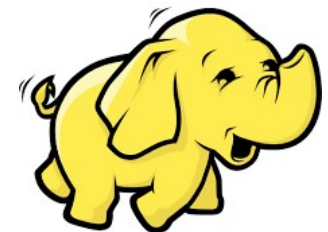


# Distributed File System

Tushar B. Kute,  
<http://tusharkute.com>



# Distributed File System

- A Distributed File System (DFS) as the name suggests, is a file system that is distributed on multiple file servers or multiple locations.
- It allows programs to access or store isolated files as they do with the local ones, allowing programmers to access files from any network or computer.

# Distributed File System

- 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.
- In DFS, a namespace is created and this process is transparent for the clients.

# Distributed File System

- DFS has two components:
  - Location Transparency –
    - Location Transparency achieves through the namespace component.
  - Redundancy –
    - Redundancy is done through a file replication component.

# Distributed File System

- In the case of failure and heavy load, these components together improve data availability by allowing the sharing of data in different locations to be logically grouped under one folder, which is known as the “DFS root”.
- It is not necessary to use both the two components of DFS together, it is possible to use the namespace component without using the file replication component and it is perfectly possible to use the file replication component without using the namespace component between servers.

# Features of DFS

- Transparency :
  - Structure transparency –
    - There is no need for the client to know about the number or locations of file servers and the storage devices. Multiple file servers should be provided for performance, adaptability, and dependability.
  - Access transparency –
    - Both local and remote files should be accessible in the same manner. The file system should be automatically located on the accessed file and send it to the client's side.

# Features of DFS

- Transparency :
  - Naming transparency –
    - There should not be any hint in the name of the file to the location of the file. Once a name is given to the file, it should not be changed during transferring from one node to another.
  - Replication transparency –
    - If a file is copied on multiple nodes, both the copies of the file and their locations should be hidden from one node to another.

# Features of DFS

- User mobility :
  - It will automatically bring the user's home directory to the node where the user logs in.
- Performance :
  - Performance is based on the average amount of time needed to convince the client requests. This time covers the CPU time + time taken to access secondary storage + network access time.
  - It is advisable that the performance of the Distributed File System be similar to that of a centralized file system.



# Features of DFS

- 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.
- High availability :
  - A Distributed File System should be able to continue in case of any partial failures like a link failure, a node failure, or a storage drive crash.
  - A high authentic and adaptable distributed file system should have different and independent file servers for controlling different and independent storage devices.

# Features of DFS

- Scalability :
  - Since growing the network by adding new machines or joining two networks together is routine, the distributed system will inevitably grow over time. As a result, a good distributed file system should be built to scale quickly as the number of nodes and users in the system grows.
- High reliability :
  - The likelihood of data loss should be minimized as much as feasible in a suitable distributed file system. That is, because of the system's unreliability, users should not feel forced to make backup copies of their files.

# Features of DFS

- Data integrity :
  - Multiple users frequently share a file system. The integrity of data saved in a shared file must be guaranteed by the file system. That is, concurrent access requests from many users who are competing for access to the same file must be correctly synchronized using a concurrency control method.
- Security :
  - A distributed file system should be secure so that its users may trust that their data will be kept private. To safeguard the information contained in the file system from unwanted & unauthorized access, security mechanisms must be implemented.

# Features of DFS

- Heterogeneity :
  - Heterogeneity in distributed systems is unavoidable as a result of huge scale. Users of heterogeneous distributed systems have the option of using multiple computer platforms for different purposes.

# History

- The server component of the Distributed File System was initially introduced as an add-on feature. It was added to Windows NT 4.0 Server and was known as “DFS 4.1”.
- Then later on it was included as a standard component for all editions of Windows 2000 Server. Client-side support has been included in Windows NT 4.0 and also in later on version of Windows.
- Linux kernels 2.6.14 and versions after it come with an SMB client VFS known as “cifs” which supports DFS. Mac OS X 10.7 (lion) and onwards supports Mac OS X DFS.

# History

- NFS –
  - NFS stands for Network File System. It is a client-server architecture that allows a computer user to view, store, and update files remotely. The protocol of NFS is one of the several distributed file system standards for Network-Attached Storage (NAS).
- CIFS –
  - CIFS stands for Common Internet File System. CIFS is an accent of SMB. That is, CIFS is an application of SIMB protocol, designed by Microsoft.

