

# Skybridge: Bounded Staleness for Distributed Caches

Robert Lyerly  
*Meta Platforms Inc.*

Scott Pruett

Kevin Doherty  
*Meta Platforms Inc.*

Greg Rogers  
*Meta Platforms Inc.*

Nathan Bronson  
*OpenAI*

John Hugg  
*Meta Platforms Inc.*



 **Meta**

# Introduction



Meta operates social media products used by billions of people around the world every day

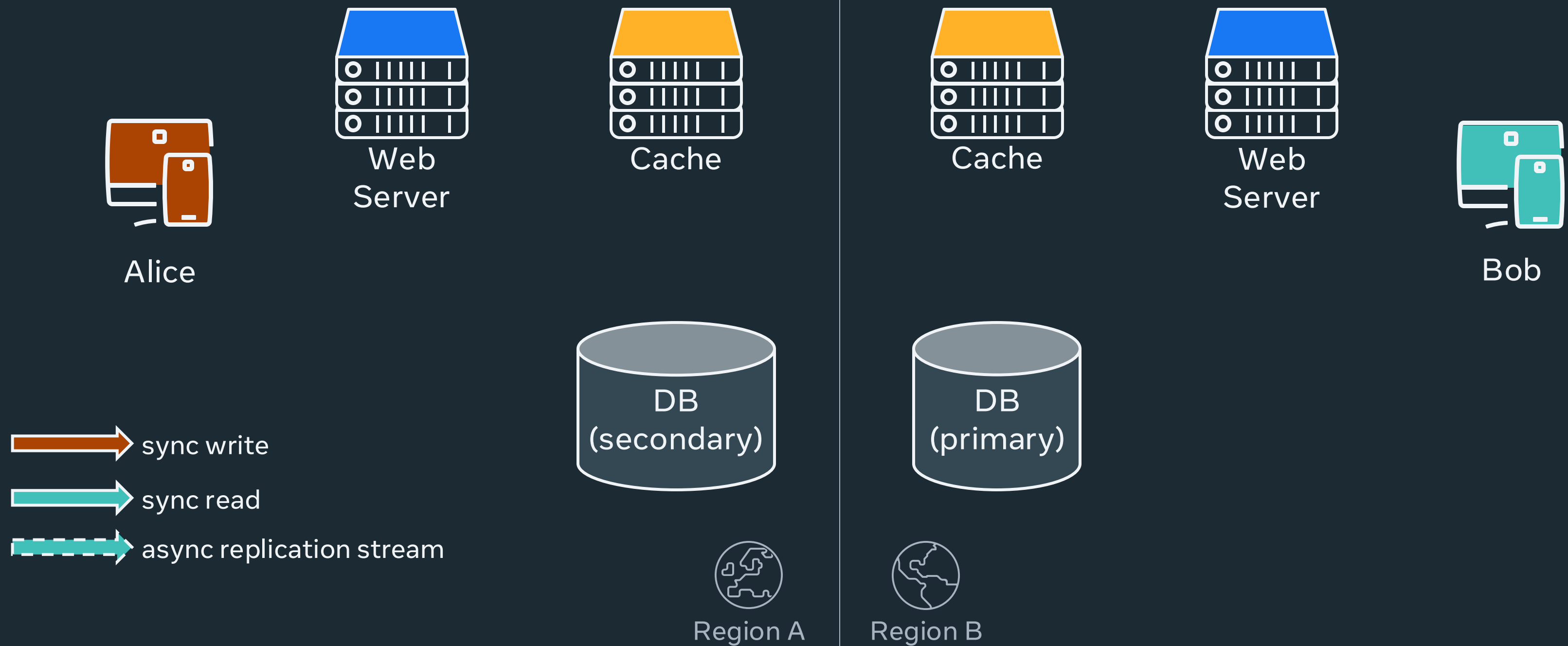


Meta's products have strict latency and high availability requirements

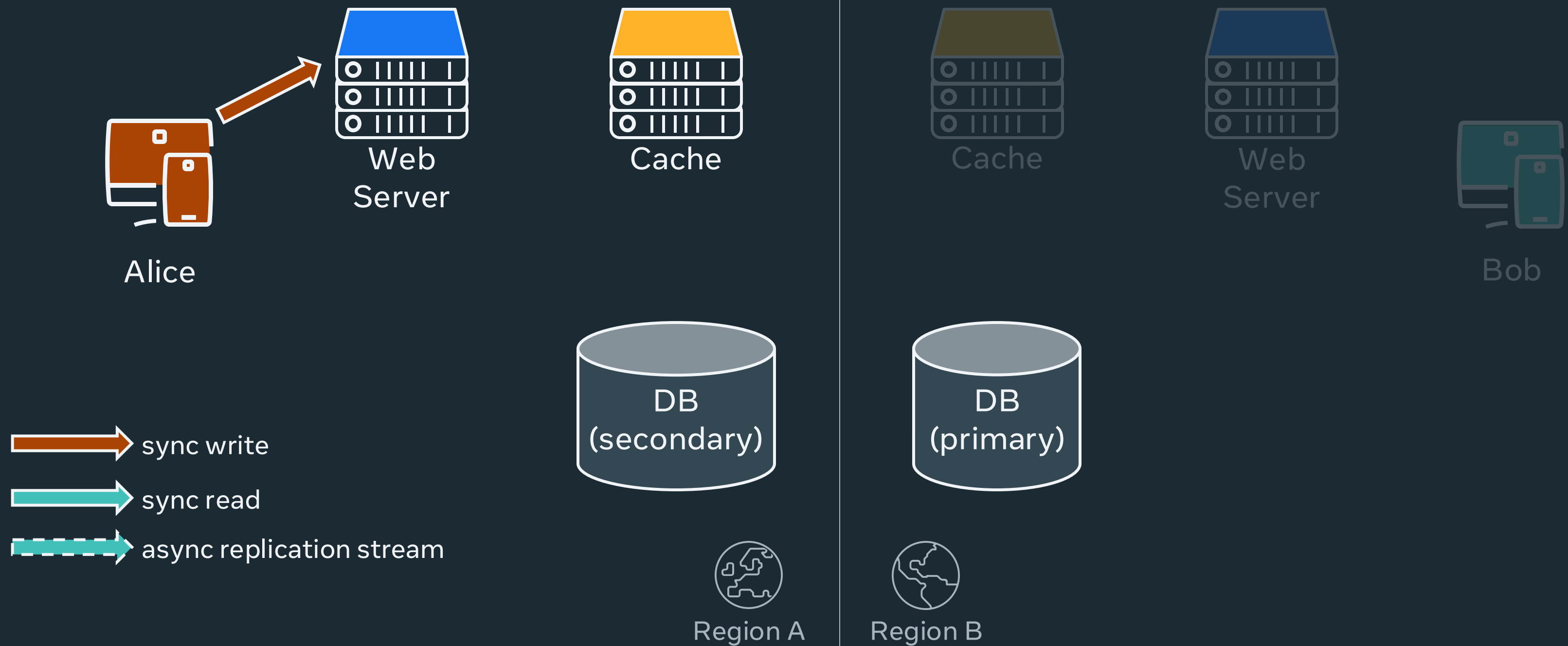


Meta operates a loosely-coupled serving stack in multiple geographic regions worldwide

