

UNIT-V

Storage System:-

- ⇒ Evolution of Storage Technology
- ⇒ Storage models
- ⇒ File Systems and database
- ⇒ Distributed file system
- ⇒ General Parallel file system
- ⇒ Google File System.

Evolution of Storage Technologies:-

- ⇒ Advancements in data storage technology have occurred at a breakneck speed, and the amount of information that is saved annually has been steadily growing.
- ⇒ Storage technology evolved from non-intelligent internal storage to intelligent networked storage.
- ⇒ Highlights of this technology in evolution include:

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, SAN and so on.

Direct-Attached Storage (DAS):-

- ⇒ This type of storage connects directly to a server or a group of servers in a cluster.
- ⇒ Storage can be either internal or external to the server.

Storage Area network (SAN):

- ⇒ this is a dedicated, high-performance Fibre channel network to facilitate block-level communication between servers and storage.
- ⇒ Storage is partitioned and assigned to a server for accessing its data.
- ⇒ SAN offers scalability, availability, performance and cost benefits.

Network Attached Storage (NAS):

- ⇒ this is dedicated storage for file serving applications.
- ⇒ It offers scalability, availability, performance and cost benefits compared to general purpose file servers.

Internet Protocol SAN (IP-SAN):

- ⇒ one of the latest evolutions in storage architecture, IP-SAN.
- ⇒ IP-SAN besides block-level communication across a local or wide area network

Evolution of Storage Devices:-

- 1967 - Floppy Disk
- 1982 - compact Disk
- 1994 - USB Drive
- 2006 - Cloud Computing

Storage models:-

- ⇒ In cloud computing, cloud storage is a virtual locker where we can remotely stash any data.
- ⇒ When we upload a file to a cloud-based server like Google Drive, One Drive, iCloud that file gets copied over the Internet into a data center.
- ⇒ Most companies have hundreds of these servers known as "server farms" spanning across multiple locations.
- ⇒ So, if our data will not lose.

Features of Cloud Storage System:

- ⇒ It has a greater availability of resources.
- ⇒ Easy maintenance.
- ⇒ Has a large network access.
- ⇒ Automatic system.

Storage Systems in the cloud:-

⇒ There are 3 types of storage systems in the cloud:

- * Block - Based Storage System

- * File - Based Storage System

- * Object - Based Storage System.

1. Block - Based Storage System:-

⇒ Hard drives are block-based storage systems.

⇒ Your operating system like windows or Linux actually sees a hard disk drive.

⇒ A drive on which you can create a volume, and then you can partition that volume and format them.

⇒ Ex: If a system has 1000 GB of volume, then we can partition it into 800 GB and 200 GB for local C' and local D' drives respectively.

2. File - Based Storage System:-

⇒ In this, you are actually connecting through a network interface card (NIC).

⇒ You are going over a network, and then you can access the network - Attached storage server (NAS).

⇒ NAS devices are file-based storage systems.

3. Object-Based Storage System:

⇒ In this, a user uploads objects using a web browser and uploads an object to a container i.e., object storage container.

⇒ This uses the HTTP protocols with the rest of the APIs.

⇒ Ex: when you connect to any website, you need to download some images, text, or anything that the website contains.

Advantages of cloud storage

* Scalability

* Flexibility

* Simpler Data migrations

* Recovery

Disadvantages of cloud storage:

⇒ Data centers require electricity and proper internet.

⇒ The long distance is not sufficient physical device.

File systems and database:-

- ⇒ The file system is basically a way of arranging the files in a storage medium like a hard disk.
- ⇒ The file system organizes the files and helps in the retrieval of files when they are required.
- ⇒ File systems consist of different files which are grouped into directories.
- ⇒ The file system performs basic operations like management, file naming, giving access rules, etc.

Eg:- NTFS (New Technology File System).

