CLOUD COMPUTING

Multi Leader Replication

Dr. Prafullata Kiran Auradkar

Department of Computer Science and Engineering

Acknowledgements:

Significant information in the slide deck presented through the Unit 3 of the course have been created by **Dr. H.L. Phalachandra** and would like to acknowledge and thank him for the same. There have been some information which I might have leveraged from the content of **Dr. K.V. Subramaniam's** lecture contents too. I may have supplemented the same with contents from books and other sources from Internet and would like to sincerely thank, acknowledge and reiterate that the credit/rights for the same remain with the original authors/publishers only. These are intended for classroom presentation only.

Multi-Leader Replication

Recap: Replication *is a means keeping a copy of the same data on multiple machines that are connected via a network. We discussed that there were three popular algorithms for replicating changes between nodes: single-leader, multi-leader, and leaderless replication.*

- *We discussed* ***Leader Based Replication*** *earlier.*
- Leader-based replication has a single bottleneck in the leader
- All writes must go through it. If there is a network interruption between the user and the leader, then no writes are allowed.
- An alternate approach to consider is , what if we have more than one leader through whom you can do the writes.

Multi-Leader Replication

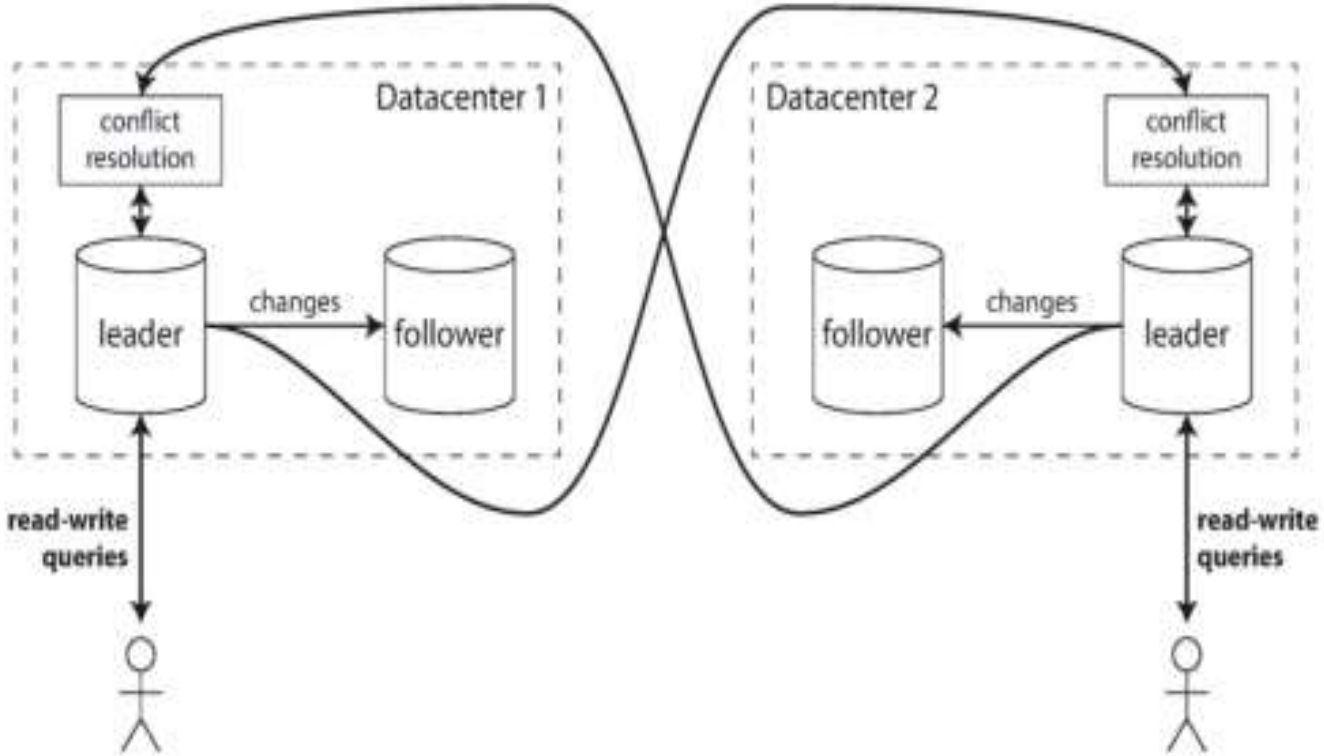
So with this natural extension of the leader-based replication model of allowing more than one node to accept writes will lead to

- Each node that processes a write must forward that data change to all the other nodes
- This approach is also known as master/master or active/active replication
- Each leader simultaneously acts as a follower to the other leaders.