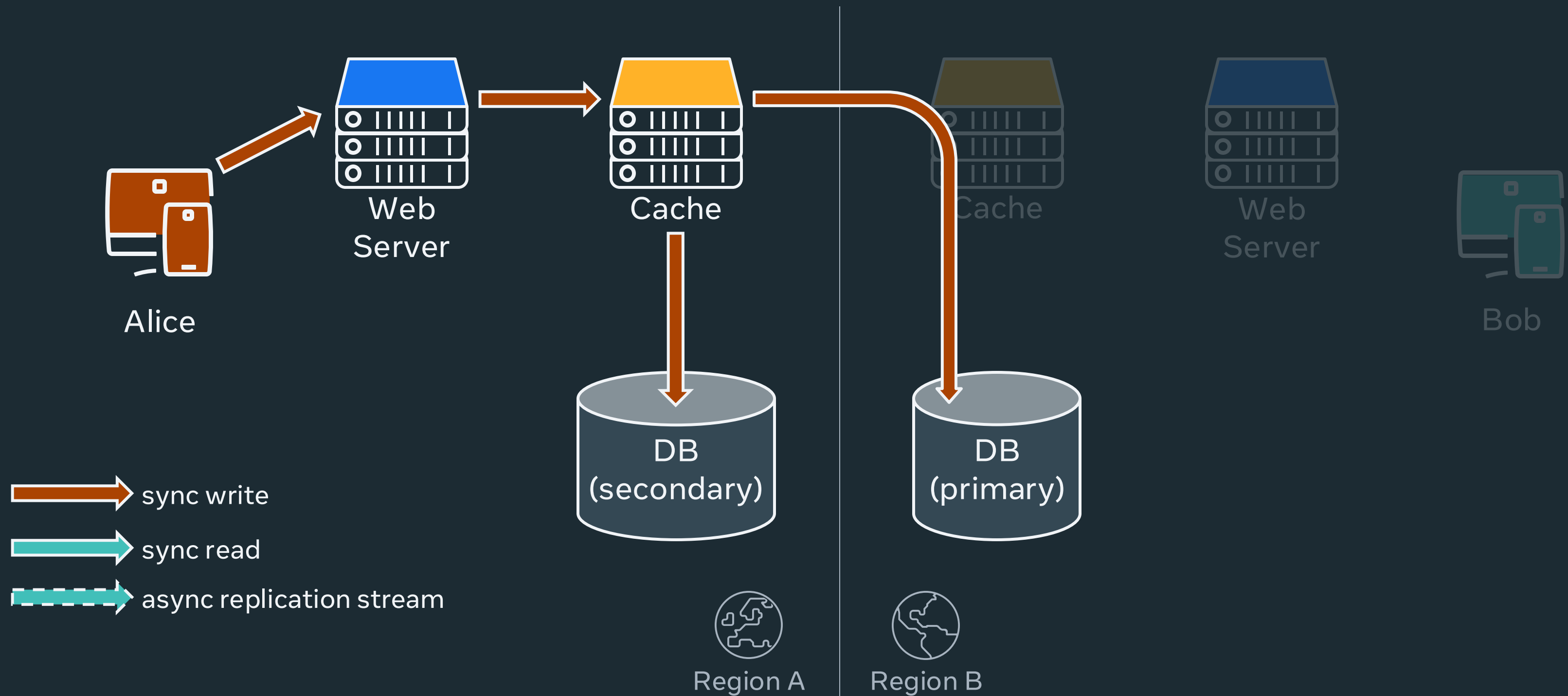
# System Architecture



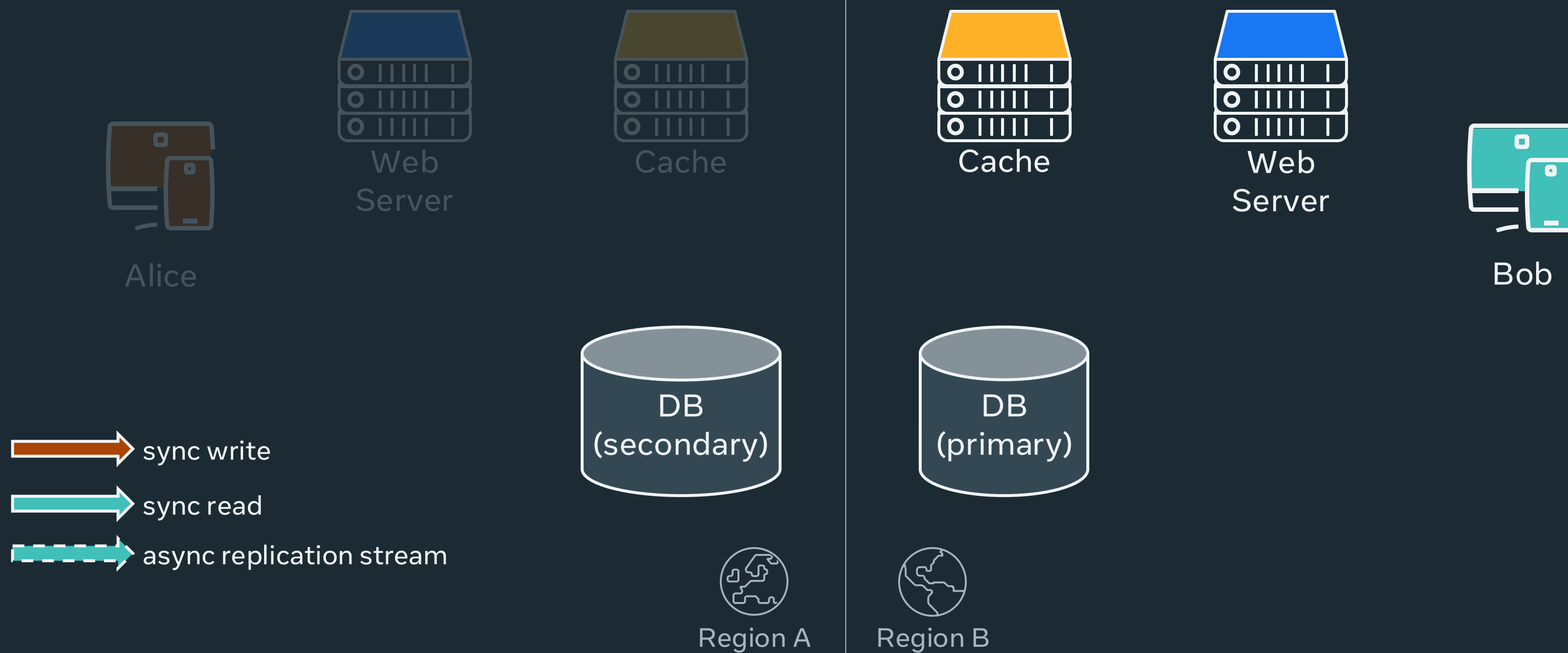