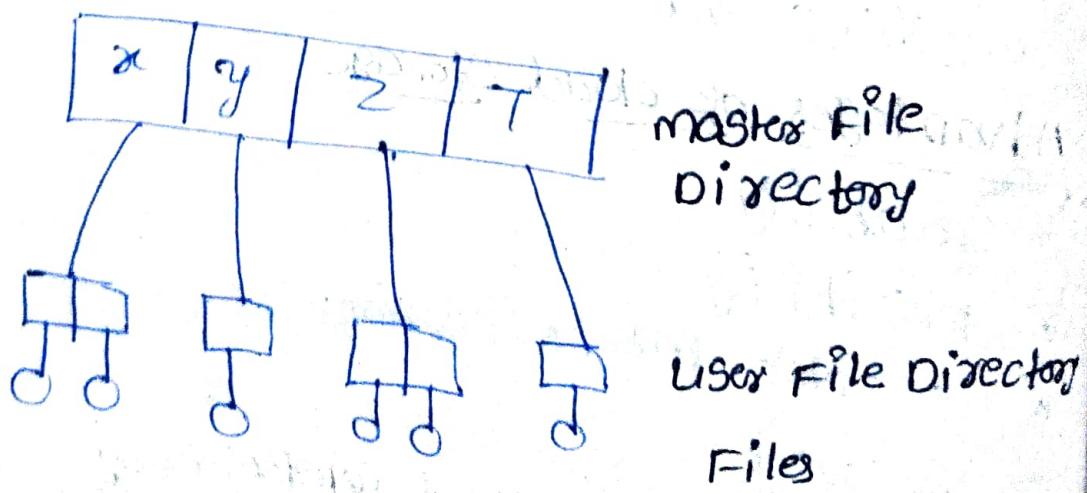


Fig: filesystem

DBMSC Database management system

- ⇒ Database management system is basically software that manages the collection of related data.
- ⇒ It is used for storing data and retrieving the data effectively.
- ⇒ It also provides proper security measures for protecting the data from unauthorized access.
- ⇒ In database management system the data can be fetched by SQL Queries and relational algebra.
- ⇒ It also provides mechanisms for data recovery and data backup.

Eg:- oracle, MySQL, MS SQL Server

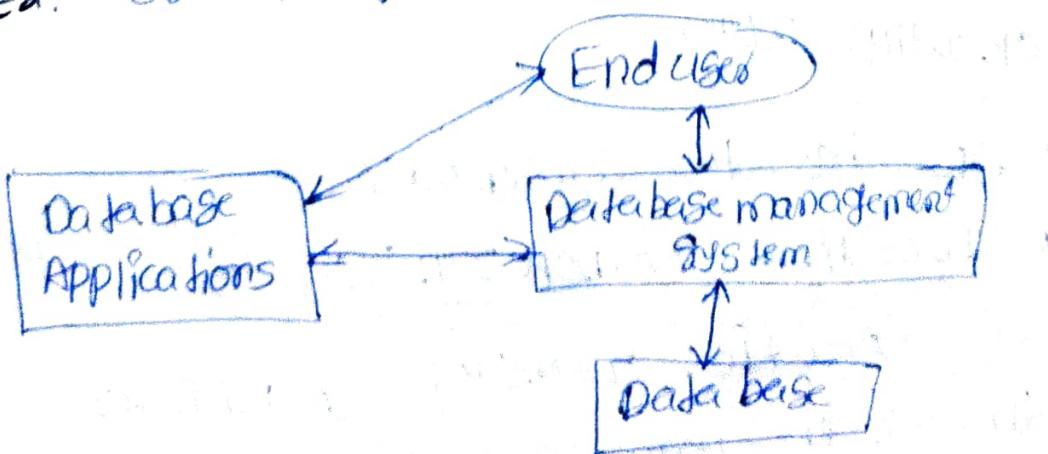


Fig: DBMS

Distributed File System

- ⇒ A Distributed File System as the name suggests, is a file system that is distributed on multiple file servers or multiple locations.
- ⇒ It allows programmers to access files from any network.
- ⇒ The main purpose of the Distributed File System (DFS) is to allow users of physically distributed systems to share their data and resources by using a common file system.
- ⇒ A collection of workstations and mainframes connected by a Local Area network (LAN) is a configuration on Distributed File System.
- ⇒ A DFS is executed as a part of the operating system.
- ⇒ DFS has two components:-
 - * Location Transparency:-
Location Transparency achieves through the name space component.