This would have different Use cases like

- **Multi-Datacenter operation** can have a leader in each datacenter. Each datacenter has a regular leader– follower replication and between datacenters, each datacenter's leader replicates its changes to the leaders in other datacenters.
- **Clients with offline operation**

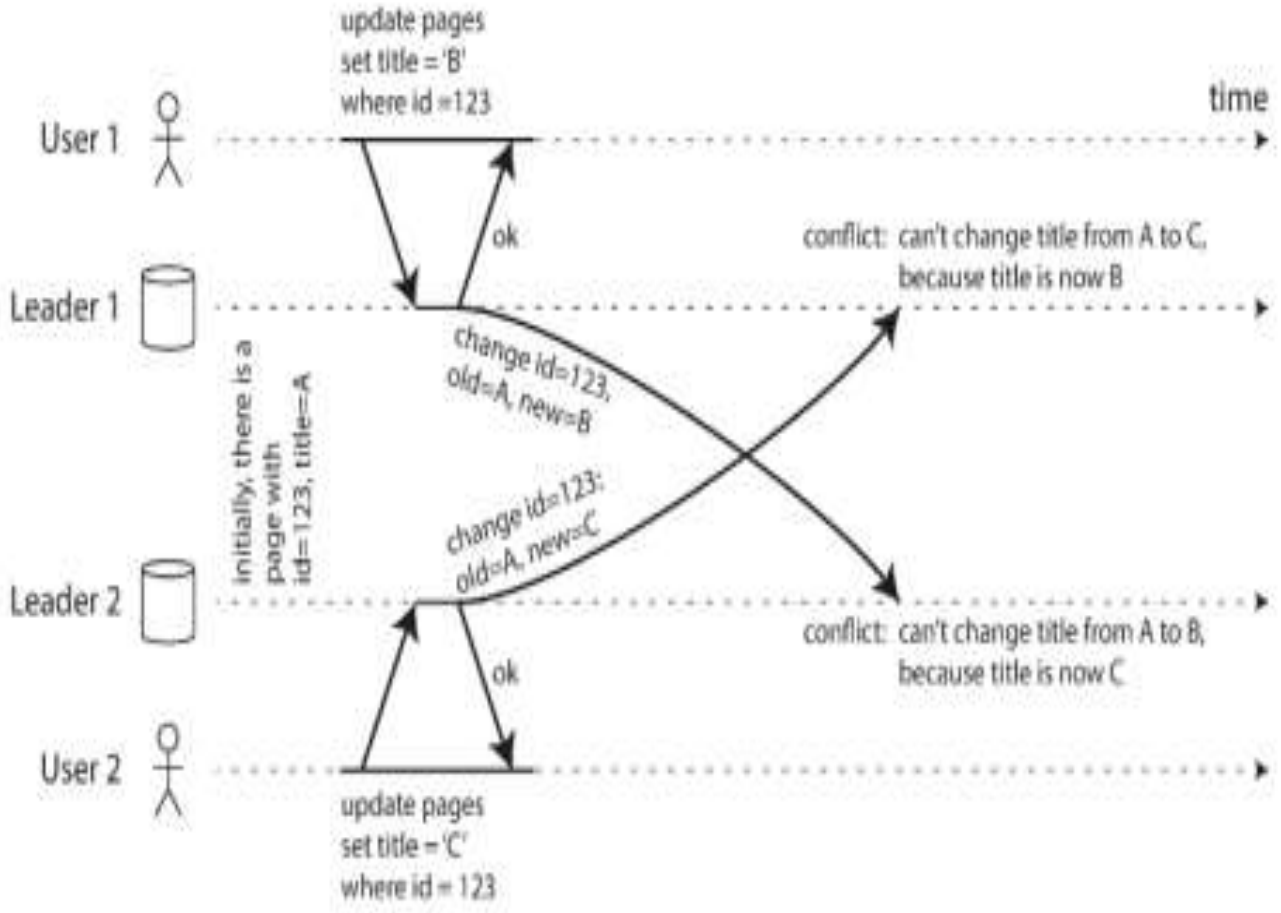
  Every device has a local database that acts as a leader
- **Collaborative editing**



