

Summary: The Google File System

The Google File System (GFS) was created to meet Google's needs to process data. Its main goals are performance, scalability, reliability, and availability. GFS was designed to account for the following: component errors being the norm, large multi-GB file sizes, and files being mutated by appending instead of overwriting data. The assumptions for the system include: that the system will be built from commodity components (similar to MapReduce), it will store a modest number of large files, it consists of streaming and random reads, there are large sequential writes that append data to files, and that high sustained bandwidth takes priority over low latency. GFS is built with a single master and many chunkservers, and can be accessed through multiple clients. Chunks are 64MB in size, which reduces the client's need to interact with the master. The master stores the file and chunk's namespaces, the mapping from files to chunks, and the location of each chunk's replicas.