

## Unit-5 storage Systems

### TOPICS:

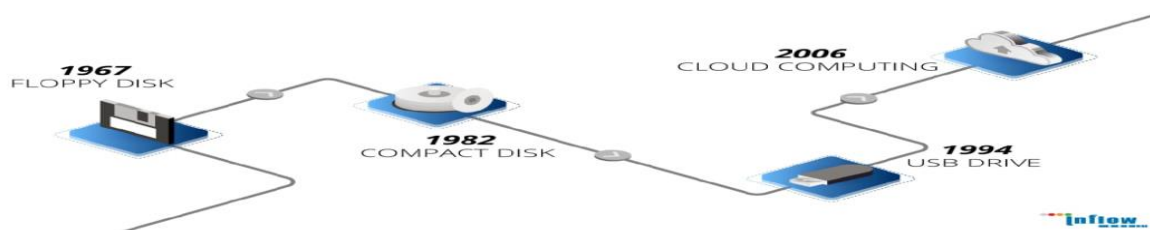
- Evolution Of Storage Technology
- Storage Models
- File Systems And Database
- Distributed File Systems
- General Parallel File Systems
- Google File System

### 1.Evolution Of Storage Technology:

Storage devices have transformed in terms of size and price, allowing us to carry voluminous amounts of

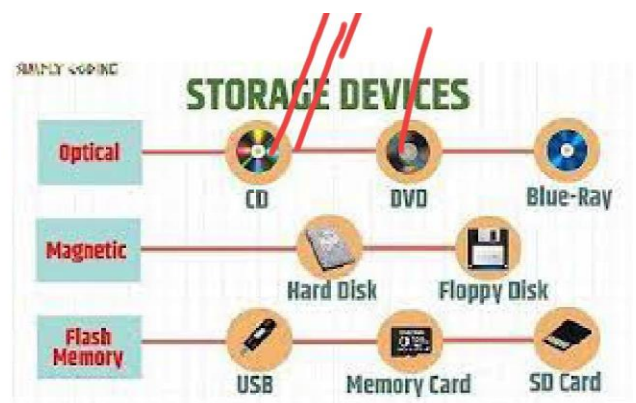
data on devices that can fit in our pockets.

Here's a look at the evolution of storage devices:



Redundant Array of Independent Disks (RAID): This technology was developed to address the cost, performance, and availability requirements of data. It continues to evolve today and is used in all storage architectures such as :

- DAS(Direct Attached Storage)
- SAN(Storage Area Network)
- NAS(Network Attached Storage)
- IP-SAN(Internetprotocol SAN)



## 2. Storage Models

### Database storage models:

- Cloud storage models are models of cloud computing that stores data on the internet via cloud computing providers. These providers manage and operate data storage as a service.
- Cloud storage is basically an online storage of data. Data that is stored can be accessed from multiple connected devices, which constitute a cloud.

### The 3 cloud storage models are:

- Instance storage: Virtual disks in the cloud
- Volume storage: SAN sans the physical
- Object storage: Web-scale NAS
- These models are pioneered in virtualization also providing new models that enable fully virtualized storage stacks.
- The cloud environment tries to provide a self-service with a precise separation between application and infrastructure
- **Instance storage: Virtual disks in the cloud**

In a traditional virtualized environment, the virtual disk storage model is the eminent one. The nomenclature of this model is based upon this very reason, instance storage, meaning storage that is used like conventional virtual disks. Its crucial to note that instance storage is a storage model, not a storage protocol.

- **Volume storage: SAN sans the physical**

Volume storage is also known as block storage. It supports operations like read/write and keeping the system files of running virtual machines. As suggested by its name, data is stored in structured blocks and volumes where files are split into equal-sized blocks. Each block has its own address.

- **Object storage: Web-scale NAS**

Cloud-native applications need space, for storing data that is shared between different VMs. However, often there's a need for spaces that can extend to various data centers across multiple geographies which is catered by Object storage.

### **Cloud Storage Models:**

- Cloud storage enables you to store your business' and customers' data on hosted servers. It eliminates the need to buy and manage your own storage infrastructure by operating through a cloud computing provider who manages your data capacity and security on your behalf, generally through a pay-as-you-go model that is delivered over the internet.
- By storing data such as audio-visual files, applications, and documents in the cloud, the need to purchase and maintain traditional, physical storage infrastructure is eliminated, as the data is stored in the cloud provider's data center and the enterprise pays only for the infrastructure they use.
- **Cloud storage can be separated into four categories:**
  1. Public cloud storage
  2. Private cloud storage
  3. Hybrid cloud storage
  4. Community cloud storage

#### **Public cloud storage:**

- Public cloud storage is a popular route for businesses needing a quick and simple way to store their important data. Users of a public cloud storage solution including those provided by hyper scalers such as AWS can connect over the internet and store data within the cloud provider.

#### **Private cloud storage:**

- Private cloud storage, on the other hand, resides within your network by leveraging physical infrastructure to create virtual servers that store your company's data.