# System Architecture



# System Architecture

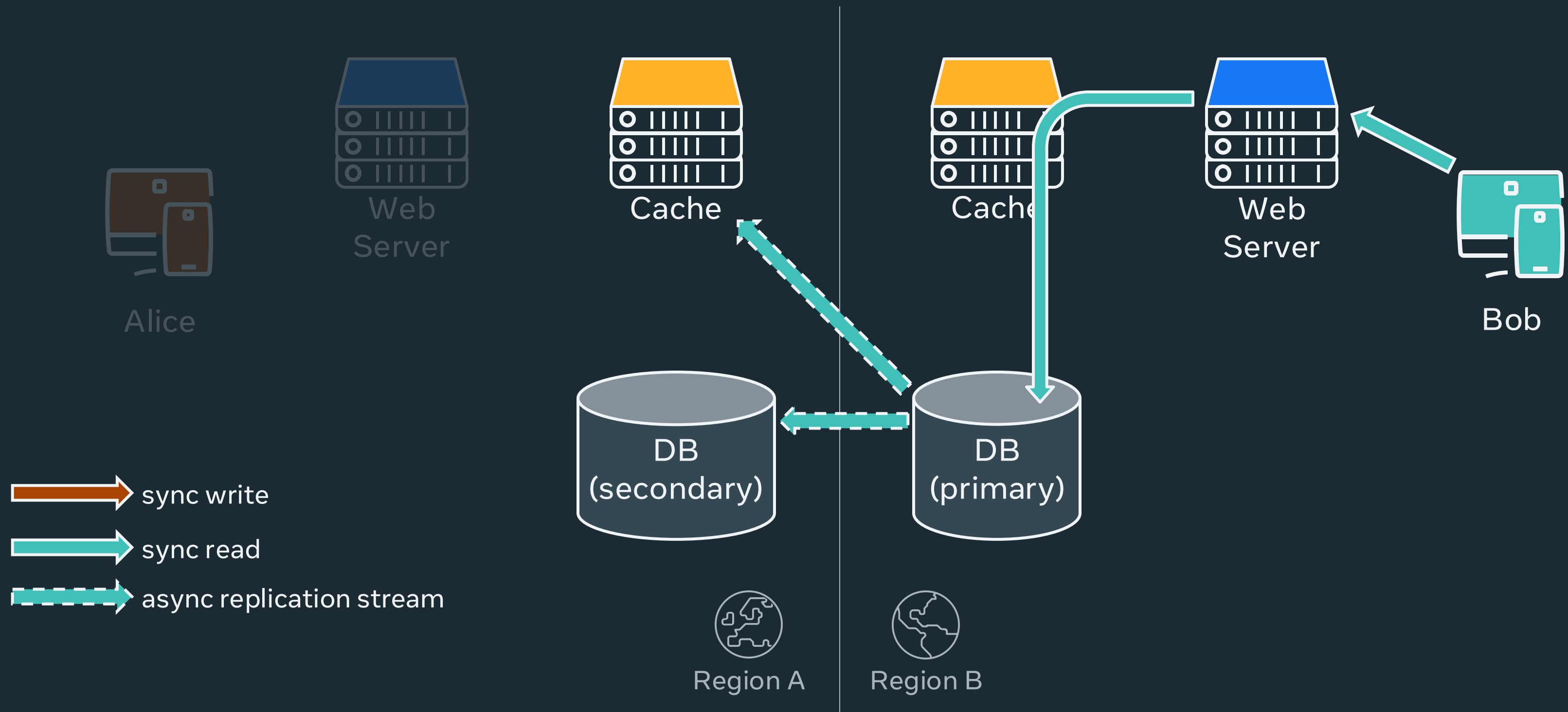


