Replication

Systems Analysis & Design

Learning Objectives

By the end of this session, you will have acquired the following information:

- Replication Concept
- Synchronous vs Asynchronous Replication
- Failover
- Write-ahead Log (WAL)

Illustrative Example

- Alireza has recently joined a new social media platform.
- He is experiencing slow loading times when attempting to read posts from his peers on his timeline.

Illustrative Example

- Alireza utilizes cloud storage for his files.
- A recent incident with the cloud provider resulted in the removal of the files.
- Unfortunately, there appear to be no backups available.

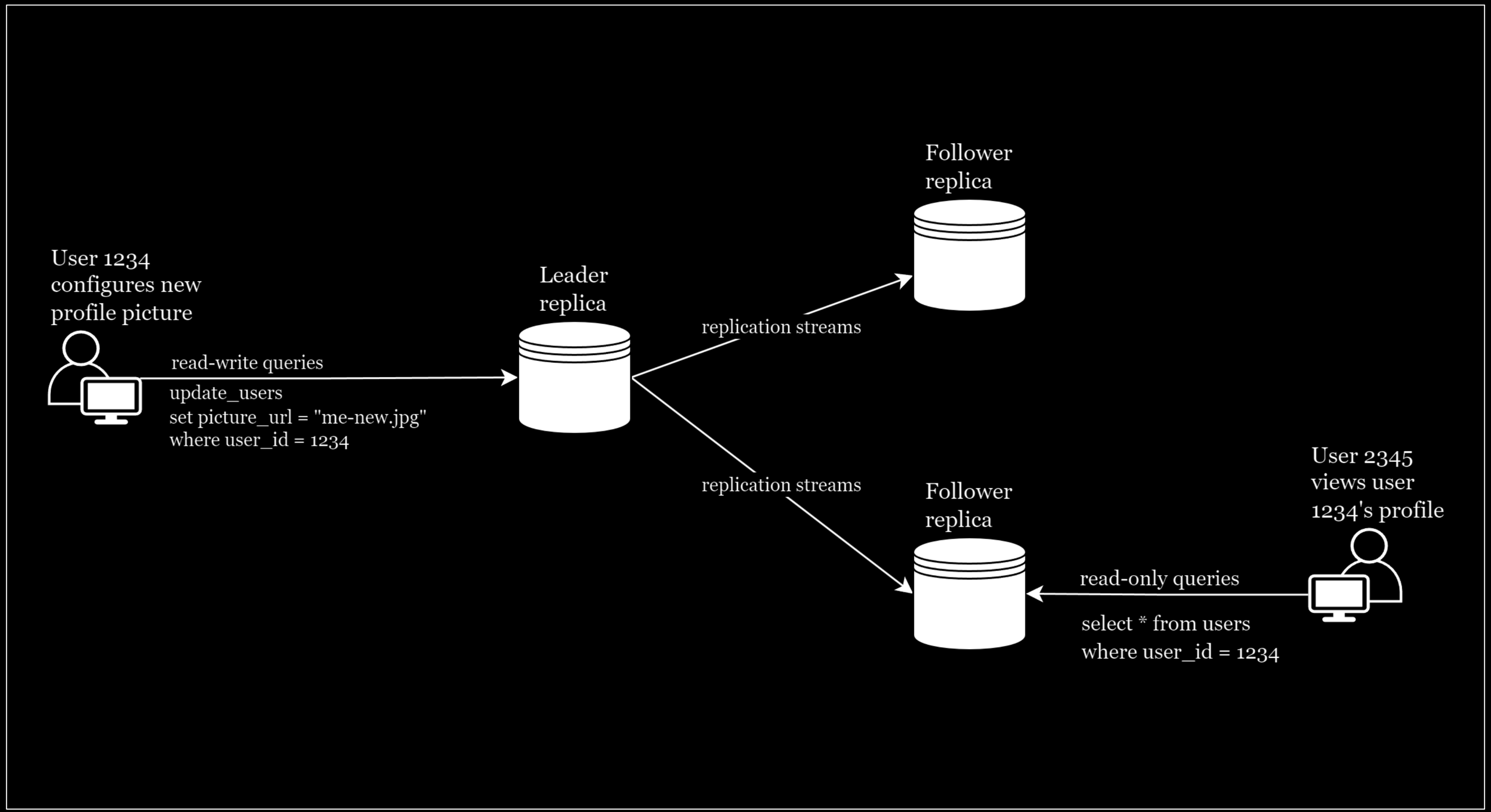
Replication

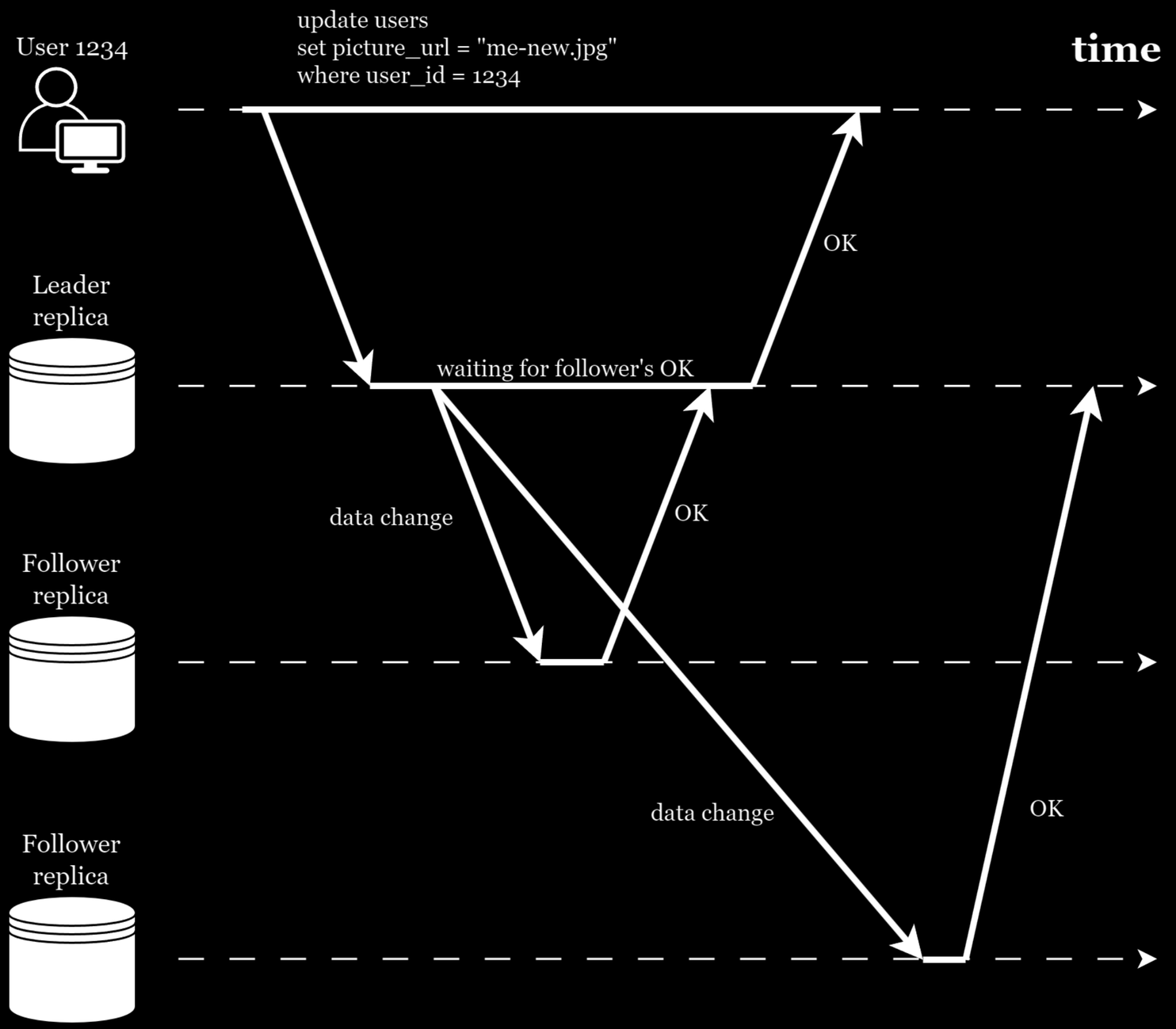
- A copy of the same data is kept on multiple machines that are connected via a network.
- Replication offers several benefits:
  - It keeps data geographically close to users, reducing latency.
  - It allows the system to continue functioning even if some parts fail, increasing availability.
  - It scales out the number of machines that can serve read queries, increasing read throughput.
- In this lecture, we'll assume that the dataset is small enough for each machine to hold a complete copy.
- A node that stores a copy of the database is referred to as a replica.

