

# Paper Note for Week 5

Paper: The Google File System

The paper covers four major challenges and motivations to build the distributed file system which is the google file system.

The first problem is the component failure. The host of the file system is made from hundreds or thousands of machines build from commodity hardware. The failure of the machine is inevitable. Therefore, it is required to build a system that can perform constant monitoring, error detection and fault tolerance.

The second problem is that nowadays the files in the file system are getting large. As the result, things like I/O operation and block sizes need to be redesigned.

The third problem is that the mutation of files is often being appended rather than overwritten. Therefore, the new file system needs to optimize for appending performance while guaranteeing atomicity.

The fourth motivation is that it is good to co-design the applications and the file system API since it benefits the overall system by increasing flexibility.

These four problems, namely component failure, large files, append-only mutation and co-design, conclude the main motivations to build the google file system.