Introduction

A **file system** in the **cloud** is a hierarchical storage **system** that provides shared access to **file** data. Users can create, delete, modify, read, and write **files** and can organize them logically in directory trees for intuitive access.

GFS

Google File System (**GFS** or **GoogleFS**) is a proprietary distributed file system developed by Google to provide efficient, reliable access to data using large clusters of commodity hardware. The last version of Google File System codenamed Colossus was released in 2010

GFS is not implemented in the kernel of an operating system, but is instead provided as a userspace library.

GFS

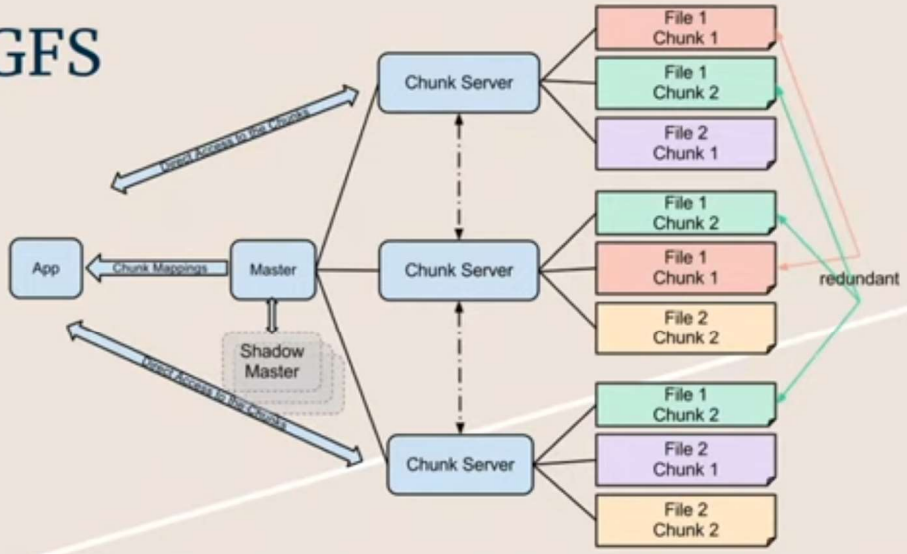
The Master server does not usually store the actual chunks, but rather all the metadata associated with the chunks, such as the tables mapping the 64-bit labels to chunk locations and the files they make up (mapping from files to chunks), the locations of the copies of the chunks, what processes are reading or writing to a particular chunk, or taking a "snapshot" of the chunk pursuant to replicate it (usually at the instigation of the Master server, when, due to node failures, the number of copies of a chunk has fallen beneath the set number). All this metadata is kept current by the Master server periodically receiving updates from each chunk server ("Heart-beat messages").

GFS

Permissions for modifications are handled by a system of time-limited, expiring "leases", where the Master server grants permission to a process for a finite period of time during which no other process will be granted permission by the Master server to modify the chunk.

Programs access the chunks by first querying the Master server for the locations of the desired chunks; if the chunks are not being operated on (i.e. no outstanding leases exist), the Master replies with the locations, and the program then contacts and receives the data from the chunkserver directly

GFS



Google File System is designed for system-to-system interaction, and not for user-to-system interaction. The chunk servers replicate the data automatically.

