cases
- Block Storage and its use cases
- File Storage and its use cases

If the reader believes themselves to already be familiar with the listed concepts, then they can feel free to skip this section and move on to the rest of the chapter.

## Memory vs Storage

As technology has become more ubiquitous in modern society, it must come as little surprise that technical terms have also found their way into the modern vocabulary. **Memory** and **Storage** are two such terms, and are often used interchangeably by people to refer to a computer's capacity to hold data. However, there do exist contrasts between the two which I find important to delve into.

Memory, sometimes also referred to as RAM, is used to refer to the hardware component(s) responsible for holding data temporarily, data typically related to the instructions that are currently being executed or operated by the computer, a task typically done by the CPU. It is quite fast, and enables quick data processing but **volatile** in nature, meaning that all data within it is lost when the device is turned off.

Storage on the other hand, is a broad term encompassing all hardware components and devices used to hold data for longer periods of time, data

typically related to the OS, installed applications and personal files like documents, photos and videos. Though slower than memory, storage is non-volatile in nature and can typically handle much larger quantities of data than memory can.

Basically, while memory serves as the computer's short-term workspace, enabling rapid access to data for active tasks, storage acts as the long-term repository, preserving data permanently.

## On-prem Storage and RAID

Now, there are two major ways in which large organizations and enterprises acquire and organize their storage infrastructure: **On-premises**, **Cloud** and **Hybrid**. Since On-prem is the oldest kid on the block, it would be helpful for us to learn about it before transitioning towards Cloud and Hybrid storage.

On-Premises storage environments have all their storage systems, devices and related infrastructure physically located within or near their facilities, providing direct oversight and security control over them, and minimal redundancy to boot. It has remained extremely popular among legacy organizations and organizations with  high data security needs, compliance requirements, or specialized performance criteria.

Also, On-prem environments often utilize **RAID (Redundant Array of Independent Disks)** setups for the organization and management of their hardware devices, but since RAID and its multitude configuration types are a complex topic, beyond the scope of a book dedicated to Cloud computing, we will not be discussing it here. Though the topic might be worth looking into for the interested.

## Cloud Storage

Cloud storage environments involve the renting of storage systems, devices and related infrastructure, usually from a cloud provider like AWS or Microsoft Azure. These rented storage systems can then be connected to and taken advantage of either through the internet or though dedicated private channels.

Unlike with traditional storage, cloud storage can scale itself according to the capacity needs, and enterprises also only have to pay for the storage capacity that they use and not a dime more, allowing them to avoid the problems of under or over-provisioning of hardware resources associated with traditional storage. These factors, combined with the variety of cloud storage types and the convenience of not having to manage any hardware on their own, make cloud storage an appealing choice for businesses with fluctuating storage needs or those aiming to avoid the high upfront costs associated with on-premises infrastructure.

Note however that it is not necessary for organizations to chose one or the other, in fact many organizations often adopt environments that utilize both on-prem and cloud storage solutions in varying capacities, usually keeping frequently accessed or low-latency access data near their facilities; An approach termed as hybrid storage.

Now there are three primary types of cloud storage: **Object Storage**, **Block Storage**, and **File Storage**. All three types, their characteristics and use cases are discussed in detail below.

## Object Storage

Object Storage (as the name may suggest) organizes data by breaking it into pieces called objects, with each object containing the data itself, metadata related to it, and a unique ID. It is highly scalable and well-suited for unstructured data like images, videos, and backups.

Particularly designed for scenarios where quick access to large volumes of data is required, it is ideal for content repositories, backup, and archiving. All the objects are stored in a non-hierarchical manner in a single “folder”, often referred to as the object bucket. Additionally, the metadata associated with the objects allow for the objects to be searched through using say, SQL Queries. It should therefore, not

come as a surprise to many that Object Storage is often used to power big data environments such as data lakes and data warehouses.

## Block Storage

Block storage is a type of cloud storage that resembles a hard drive in traditional computers. It places data into blocks and then stores those blocks as separate pieces with each block having a unique identifier. A hierarchical storage method, it follows the nested file and folder structure similar to desktop file systems, placing the blocks of data wherever most prudent or efficient capacity wise.

## File Storage

File storage is the traditional file systems we run on our desktop machines, just now being run on the cloud i.e. now being run in someone else's data center. It is mostly used to provide shared file systems for a network of devices running the same Operating System. Examples of this include NTFS-based volumes for Windows systems and NFS-based volumes for Linux/UNIX systems.

They are often adopted due to their ability to be mounted and directly accessed by numerous computers and agents at the same time.