Vidya Academy of Science & Technology, Thrissur

Google File System

CS402

Mr. Sivadasan E.T.
Associate Professor
Computer Science & Engineering

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.

Google File System

- 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.

Google File System

- GoogleFS is another name for GFS. It 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.

- On local discs, chunk servers keep data in the form of Linux files.

Google File System

- Large (64 MB) pieces of the stored data are split up and replicated at least three times around the network.
- Reduced network overhead results from the greater chunk size.

Google File System

- Without hindering applications, GFS is made to meet Google's huge cluster requirements.

- Hierarchical directories with path names are used to store files.

Google File System

- The master is in charge of managing metadata, including namespace, access control, and mapping data.

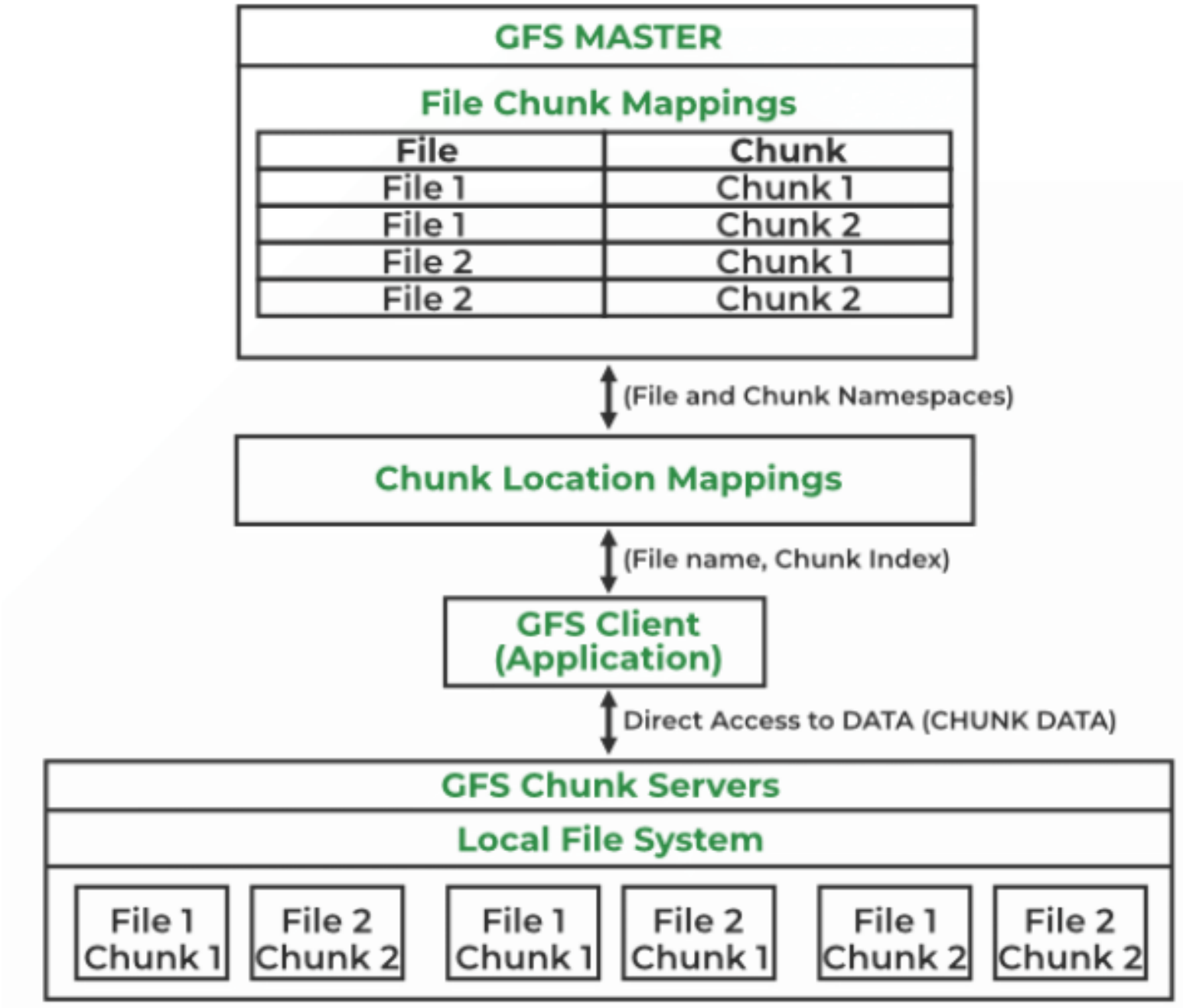
- The master communicates with each chunk server by timed heartbeat messages and keeps track of its status updates.