* Redundancy :-

Redundancy is done through a file replication component.

Features of DFS :-

1. Transparency:

* Structure Transparency:-

multiple File servers should be provided for performance, adaptability and dependency.

* Access Transparency:-

Both local and remote files should be accessible in the same manner.

* Naming Transparency:-

once a name is given to the file, it should not be changed.

* Replication Transparency:-

If a file is copied on multiple nodes, both the copies of the file and their locations should be hidden.

2. User mobility:-

It will automatically bring the user's home

3) Performance:-

Performance is based on the average amount of time needed to complete the client requests.

4) Simplicity and ease of use:-

The user interface of a file system should be simple and the number of commands in the file should be small.

5) High availability:-

The data stored in a cloud storage is always available to users.

6) Scalability:-

The data grows in daily the DFS provide high scalable data facilities.

7) High reliability:-

Data stored in a DFS highly reliable.

8) Data Integrity:-

multiple users frequently share file system
The data can not be change.

9) Security:-

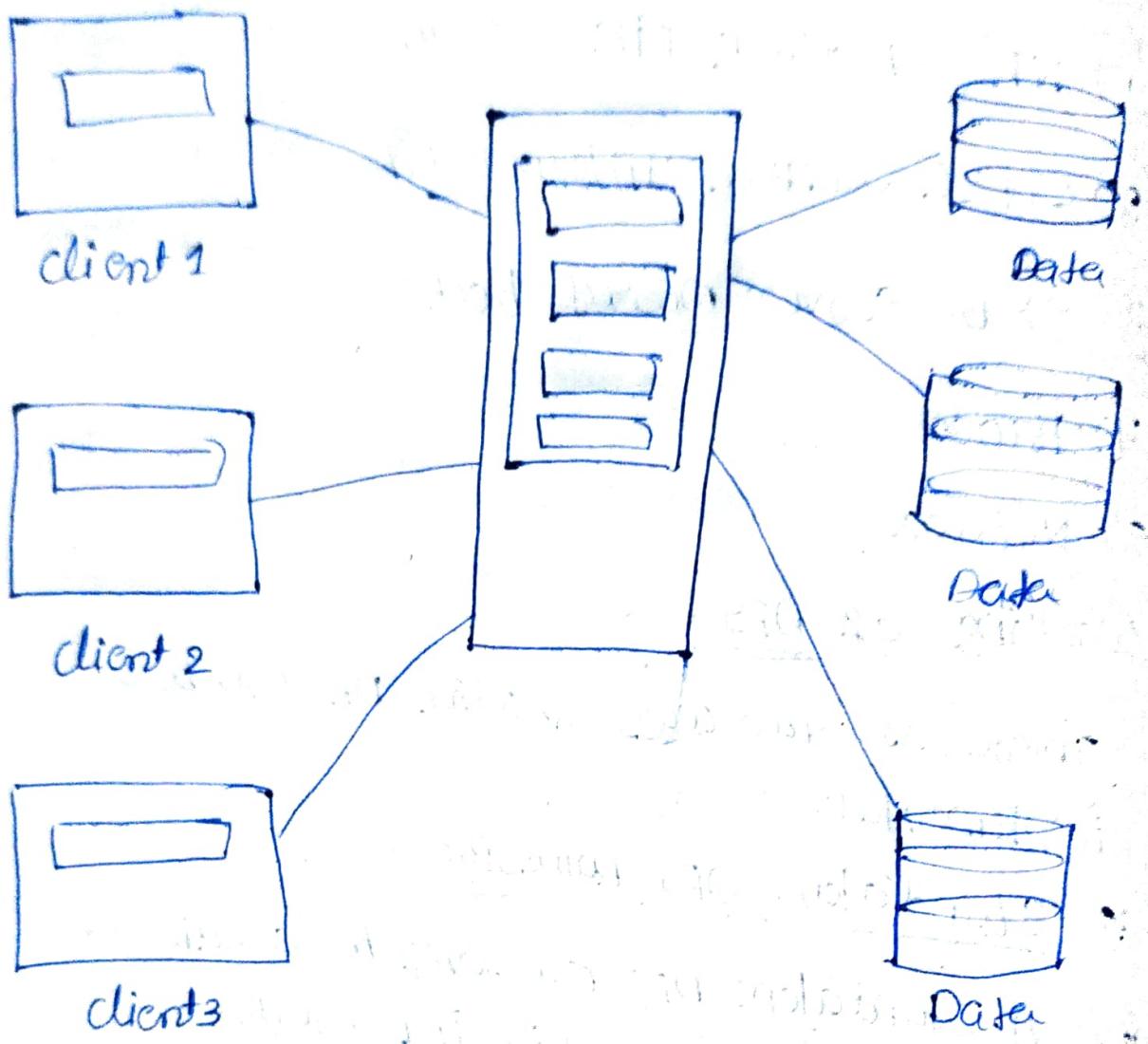
10) Heterogeneity:- unavoidable result on huge scale