# System Architecture





# System Architecture

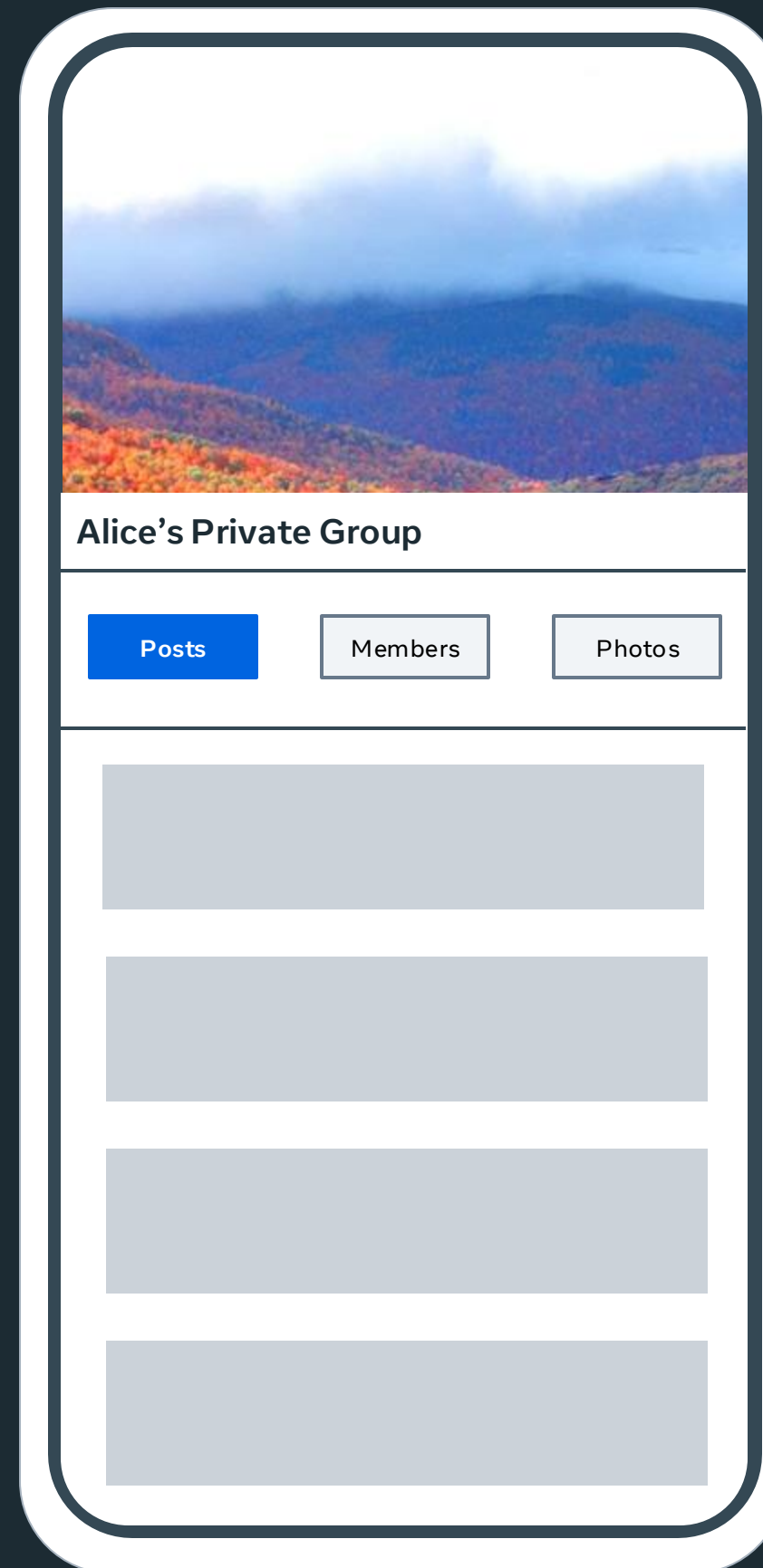




# Problem

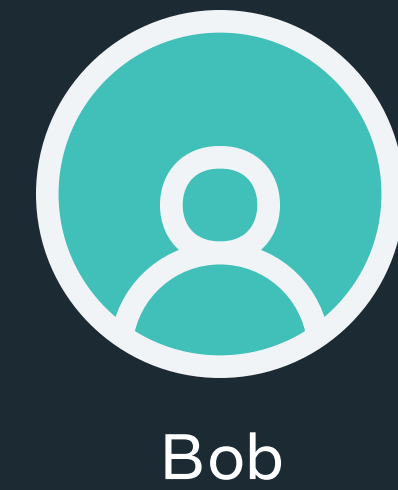
