GFS (SOSP '03) MR OSDI '04.

Goal

Shared : M/R HTML

Performance.

Fault tolerance.

1 failure / year * machine

300TB : 1000 machines ⇒ 3 failure / day

GFS Approach.

- FS-like API. (POSIX) mount.

  "lib" < write / append / read

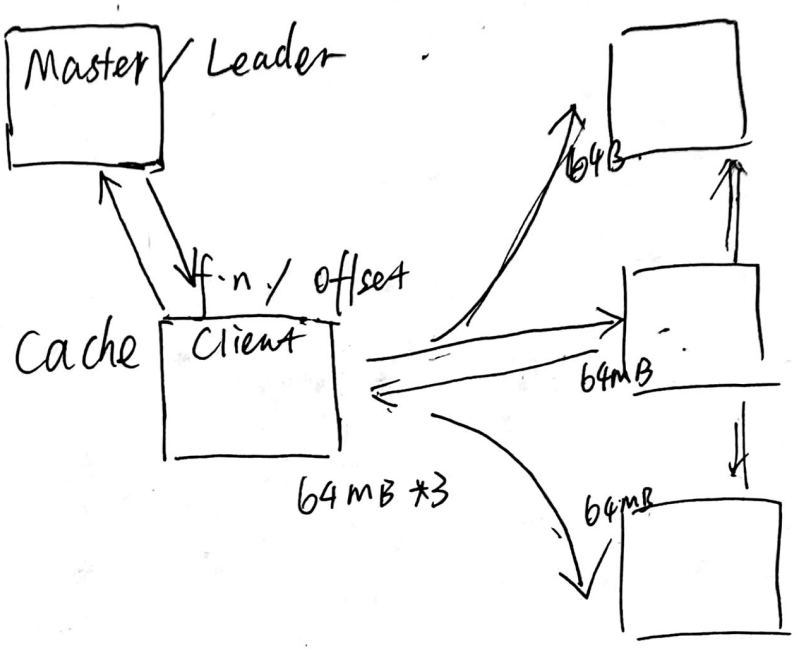
- Single Master. / Meta data servers..

  filename, offset → chunks

- 64 MB (2^26) 16,000 vs. 4 kB (2^12). { page size / block size disk

- 3-way Replication.

How GFS works



Performance

Why single master + 64 MB sufficient?

large files. sequential read/write

Not good for

small files. → aggregate

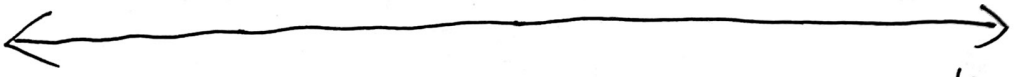
random read/write git compile

Consistency.

correctness : outcome = expectation

concurrent failures.

