Title of Paper: The Google File System
Authors: Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung

This paper is focused on the Google File System(GFS), a distributed file system for big data applications that provides constant monitoring, error detection through checksum, fault tolerance, automatic recovery through chunk replication. The GFS is created based on a few observations, such as files are generally mutated by appending and usually 100MB or larger, also the fact that throughput is more essential than latency. With these assumptions in mind, the GFS is build with a single master and multiple chunkservers, with metadatas such as namespaces, file to chunk mapping, and replica location stored in the master's memory. Persistent records of replica location are constantly polled from chunkservers through heartbeat messages. The GFS is also designed with a relatively high file system block size at 64MB, which provides several advantages that lower the workload and overhead of the system. Using inexpensive and scalable components, GFS has successfully fulfilled Google's storage needs and is used as the processing and R&D storage platform within Google.