

Venti

- Goal?**
- Write-once archival store
 - keep all snapshots forever

Primary technique? Fingerprint hash to id data

- Advantages?**
- 1) Coalesce duplicate blocks
 - 2) No extra space overhead for full backup vs. incremental
 - 3) Allows for integrity checking

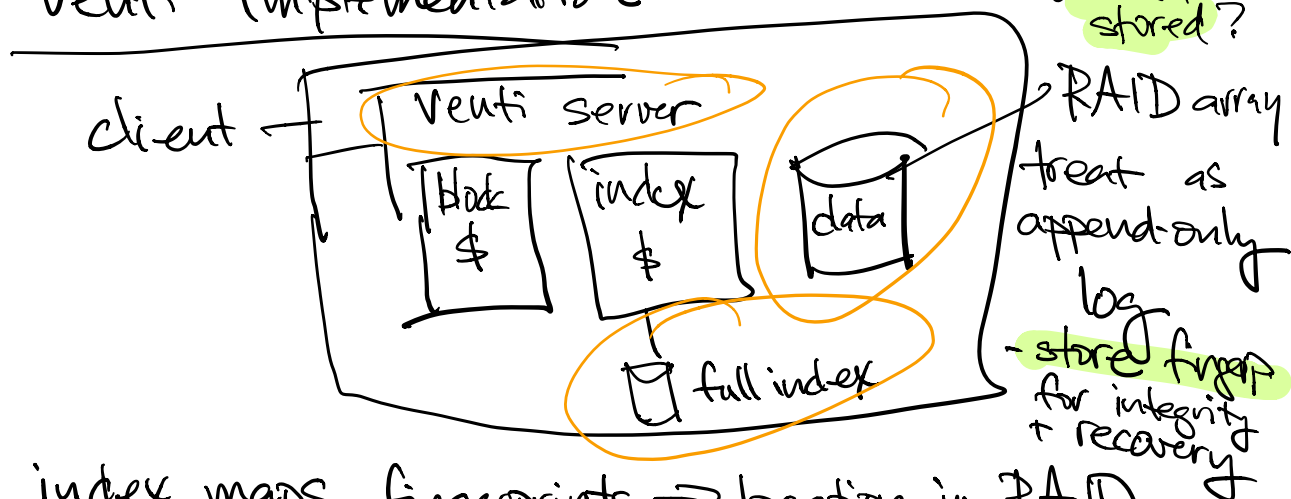
Fingerprint details?

- Use cryptographic hash (computationally infeasible to find 2 distinct inputs w/ same hash)
- low prob. of collisions
- Sha-1, 20 byte hash value

Application responsibilities

- Track mapping between files/dirs/objs and hash values
- Brief but non-trivial (sfsrc)
- Merkle trees

Verti Implementation



- index maps fingerprints \rightarrow location in RAM
- hash buckets \rightarrow Random accesses, poor performance

add index \$ to avoid disk for index
add block \$ to skip all indexing

Performance?

Table 1

— random reads

poor read performance when uncoached
writes not so bad (seq. log) ^{finds} _{dups ok}

⇒ Room for improvement