Multi-Leader Replication Challenges

A problem with multi-leader replication is that it can lead to write conflicts

- **Conflict avoidance** - ensure that all writes for a particular record go through the same leader
- **Converging toward a consistent state** - Give each write a unique ID (e.g., a timestamp, a long random number, a UUID, or a hash of the key and value), pick the write with the highest ID as the winner and throw away the other writes.
- **Custom conflict resolution logic -** Write conflict resolution logic in application code. That code may be executed on write or on read



Single-Leader vs. Multi-Leader Replication

Performance

- In a single-leader configuration, every write must go over the internet to the datacenter with the leader. This can add significant latency to writes and might contravene the purpose of having multiple datacenters in the first place.
- In a multi-leader configuration, every write can be processed in the local datacenter and is replicated asynchronously to the other datacenters. Thus, the inter-datacenter network delay is hidden from users, which means the perceived performance may be better.

Tolerance of datacenter outages

- In a single-leader configuration, if the datacenter with the leader fails, failover can promote a follower in another datacenter to be leader.
- In a multi-leader configuration, each datacenter can continue operating independently of the others, and replication catches up when the failed datacenter comes back online.

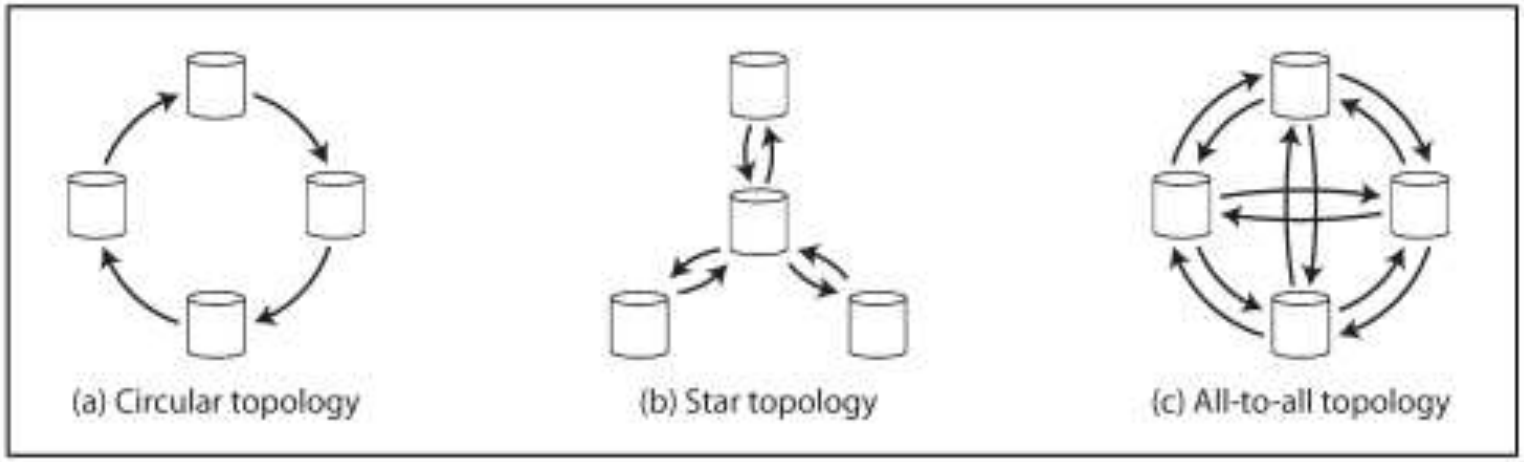
Single-Leader vs. Multi-Leader Replication

Tolerance of network problems

- Traffic between datacenters usually goes over the public internet, which may be less reliable than the local network within a datacenter.
- A single-leader configuration is very sensitive to problems in this inter-datacenter link, because writes are made synchronously over this link.
- A multi-leader configuration with asynchronous replication can usually tolerate network problems better: a temporary network interruption does not prevent writes being processed.

Multi-Leader Replication Topologies

- A replication topology describes the communication paths along which writes are propagated from one node to another.



- **Circular topology** - each node receives writes from one node and forwards those writes (plus any writes of its own) to one other node
- **Star topology** - one designated root node forwards writes to all of the other nodes.
- **All-to-all topology** - allows messages to travel along different paths, avoiding a single point of failure.

THANK YOU

Prafullata Kiran Auradkar

Department of Computer Science and Engineering

prafullatak@pes.edu