



Criticism on GFS

Here is the summary of criticism on GFS's architecture.

We'll cover the following



- Problems associated with single master
- Problems associated with large chunk size

Problems associated with single master#

As GFS has grown in usage, Google has started to see the following problems with the centralized master scheme:

- Despite the separation of control flow (i.e., metadata operations) and data flow, the master is emerging as a bottleneck in the design. As the number of clients grows, a single master could not serve them because it does not have enough CPU power.
- Despite the reduced amount of metadata (because of the large chunk size), the amount of metadata stored by the master is increasing to a level where it is getting difficult to keep all the metadata in the main memory.

Problems associated with large chunk size#



Large chunk size (64MB) in GFS has its disadvantages while reading. Since a small file will have one or a few chunks, the ChunkServers storing those chunks can become hotspots if a lot of clients are accessing the same file. As a workaround for this problem, GFS stores extra copies of small files for distributing the load to multiple ChunkServers. Furthermore, GFS adds a random delay in the start times of the applications accessing such files.

[< Back](#)[Garbage Collection](#)[Next >](#)[Summary: GFS](#)[Mark as Completed](#)[Report an Issue](#)