Summary of “Google File System”

Hou, Jue

014695647

[jue.hou@helsinki.fi](mailto:jue.hou@helsinki.fi)

In the article “Google File System”, authors introduced the basic mechanism of Google File System (GFS) and main assumptions used for design. Based on those assumptions, such as the constant failing of the system caused by cheap components, the designers make GFS not only to provide the same fundamental features as other distributed systems do, but also to get simplified and optimized in some key implementation. There are two different kinds of server in the GFS: one single logical master server and a large number of chunk servers for data storage. Client will only have to interact with master once and then it can have the direct access to the chunk servers. And after client pushes the data, it will be passed from one chunk server to another but closest chunk server. On the other hand, master server will keep the record of metadata of the system and be responsible for the consistency of the data in the long term, the management of the chunk and other important functions.

Overall, authors introduced GFS from different aspect thoroughly. Although the mechanism of some features such as the region states after mutation and locking turn out to be a little bit confusing for me, I still manage to grab a general idea of those features with the help of other articles.