Google File System
Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung

Mohammed

11/25/2013

Main Idea

- Use a system which can work on inexpensive commodity hardware

- Designed to provide fast access to data using commodity hardware.

Implementation

- Files are divided into fixed-size chunks (64 bit chunks)
- Each chunk is replicated on multiple chunk servers, and replicated 3 times by default.
- Make the master operations fast by using memory.

Analysis

- Very clever system to store huge amounts of data.
- Even though data is cloned multiple times it always guarantees that data is never destroyed.

Advantages

- Chunk file reduces the client's need to interact with the master.
- Metadata is stored in the memory, which makes master's operations very fast.

Disadvantages

- Lazy space allocation
- Hotspots – allowing many clients to access same file.

Real World Uses

- The system can be used to maintain large amounts of data.
- System only suitable for Google’s needs