

# Distributed OS

Hermann Härtig

## File System Scalability

(NFS → AFS →) GFS

# Network File System (NFS, SUN v2)

## Design Principles

- Interface: as close as possible to Unix
- names/"open": message to file server for each part of pathname
- Read/Write: use block cached in client
- Cache: blocks
- Replication: none (only very rudimentary)
- Consistency: exchange consistency messages every few s
- Faults: write through on server (v2),  
hold up client when server crashes

# Andrew File System

## Design Principles

- Interface: as close as possible to Unix
- names/"open": name resolution on client using caches directories
- Read/Write: use file cached in client
- Cache: whole files, use server replicates to fetch files to cache
- Replication: some (performance)
- Consistency: explicit invalidation of cache through "call backs"
- Faults: ?

# Google File System

## Design Principles

- Interface: specialized for google applications
- names/"open": name resolution on "Master server"
- Read/Write: messages with "chunk servers"
- Cache: files too large to cache, some metadata are cached
- Replication: "chunks" are replicated
- Consistency: relaxed consistency (replicates need not be identical)
- Faults: all servers regularly crash → replicates and fast start

Details: see slides by authors and paper

---

# Architecture

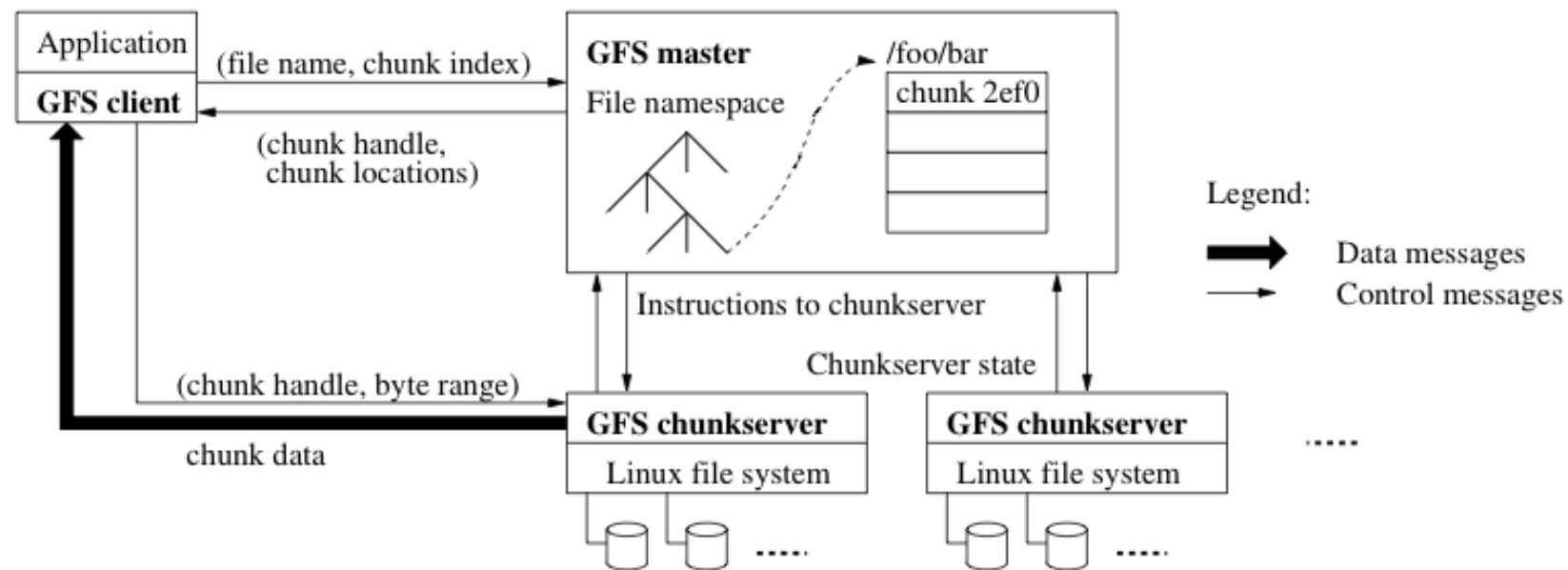


Figure 1: GFS Architecture

Figure from Ref 2(see last slide)

# Write

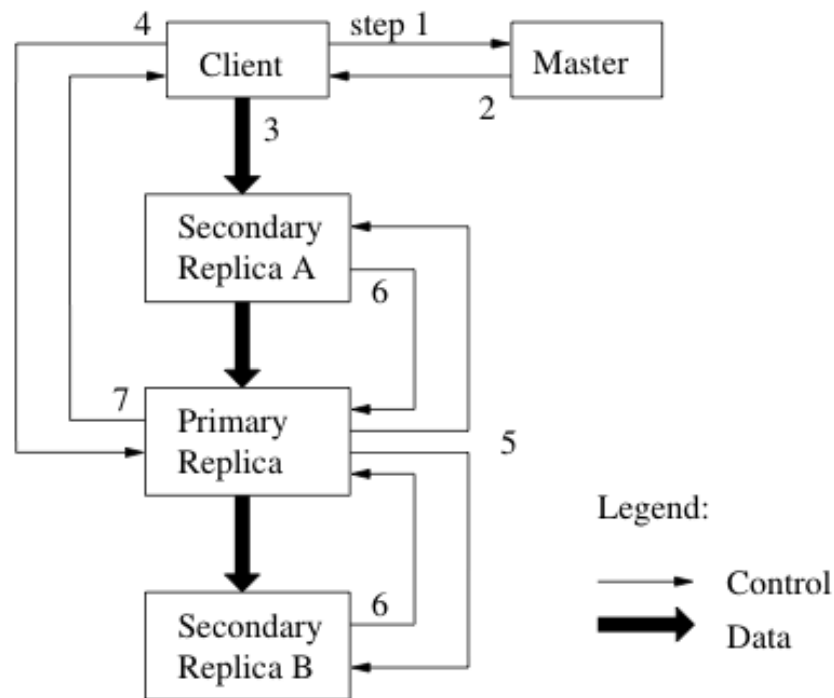


Figure 2: Write Control and Data Flow

Figure from Ref 2(see last slide)

# Write 2

- No concurrency control → inconsistencies
- But Atomic “Append” with at least once semantics
- Replicates not identical → checksum per chunk-replicate
- Applications' data structures must be designed to tolerate this

# Replicas

- Replicas are not always identical
- Replicas may become stale → version numbers
- Master stores current version in operation log
- Chunk servers register with master on start up



# References

1) NFS/AFS: many text books

2) Google FS:

The Google File System

Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung

SOSP 2003