Applications:-

- 1) NFS (Network File System)
- 2) CIFS (Common Internet File System)
- 3) SMB (Server Message Block)
- 4) Hadoop
- 5) Netware

Working of DFS:-

- There are two ways in which DFS can be implemented:-
- * Standalone DFS Namespace:-
 - ⇒ A standalone DFS can only be accessed on those computers on which it is created.
 - ⇒ It does not provide any fault isolation and cannot be linked to any other DFS.
 - * Domain-based DFS Namespace:-
 - ⇒ It stores the configuration of DFS in Active Directory.



Advantages:

- DFS allows multiple users to access or store the data.
- It allows the data to be share remotely.
- It improves the availability of file, Access Time, and network efficiency.
- DFS provides Transparency of data

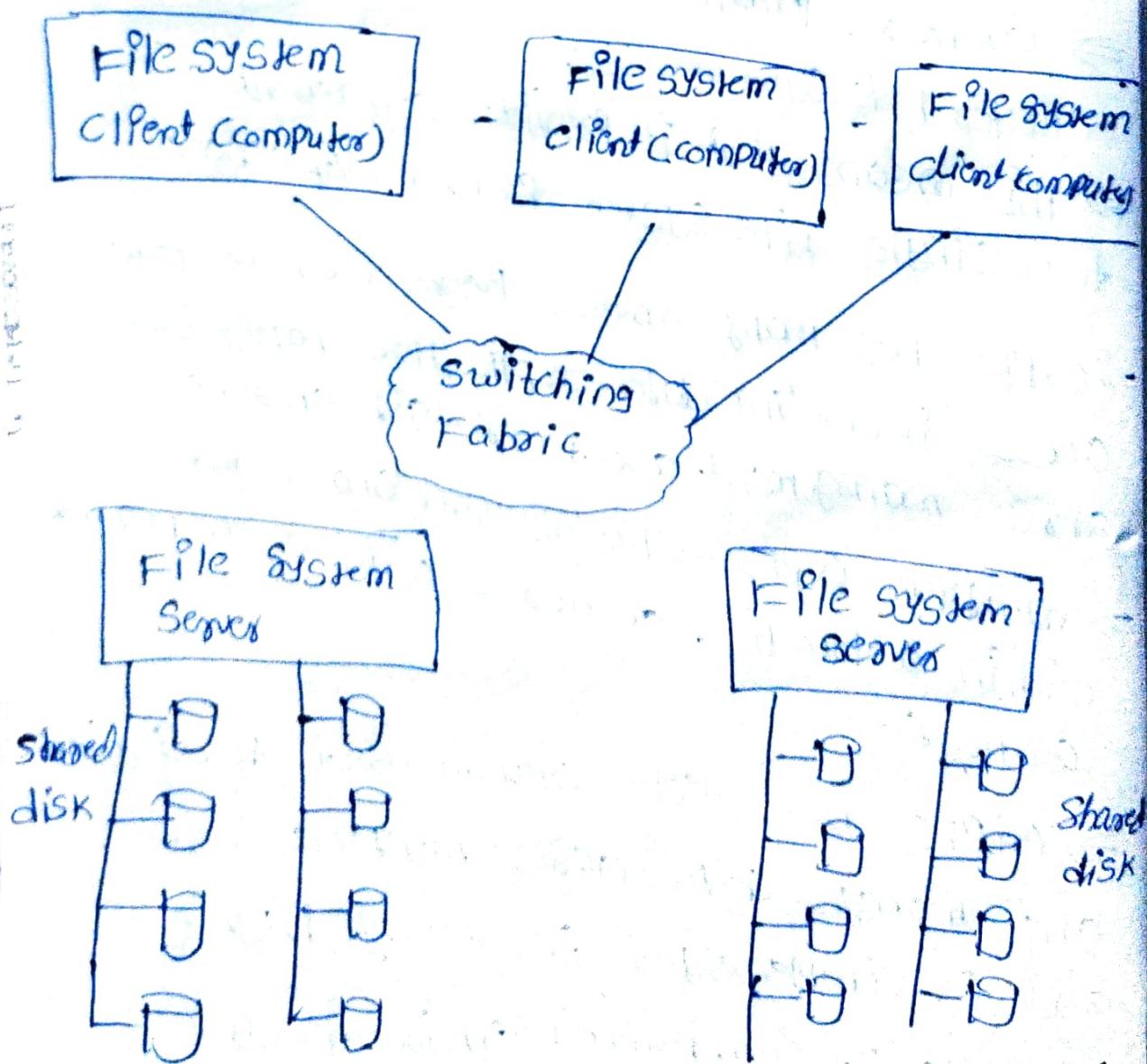
Disadvantages:-

- ⇒ In DFS Nodes and connections needs to be secured.
- ⇒ There is a possibility of loss of messages and data in the network.
- ⇒ Data base connection in DFS is complicated.

General Parallel File System :-

- ⇒ General Parallel File System(GPFS) is a cluster file system.
- ⇒ This means that it provides concurrent access to a single file system from multiple nodes.
- ⇒ GPFS has many features beyond common data access including data replication, policy based storage management and multi-site operations.
- ⇒ Multiple GPFS clusters can share data within a location or across wide area network connections.
- ⇒ GPFS is a file system used to distribute and manage data across multiple servers and is implemented in many high-performance computing environments.

- It is describe as a parallel file system because GPFS data is broken into blocks and striped across multiple disks in an array.
- GPFS also provide other management features such as high availability, replication, mirroring, automation and disaster recovery.