#### **Hybrid cloud storage:**

- Hybrid cloud storage combines elements of both private and public cloud storage solutions. This means businesses have more choice over which data they store in which cloud.

**Community cloud storage:** With community cloud storage, private cloud providers offer their cloud architecture and software to meet the requirements of specific businesses or communities, such as a network of healthcare companies

### **3.FileSystems And Database**

- File systems and databases are crucial components in cloud computing that allow for efficient storage, retrieval, and management of data. File systems provide a way to organize and store files in a structured manner, making it easier for users to access and share data across different devices and locations.
- Databases, on the other hand, are specialized systems for storing, managing, and retrieving structured data efficiently.
- The file system in cloud computing refers to the way data is organized and stored on the cloud servers. It allows users to access, store, and manage files efficiently in a distributed environment.
- Cloud file systems provide scalability, flexibility, and reliability to accommodate the storage needs of different users and applications.
- The database in cloud computing is a structured collection of data that enables users to retrieve, update, and manage information in a systematic way.
- Cloud databases offer features such as data security, high availability, and scalability to handle large amounts of data efficiently.
- Understanding file systems and databases in cloud computing is essential for individuals looking to build scalable and reliable applications in the cloud.

## **4.Distributed File Systems**

- A distributed file system (DFS) is a file system that is distributed on various file servers and locations.
- It permits programs to access and store isolated data in the same method as in the local files.
- It also permits the user to access files from any system.
- It allows network users to share information and files in a regulated and permitted manner.
- DFS's primary goal is to enable users of physically distributed systems to share resources and information through the Common File System (CFS).
- It is a file system that runs as a part of the operating systems
- DFS has two components in its services, and these are as follows:

### **1. Local Transparency**

### **2. Redundancy**

## **History of Distributed File System**

- The DFS's server component was firstly introduced as an additional feature. When it was incorporated into Windows NT 4.0 Server, it was called "DFS 4.1". Later, it was declared a standard component of all Windows 2000 Server editions.

## **Distributed File System Replication(DFSR)**

- Initial versions of DFS used Microsoft's File Replication Service (FRS), enabling basic file replication among servers. FRS detects new or altered files and distributes the most recent versions of the full file to all servers.
- Windows Server 2003 R2 developed the "DFS Replication" (DFSR).

## Working of Distributed File System

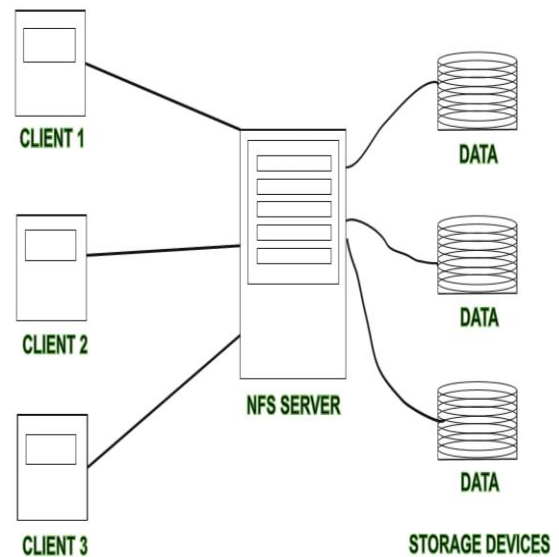
- There are two methods of DFS in which they might be implemented, and these are as follows:

**1. Standalone DFS namespace**

**2. Domain-based DFS namespace**

### There are various features of the DFS:

- Transparency
- Scalability
- Data Integrity
- High Reliability
- High Availability
- Ease of use
- Performance

