

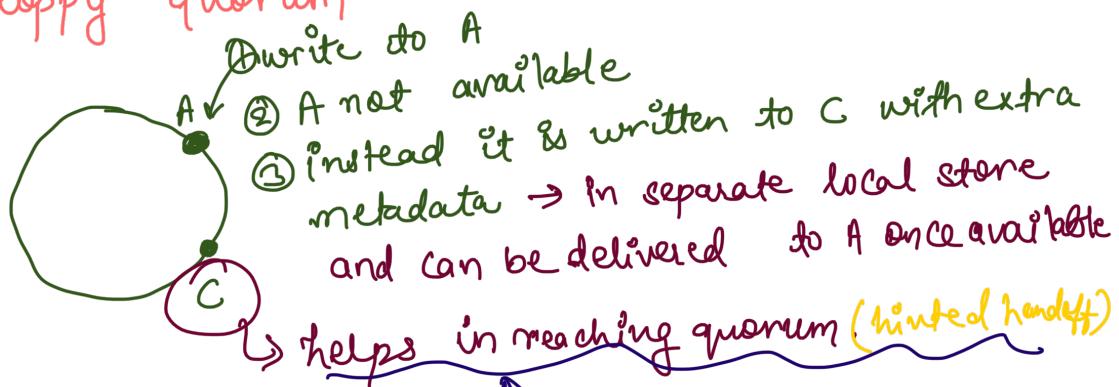
Dynamo Architecture Analysis

Part III - System Architecture 2

① Failure Handling - Hinted Handoff

what if ddb cluster is not able to reach quorum?
should write be rejected? \Rightarrow Durability $\uparrow\uparrow\uparrow$

ddb uses sloppy quorum



↳ DDB cluster nodes are spread across data centers to provide high availability
works well for transient issues

what if Permanent Failures?

ddb implements anti-entropy (replica synchronization) for keeping replicas synchronized.

Detection of replica inconsistency

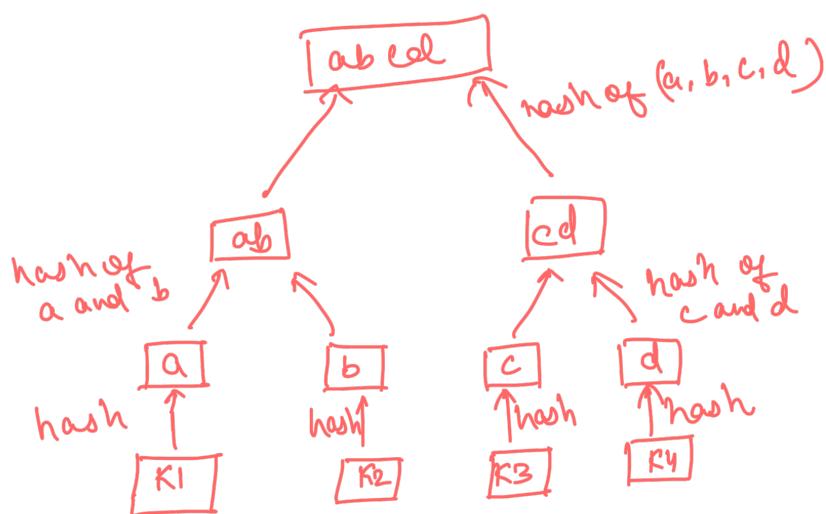
\hookrightarrow Merkle Trees \rightarrow helps in faster identification

Problem - identify inconsistency

- ① Check contents of replica → can take too much time
- ⓐ Traverse both replica contents
- ⓑ Create hash of replica content and compare hash
 ↳ If hash don't match, you've to sync entire replica.
 Merkle Tree helps in syncing only unsynchronized section.

How? Each key maps to leaf node

Assume DDB table has 4 keys



key changes on nodes will change → node added / removed

↳ Tree to be re-calculated.

We'll discuss this in DDB partitioning schemes.

② Membership

ⓐ Ring Membership

→ node is not available don't mean server is down.
 ↳ It may come back up.

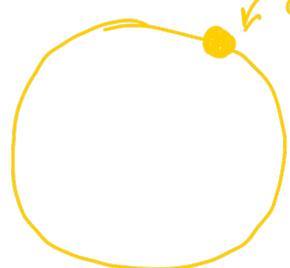
↳ partition rebalancing in ring should only be done once we're absolutely sure.

Assume Two Replicas R1 & R2
 ⓐ If root hash of R1 == R2
 ↳ all in sync

ⓑ If #a == false
 Traverse the tree to find out keys that are out of sync.

→ handled by administrator.

→ membership changes are propagated via gossip protocol to all nodes in ring.



add new node

→ @ choose

virtual nodes in
Hash ring

Set of tokens and
map nodes to token sets.

Only contain local node & token set
initially

this helps a node
to forward request to
specific node.

② Other mappings reconciled over
time via gossip protocol.

③ Existing nodes which were maintaining
the keys will transfer to new node.

could result in logically partitioned ring.

→ Add Node A
→ Add Node B } initially they'll be unaware
of each other

Some nodes are chosen to act as seeds.

↳ via configuration Service/static config

→ Known to all nodes
→ can be discovered via
external mechanism

→ All nodes reconcile membership with seed → avoid logical partitioning

Hope you enjoyed the content

will discuss 'Architectural Decisions and Learnings' in next video

Happy Learning 😊

YouTube - MsDeep Singh