- ⇒ There are two main types of parallel file systems:
- * Cloud-based Parallel File System
 - * Traditional Parallel File System.

Advantages:-

- ⇒ Data Integrity
- ⇒ Data Security
- ⇒ Disaster Recovery

Disadvantages:-

- ⇒ Low Scalability
- ⇒ Less Performance.

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 (errors) while gains of commercially available servers.
- ⇒ GFS manages two types of data namely file metadata and file data.
- ⇒ The GFS node cluster consists of a single master and several chunk servers that various client systems regularly access.
- ⇒ 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 name space, access control, and mapping data.
- ⇒ The master communicates with each chunk server by timed heartbeat messages.

GFS master

File chunk mappings

file	chunk
File 1	chunk 1
File 1	chunk 2
File 2	chunk 1
File 2	chunk 2

(file and chunk name space)

chunk Location mappings

(file name, chunk index)

GFS client
(Application)

(Direct Access to DATA)

GFS chunk servers

Local File System



Components of GFS :-

- * GFS Clients:-

computer Applications which may be used to request Files.

* GFS master Server:-

⇒ It serves as the cluster's coordinator.

* GFS chunk Server:

⇒ They are the GFS's work horses.

⇒ They keep 64MB-sized file chunks.

⇒ The master server does not receive any chunks from the chunk servers.

Features of GFS:-

⇒ Namespace management and locking.

⇒ Fault tolerance.

⇒ High availability.

⇒ Critical data replication.

⇒ Efficient Data recovery.

⇒ High aggregate throughput.

Advantages:-

⇒ High accessibility Data

⇒ Excessive throughput

⇒ Dependable Storing. (Duplication data stored)

Dis Advantages:-

⇒ not the best fit for small files

⇒ unable to type at random

⇒ Suitable for procedures. Data written only once.