# History

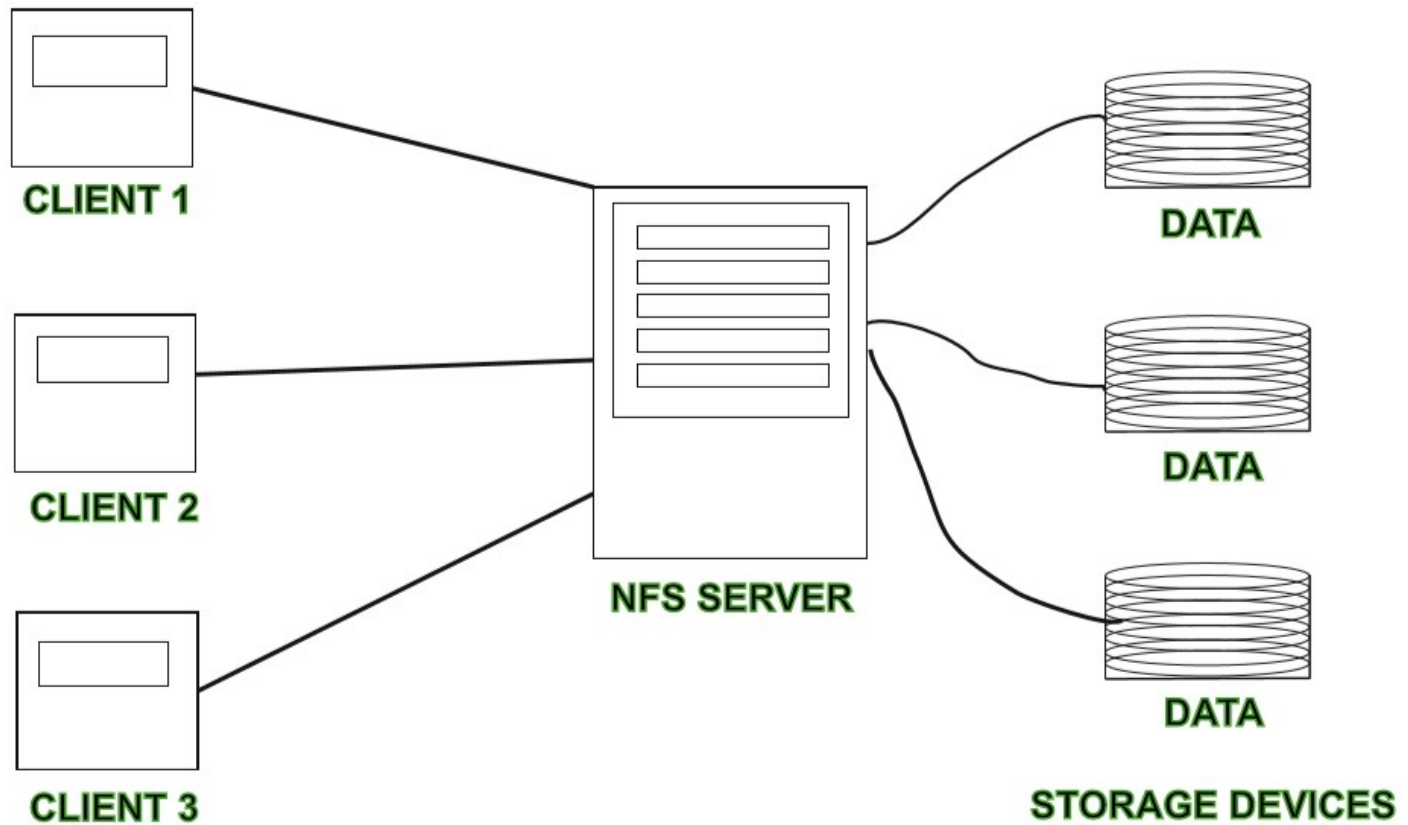
- SMB –
  - SMB stands for Server Message Block. It is a protocol for sharing a file and was invented by IMB. The SMB protocol was created to allow computers to perform read and write operations on files to a remote host over a Local Area Network (LAN). The directories present in the remote host can be accessed via SMB and are called as “shares”.
- Hadoop –
  - Hadoop is a group of open-source software services. It gives a software framework for distributed storage and operating of big data using the MapReduce programming model. The core of Hadoop contains a storage part, known as Hadoop Distributed File System (HDFS), and an operating part which is a MapReduce programming model.

# Working of DFS

- Standalone DFS namespace –
  - It allows only for those DFS roots that exist on the local computer and are not using Active Directory. A Standalone DFS can only be acquired on those computers on which it is created.
  - It does not provide any fault liberation and cannot be linked to any other DFS. Standalone DFS roots are rarely come across because of their limited advantage.
- Domain-based DFS namespace –
  - It stores the configuration of DFS in Active Directory, creating the DFS namespace root accessible at \\<domainname>\<dfsroot> or \\<FQDN>\<dfsroot>



# Working of DFS



# Advantages

- DFS allows multiple user to access or store the data.
- It allows the data to be share remotely.
- It improved the availability of file, access time, and network efficiency.
- Improved the capacity to change the size of the data and also improves the ability to exchange the data.
- Distributed File System provides transparency of data even if server or disk fails.

# Disadvantages

- In Distributed File System nodes and connections need to be secured therefore we can say that security is at stake.
- There is a possibility of loss of messages and data in the network while movement from one node to another.
- Database connection in case of Distributed File System is complicated.
- Also handling of the database is not easy in Distributed File System as compared to a single user system.
- There are chances that overloading will take place if all nodes try to send data at once.

# Thank you

*This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License*



@mitu\_skillologies



/miTuSkillologies



@mitu\_group



/company/mitu-  
skillologies



MITUSkillologies

## Web Resources

<https://mitu.co.in>  
<http://tusharkute.com>

[contact@mitu.co.in](mailto:contact@mitu.co.in)  
[tushar@tusharkute.com](mailto:tushar@tusharkute.com)