Google File System

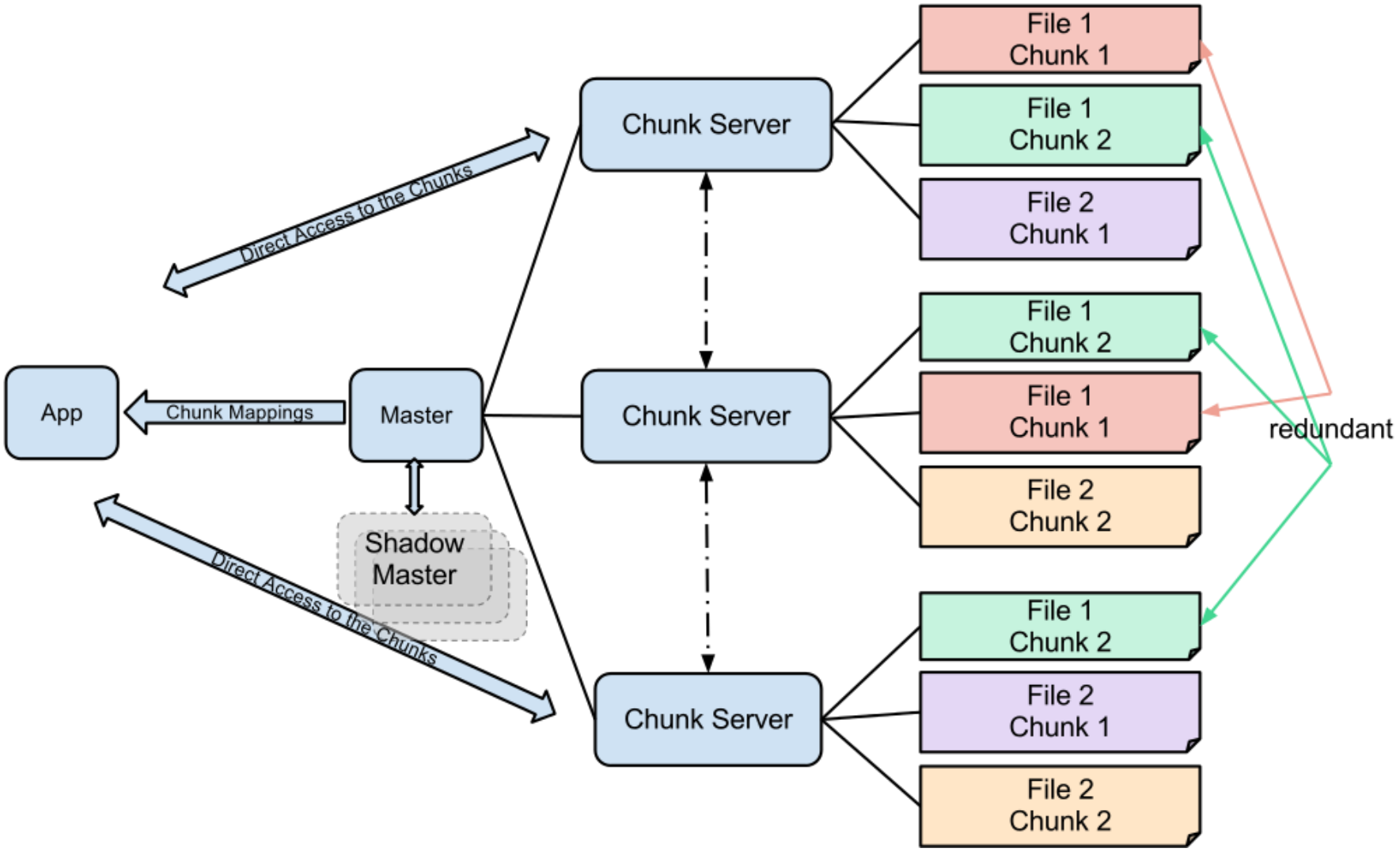
- More than 1,000 nodes with 300 TB of disc storage capacity make up the largest GFS clusters.

- This is available for constant access by hundreds of clients.

Google File System



Google File System



Components of GFS

GFS Clients:

They can be computer programs or applications which may be used to request files.

Requests may be made to access and modify already-existing files or add new files to the system.

Components of GFS

GFS Master Server:

- It serves as the cluster's coordinator.
- It preserves a record of the cluster's actions in an operation log.
- Additionally, it keeps track of the data that describes chunks, or metadata.
- The chunks' place in the overall file and which files they belong to are indicated by the metadata to the master server.

Components of GFS

GFS Chunk Servers:

- They are the GFS's workhorses.
- They keep 64 MB-sized file chunks.
- The master server does not receive any chunks from the chunk servers.
- Instead, they directly deliver the client the desired chunks.
- The GFS makes numerous copies of each chunk and stores them on various chunk servers in order to assure stability; the default is three copies.

Components of GFS

GFS Chunk Servers:

- They are the GFS's workhorses.
- They keep 64 MB-sized file chunks.
- The master server does not receive any chunks from the chunk servers.
- Instead, they directly deliver the client the desired chunks.
- The GFS makes numerous copies of each chunk and stores them on various chunk servers in order to assure stability; the default is three copies.

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.

Thank you !..