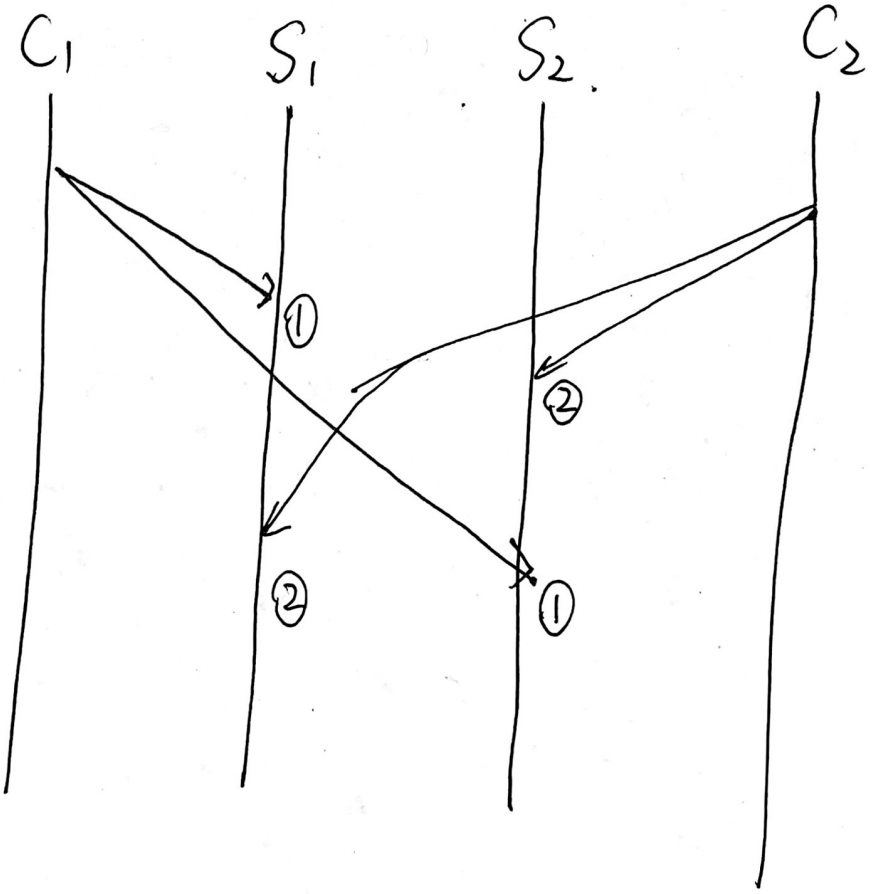
Weak. Strong

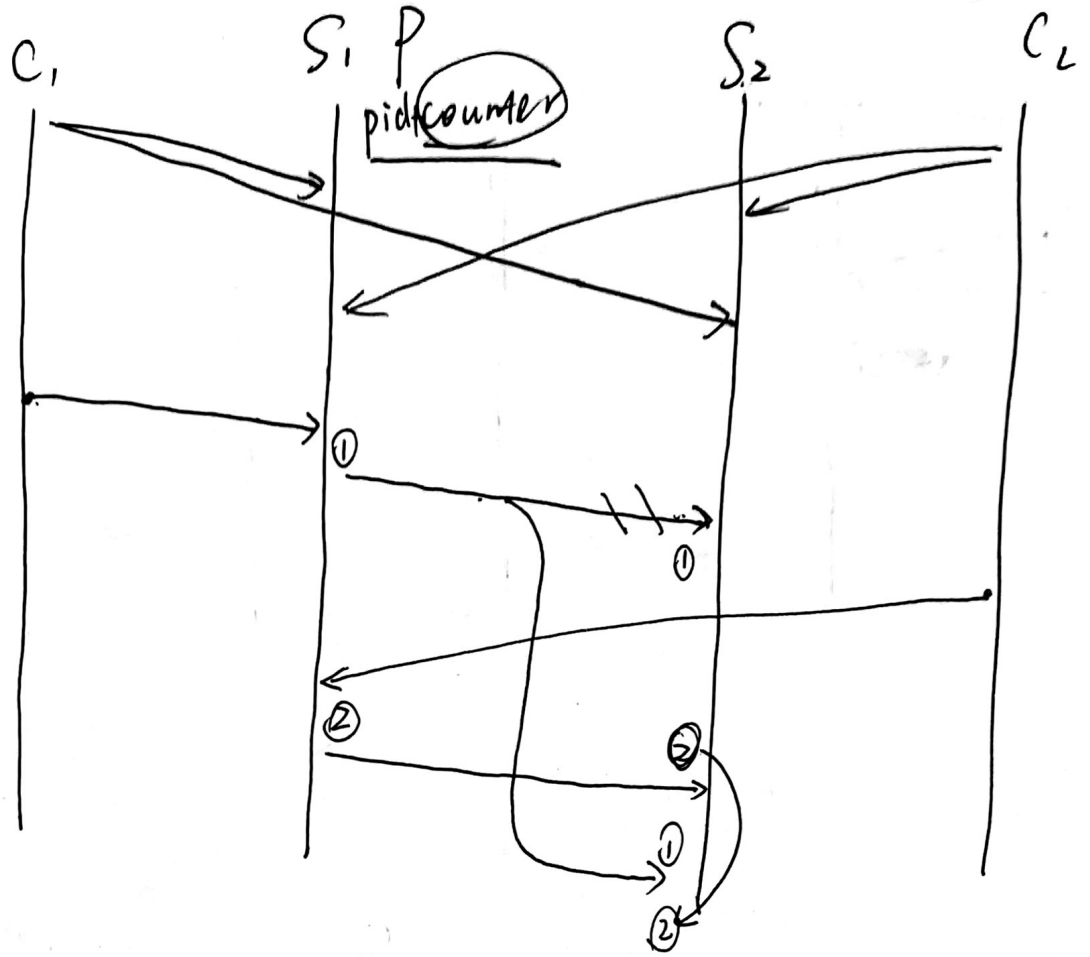


implement