Dr. Neelesh Jain

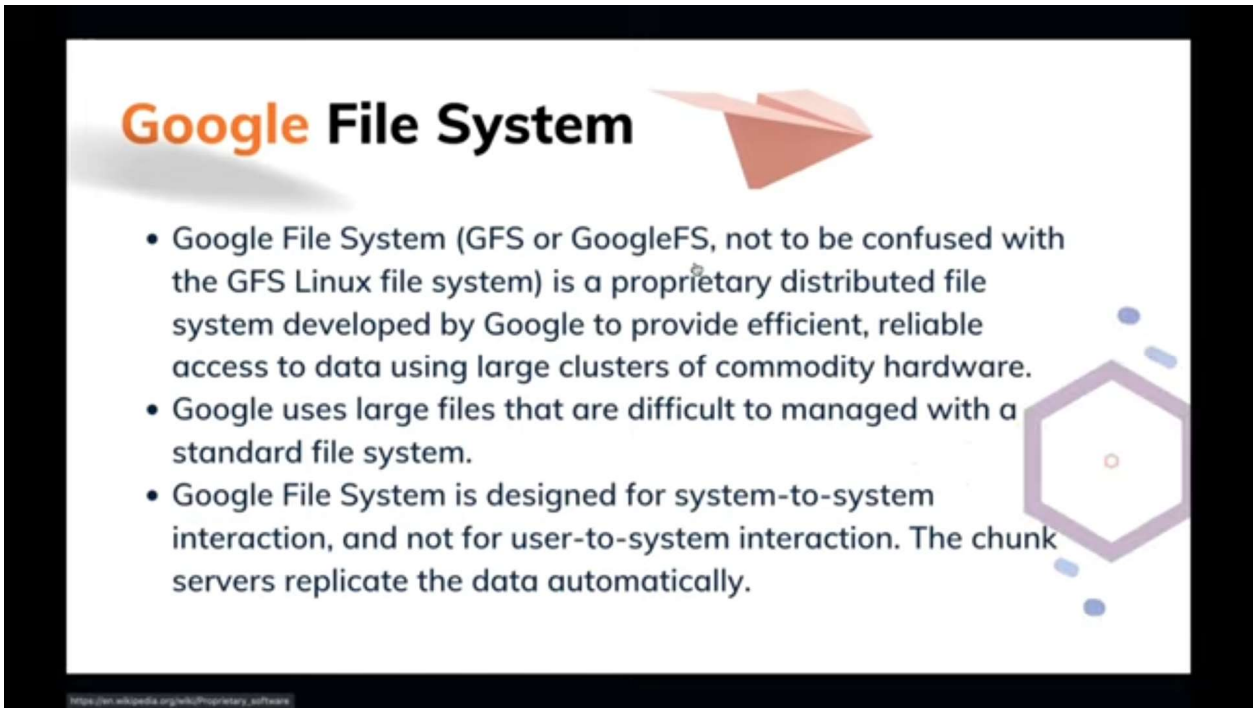
GFS Features

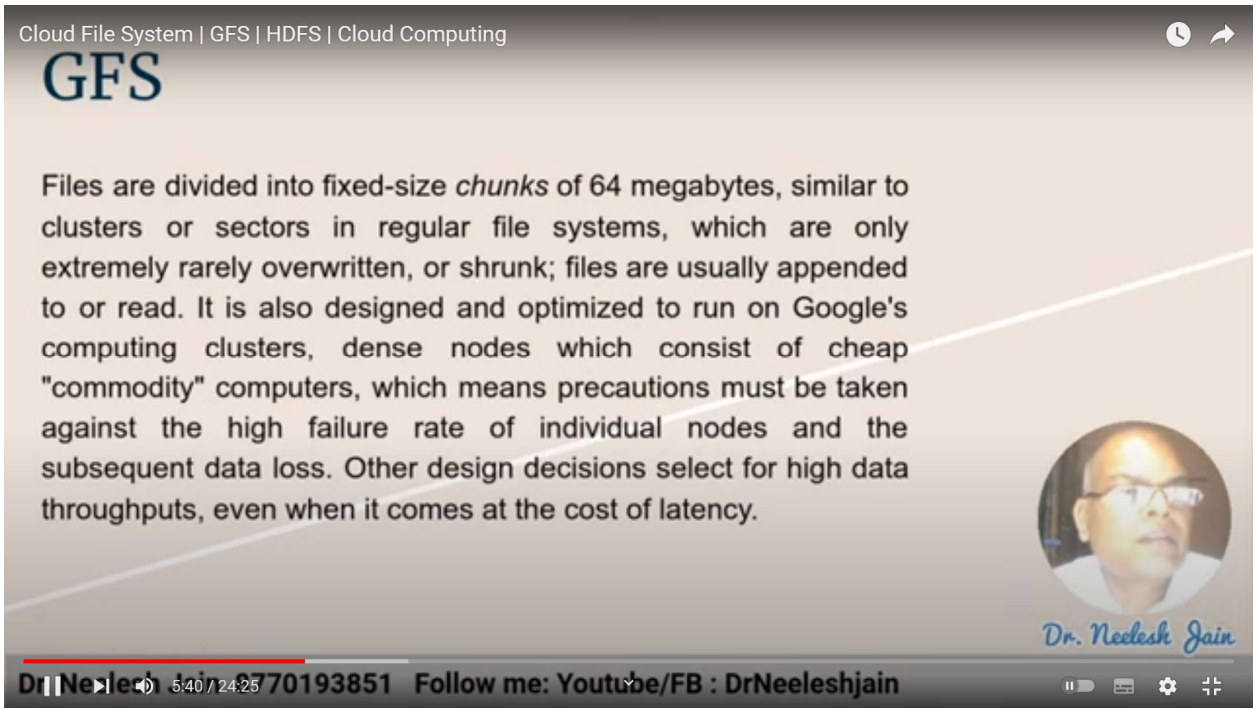
GFS features include:

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

The largest GFS clusters have more than 1,000 nodes with 300 TB disk storage capacity. This can be accessed by hundreds of clients on a continuous basis.

Dr. Neelesh Jain


Google File System
• Google File System (GFS or GoogleFS, not to be confused with the GFS Linux file system) is a proprietary distributed file system developed by Google to provide efficient, reliable access to data using large clusters of commodity hardware.
• Google uses large files that are difficult to managed with a standard file system.
• Google File System is designed for system-to-system interaction, and not for user-to-system interaction. The chunk servers replicate the data automatically.


Cloud File System | GFS | HDFS | Cloud Computing
GFS
Files are divided into fixed-size chunks of 64 megabytes, similar to clusters or sectors in regular file systems, which are only extremely rarely overwritten, or shrunk; files are usually appended to or read. It is also designed and optimized to run on Google's computing clusters, dense nodes which consist of cheap "commodity" computers, which means precautions must be taken against the high failure rate of individual nodes and the subsequent data loss. Other design decisions select for high data throughputs, even when it comes at the cost of latency.
Dr. Neelesh Jain
5:40 / 24:25
Follow me: Youtube/FB : DrNeeleshjain