Replication Algorithms

- Single Leader
- Multi Leader
- Leaderless

Single Leader


Follower
replica
User 1234
configures new
profile picture
Leader
replica
replication streams
read-write queries
update_users
set picture_url = "me-new.jpg"
where user_id = 1234
replication streams
Follower
replica
User 2345
views user
1234's profile
read-only queries
select * from users
where user_id = 1234


User 1234
update users
set picture_url = "me-new.jpg"
where user_id = 1234
time
Leader replica
OK
waiting for follower's OK
data change
OK
Follower replica
data change
OK
Follower replica

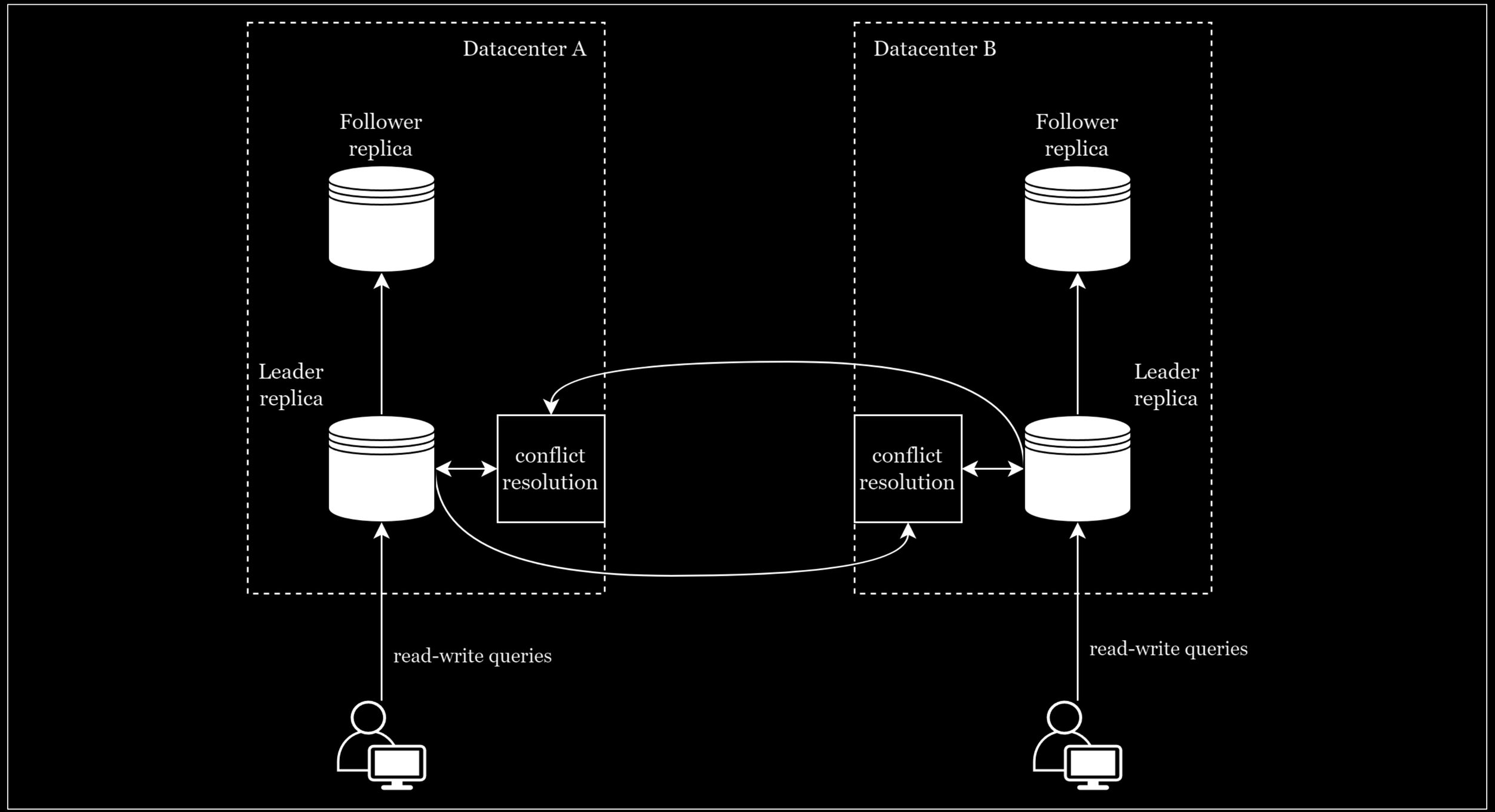
Leader Failure: Failover

- One of the followers needs to be promoted to become the new leader.
- Users need to be reconfigured to send their writes to this new leader.
- The other followers need to start consuming data changes from the new leader.

