Chapter 5 : Replication(Leaderless replication)

Leaderless Replication

- Client sends a write request to one node, database copy that write to other replicas
- Leader determines how the writes needs to be processes in what order and follower apply the writes in the same order.
- Allowing any replica to take writes from clients.
- In leaderless implementation, the client directly sends writes to several replicas or a coordinator node might send writes to replicas. But unlike leader, coordinator does not enforce a particular ordering of write.

What happens when a node is down?

- Single-leader or multi leader – failover
- Leaderless – no failover
- Read requests are sent to many nodes in parallel.
- So version numbers determine the latest value.

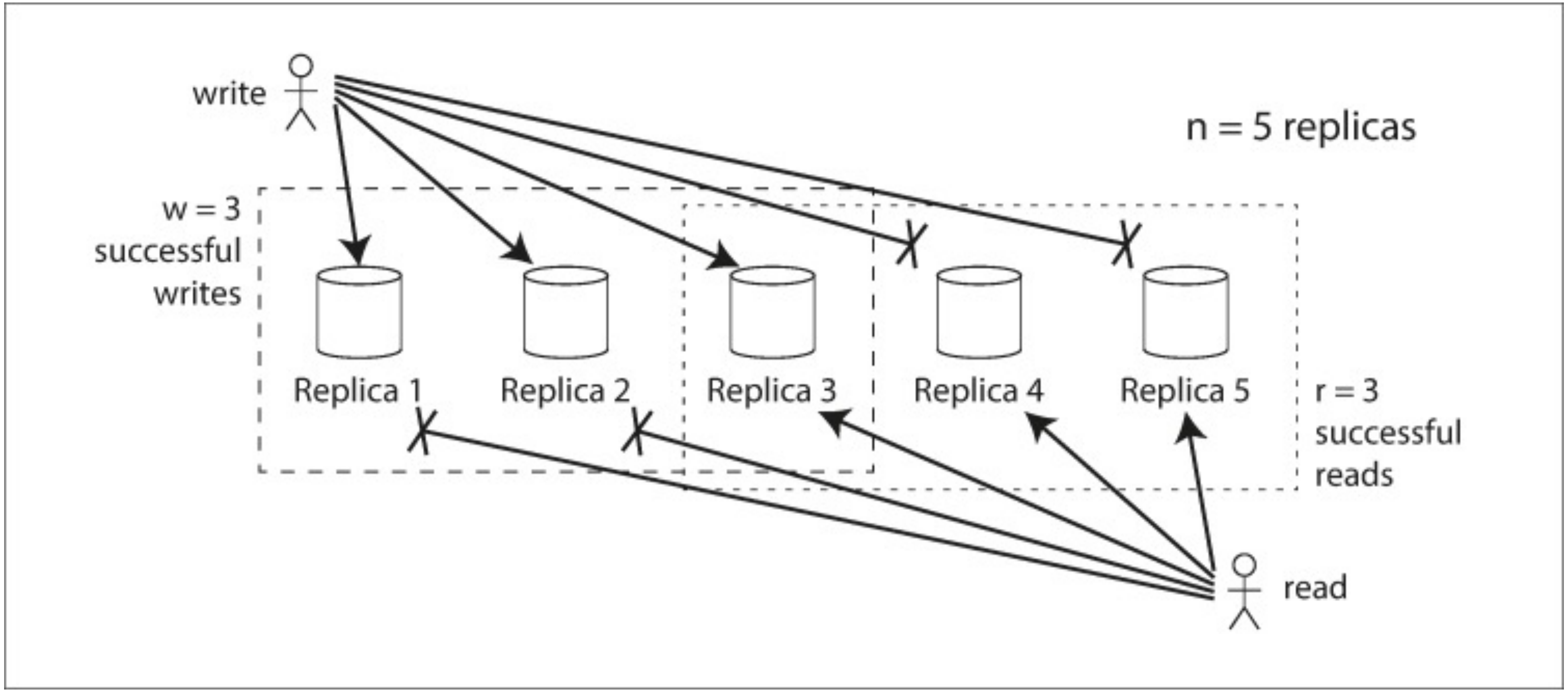
How does an unavailable node catch up?

Read repair

Anti-entropy process

Quorums for reading and writing

- N replicas
- Every write must be confirmed by w nodes to be successful
- Must query at least r nodes for each read.
- w + r > n
- Example: n = 3, w = 2, n = 2
- Minimum number of votes required for the read and write to be valid
- Commonly, n = odd number(typically 3 or 5) and w= r=(n+1)/2
- Few writes and many reads , w = n, r = 1
- Quorum condition w + r > n allows:
- If w < n, we can still process writes if a node is unavailable.
- If r < n, we can still process reads if a node is unavailable.
- Ex, n=3, w=2, r=2 we can tolerate 1 unavailable node
- Ex n=5,w=3,r=3, we can tolerate 2 unavailable node


write
n = 5 replicas
w = 3
successful
writes
Replica 1
Replica 2
Replica 3
Replica 4
Replica 5
r = 3
successful
reads
read

Limitations of Quorum Consistency

- Often r and w is chosen to be more than n/2
- What matters is – Set of nodes used by write and read operations overlap in at least one node.
- We can also set w and r to smaller numbers such that w + r <=n
- More likely to read stale values
- Lower latency and higher availability
- Limitations:
  - If two writes occur concurrently
  - If write happens concurrently with a read
  - If write succeeded on some replicas but failed on other, overall less than w, and not rolled back.
  - If a node having new value fails

Sloppy Quorums and Hinted Handoff

- Is it better to return errors to all requests for which we cannot reach a quorum of w or n nodes?
- Or should we accept writes and write them to some nodes that are reachable but are not among the n nodes on which the value usually lives? – Sloppy quorum – writes and reads still require w and r nodes to be successful but those may not be the n “home” nodes.
- Once network interruption is fixed, any writes that one node temporarily accepted on behalf of another node are sent to appropriate home nodes – hinted handoff
- Sloppy quorums – assurance of durability

Summary

- ✓ Leaderless Replication?
- ✓ What happens when a node is down?
- ✓ Read repair and Anti-entropy process
- ✓ Quorums for reading and writing
- ✓ Limitations of Quorum consistency
- ✓ Sloppy Quorums

Thank You!