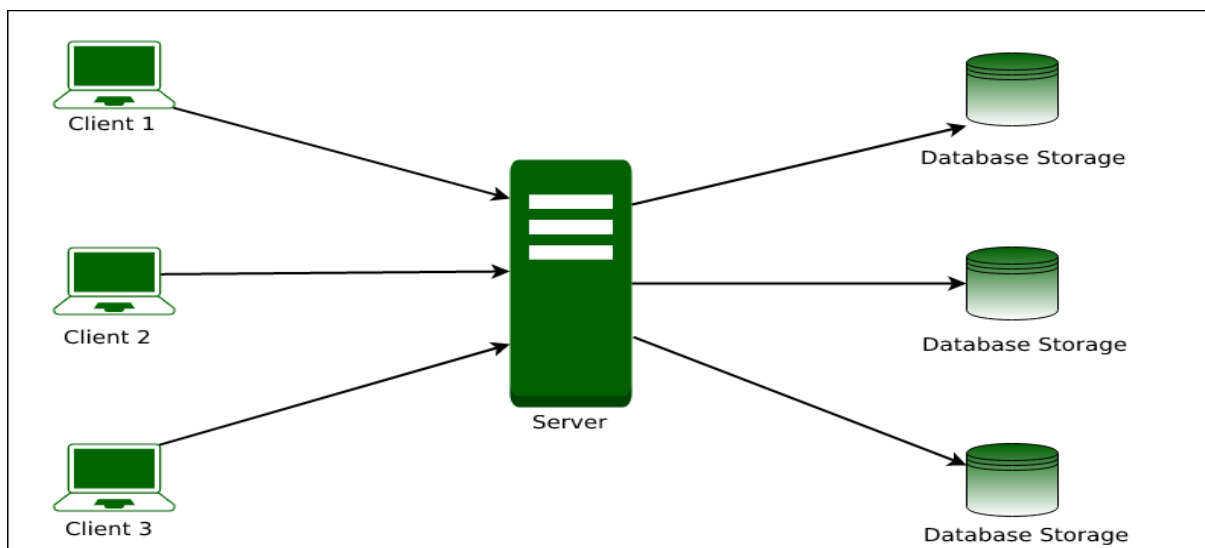


## 5.General Parallel File Systems

- Cloud computing is a popular choice among IT professionals and companies in the digital marketing industry.
- It allows users to access shared resources through the Internet with little to no up-front investment.
- Companies that offer cloud computing services typically charge clients a flat fee per month or yearly contract, but they also might offer free cloud hosting options for individuals who want to try it out before paying for a subscription plan.

### Terminologies in Cloud Computing:

- **Parallel File System:** The parallel file system is a system that is used to store data across multiple network servers.
- It provides high-performance network access through parallel coordinated input-output operations. This is a file system that allows concurrent access to data by more than one user.
- **Flock:** A group of processes (corresponding to a group of threads) sharing the same memory image.
- **Flock semantics:** The properties describe how an entity can be accessed by other processes within the flock when it is not active. In flock semantics, only one process at a time may have exclusive access to an entity and all other processes must share the same view of the entity, even if it is active or protected.



### How PFS Relates to Cloud Computing:

- Cloud computing gives users a lot of freedom to access the data and resources that they need on demand.
- However, when it comes to accessing data, it's important that we shouldn't lose the data from different machines at the same time. Without locking down file system access between different machines, there is a high risk of losing or corrupting important data across multiple computers at once.
- This can make managing files difficult because certain users may end up accessing a file while others are trying to edit it at the same time.

**There are two main types of parallel file systems:**

- Cloud-based parallel file systems.
- Traditional parallel file systems.

## **6.Google File System**

- Google Inc. developed the Google File System (GFS), a scalable distributed file system (DFS), to meet the company's growing data processing needs. GFS offers fault tolerance, dependability, scalability, availability, and performance to big networks and connected nodes.
- GFS is made up of a number of storage systems constructed from inexpensive commodity hardware parts.
- The Google File System reduced hardware flaws while gains of commercially available servers.
- GoogleFS is another name for GFS. It manages two types of data namely File metadata and File Data.
- GFS is made to meet Google's huge cluster requirements.
- Hierarchical directories with path names are used to store files. The master is in charge of managing metadata, including namespace, access control, and mapping data.

### **Components of GFS:**

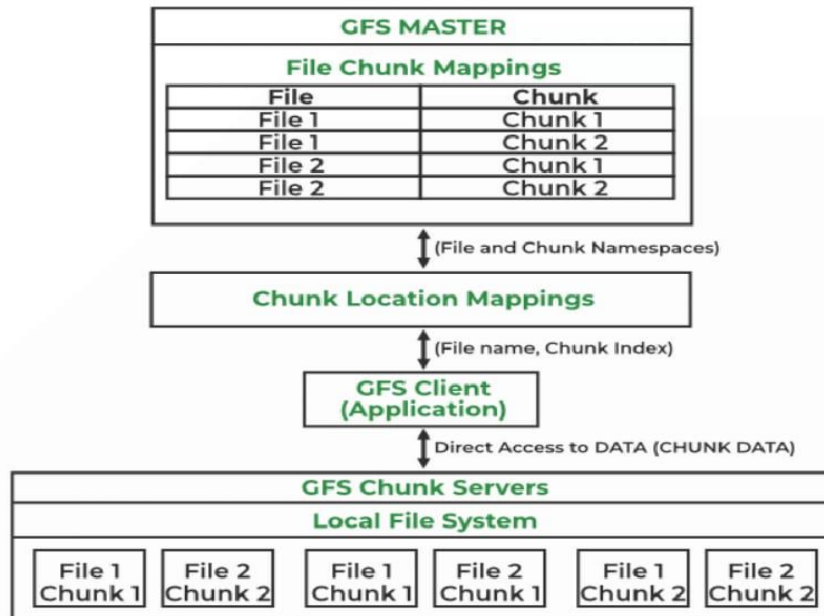
- A group of computers makes up GFS. A cluster is just a group of connected computers. There could be hundreds or even thousands of computers in each cluster.
- 
- **There are three basic entities included in any GFS cluster as follows:**

#### **1.GFS Clients**



## 2.GFS Master server

## 3.GFS Chunck servers



### Features of GFS:

- Namespace management and locking.
- Fault tolerance.
- Reduced client and master interaction because of large chunk server size.
- High availability.
- Critical data replication.
- Automatic and efficient data recovery.
- High aggregate throughput.