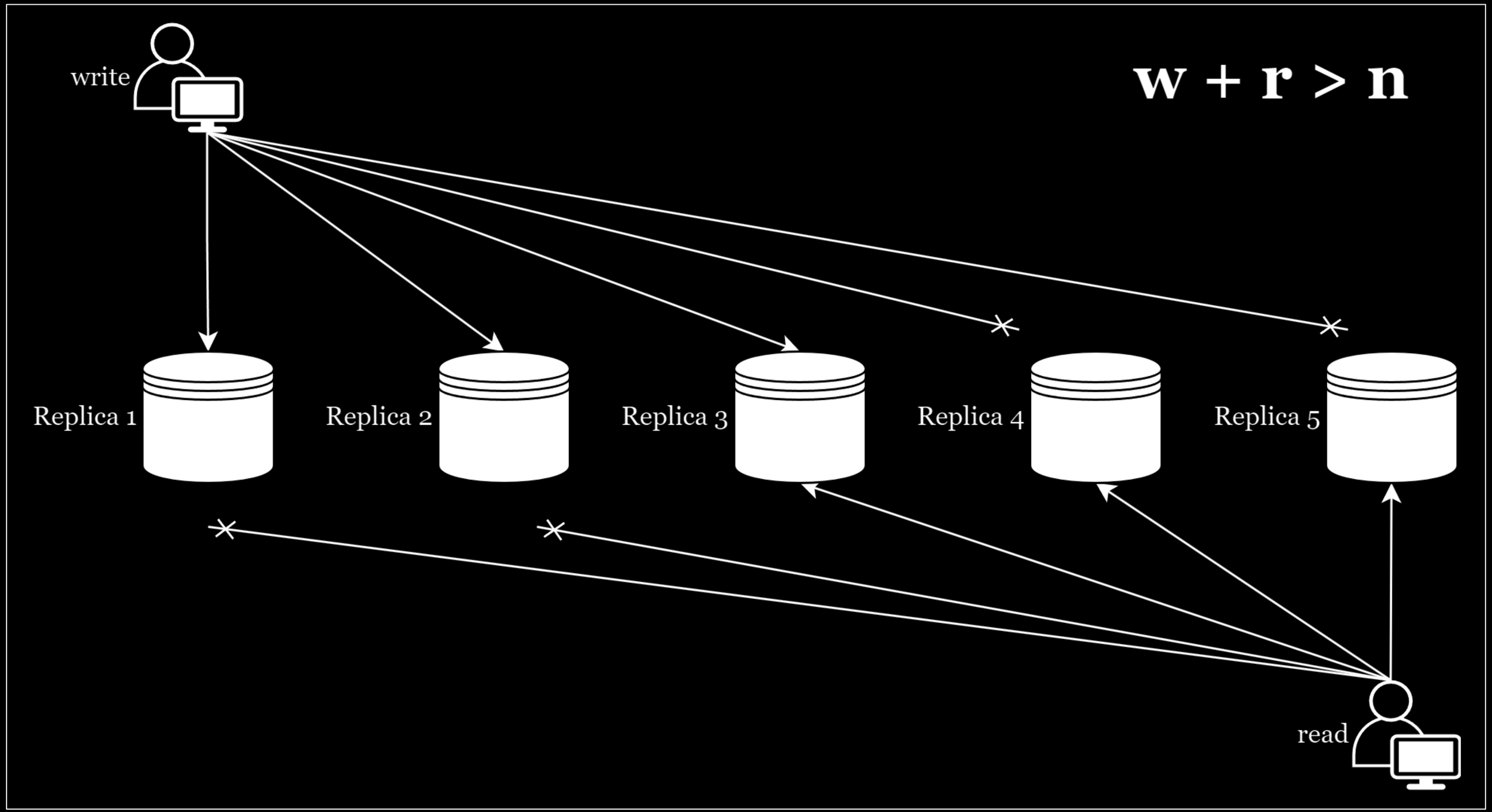
Write-ahead Log (WAL) Shipping

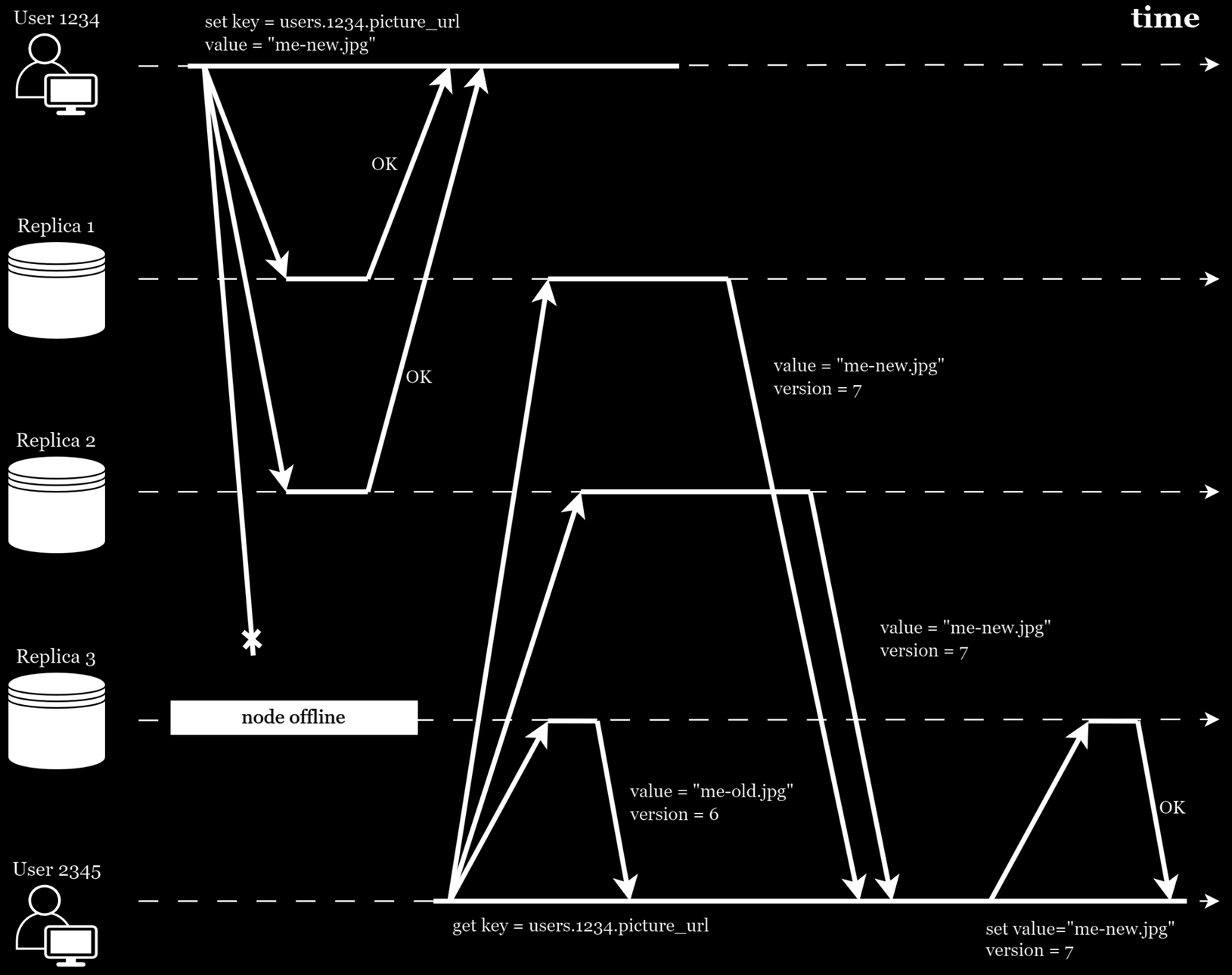
- The log, an append-only sequence of information, contains all changes to the database.
- Changes to the database, where tables and indexes reside, must be applied only after those changes have been logged.
- We can use the exact same log to build a replica on another node.

Multi Leader


Datacenter A
Datacenter B
Follower replica
Follower replica
Leader replica
Leader replica
conflict resolution
conflict resolution
read-write queries
read-write queries

Leaderless


write
w + r > n
Replica 1
Replica 2
Replica 3
Replica 4
Replica 5
read


User 1234
set key = users.1234.picture_url
value = "me-new.jpg"
time
OK
Replica 1
OK
value = "me-new.jpg"
version = 7
Replica 2
value = "me-new.jpg"
version = 7
Replica 3
node offline
value = "me-old.jpg"
version = 6
OK
User 2345
get key = users.1234.picture_url
set value="me-new.jpg"
version = 7

Further Resources

- Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems (pages 173-215)