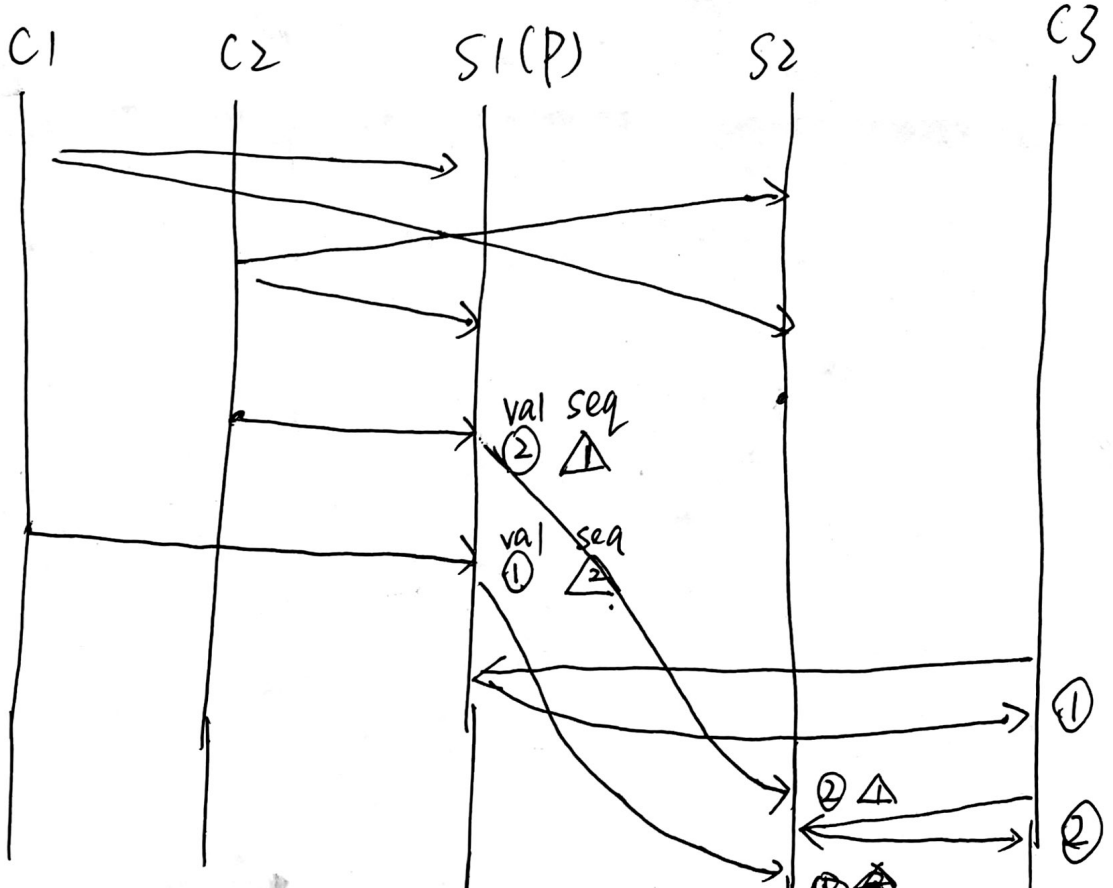
linear
/ no concurrent
/ no failures

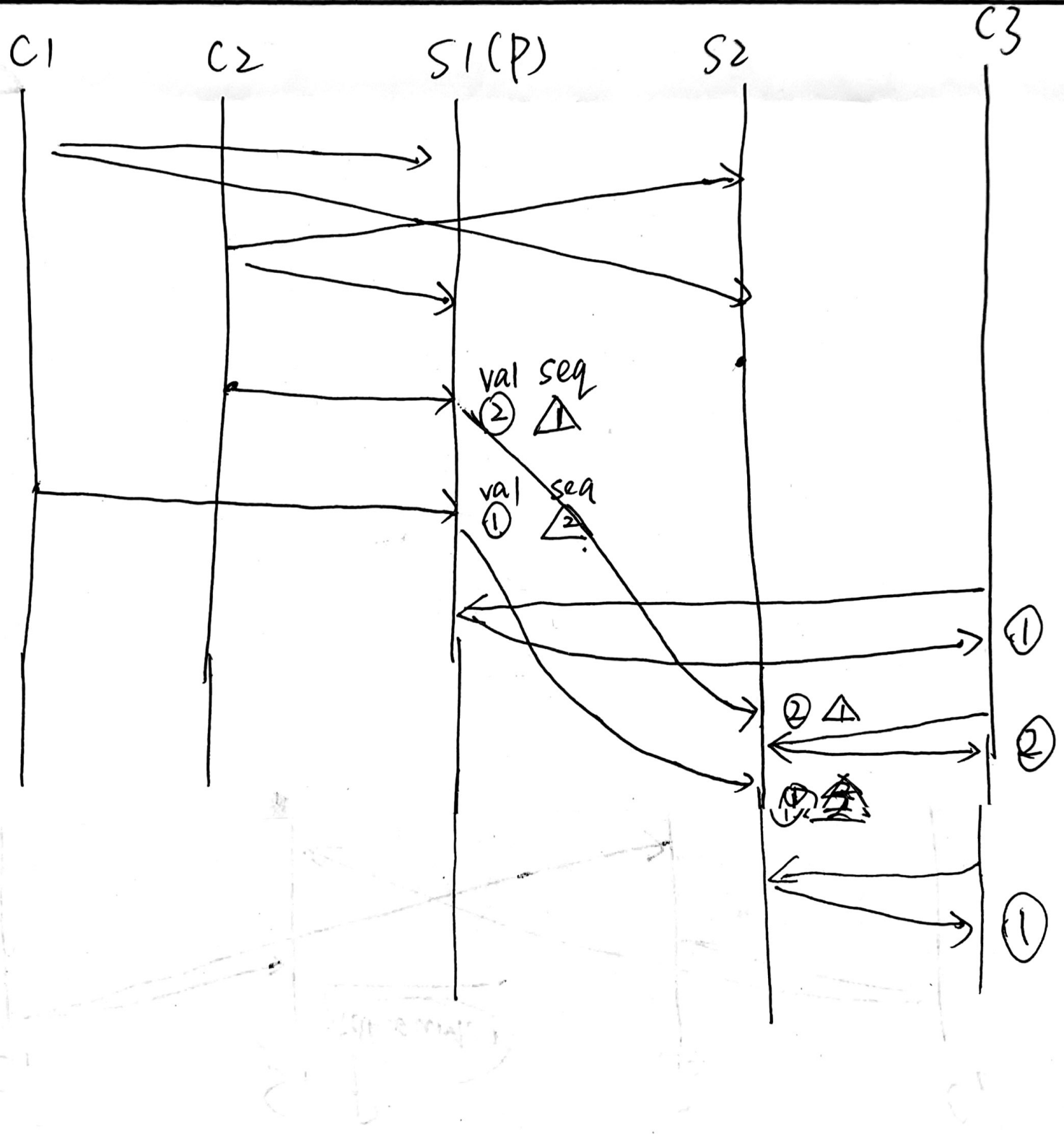
use





W1 W2. $R^3(1)$ $R^3(2)$ $R^3(1)$




C1
C2
S1(P)
S2
C3
val
seq
2
1
val
seq
1
2
1
2
2
1
1