|  |  |
| --- | --- |
| Name: Sayyed Sohail Rashid | Course Name: DC-LAB |
| Class: BE-CO | Batch: 01 |
| Roll no: 18CO48 | Assignment No: 02 |

**Aim** : Case Study on Google File System(GFS).

**Theory**:

**What is Google File System?**

- Google ﬁle system is a scalable distributed ﬁle system developed by Google to provide

efﬁcient and reliable access to data using large clusters of commodity hardware.

- It is designed to meet the rapidly growing demand of Google's data processing need.

- It provides performance, scalability, reliability and availability of data across distributed

System for handling and processing big data.

**Characteristics:**

1. Files are organized hierarchically in directories and identiﬁed by path name.

2. It supports all the general operations on ﬁles like read, write, open, delete and so on.

3. It provides atomic append operation known as record append.

4. The concurrent writes to the same region are not serializable.

5. It performs two operations: snapshot and record append.

**Goals**:

1. Performance

2. Reliability

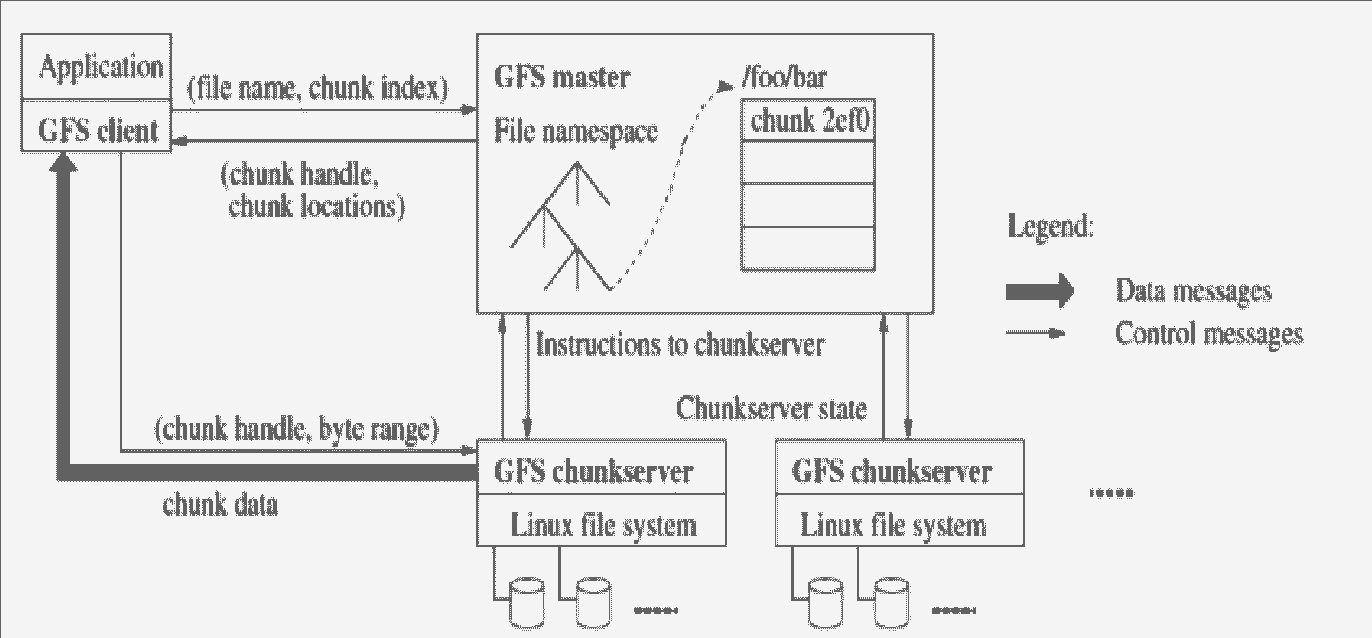
3. Automation

4. Fault Tolerance

5. Scalability

6. Availability

**GFS Architecture:**

****

1. Master Node:

* It is responsible for the activities of the system such as managing chunk leases, load balancing and so on.
* It maintains all the file system metadata.
* It contains an operation log that stores namespaces and files to chunk mappings.
* It periodically communicates with chunk server to determine chunk locations and assesses state of the overall system.
* Each node on the namespace tree has its own read-write lock to manage concurrency.

2. Chunk and Chunk Server

* The files are divided into fixed sized chunks.
* Each chunk has an immutable and globally unique 64-bit chunk handle.
* Chunk server is responsible for storing chunks on local disk as linux files.

3. Client Node

* Client node is linked with the application that implements GFS API.
* It communicates with the master and the chunk server to read chunk server

# **Operation Log and MetaData**:

* Operation log is the persistent records of metadata.
* It defines the logical timeline about serialized order of concurrent operations.
* The state is recovered by the master by replaying the operation log.
* The metadata stored in GFS master are as follows:

1. Namespace (directory hierarchy)
2. Access control information per file
3. Mapping from file to chunk
4. Current location of chunks (Chunk servers)

**Conclusion**:

We have Successfully Performed the Case Study on Google File System.