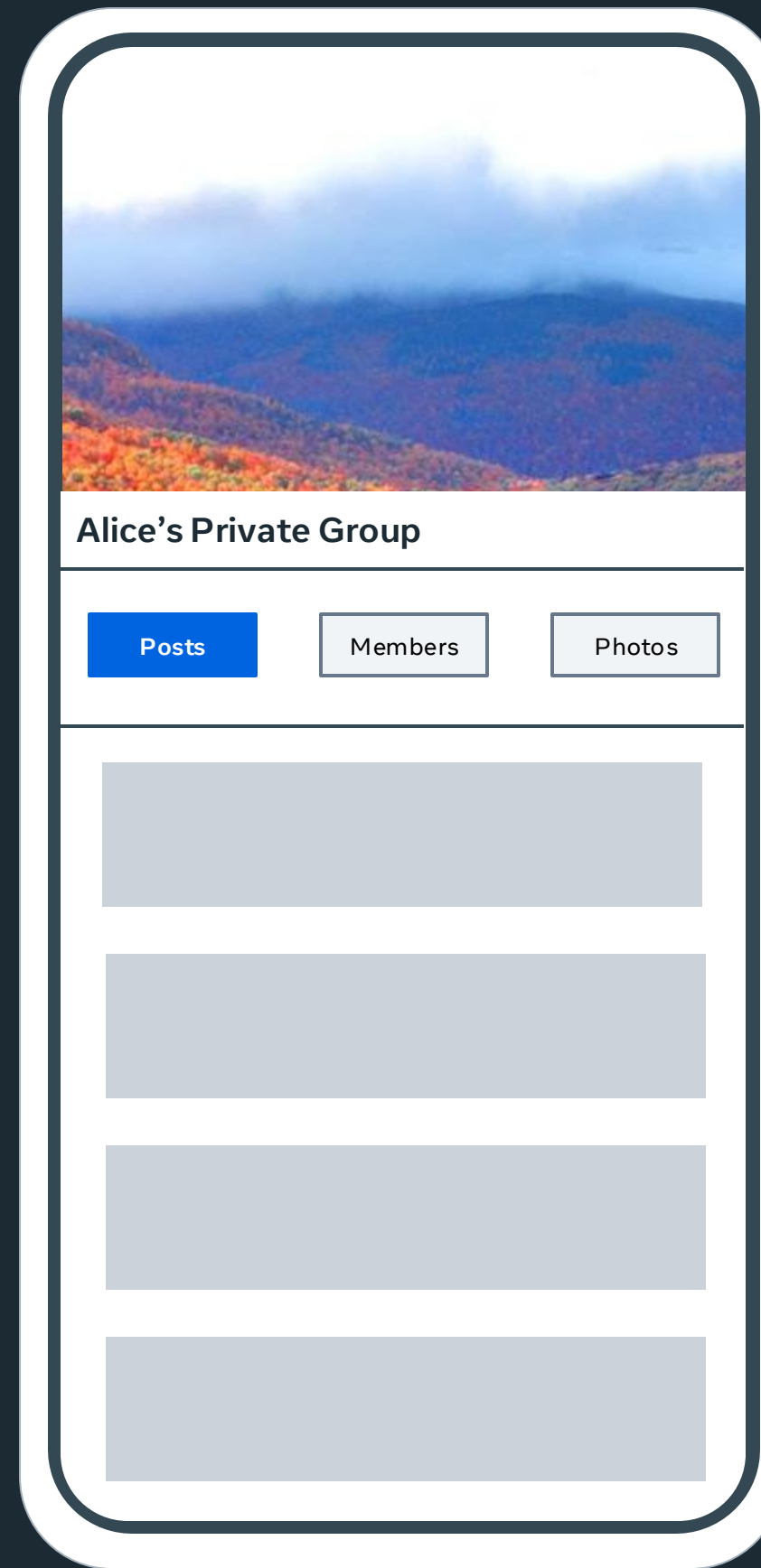
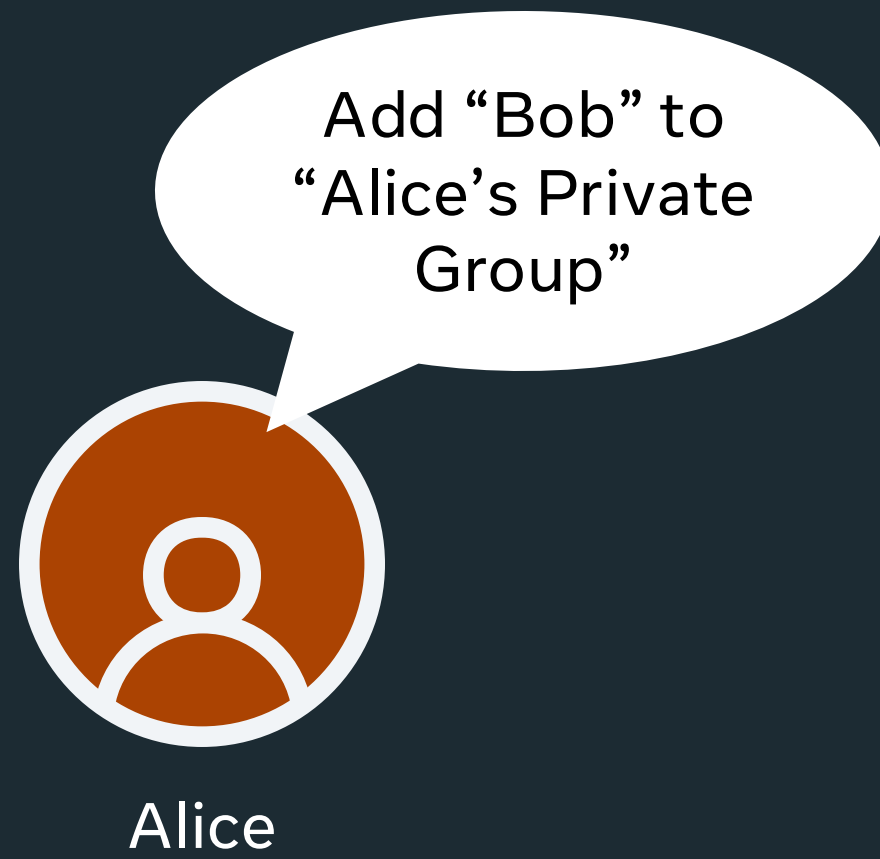


Alice



Bob

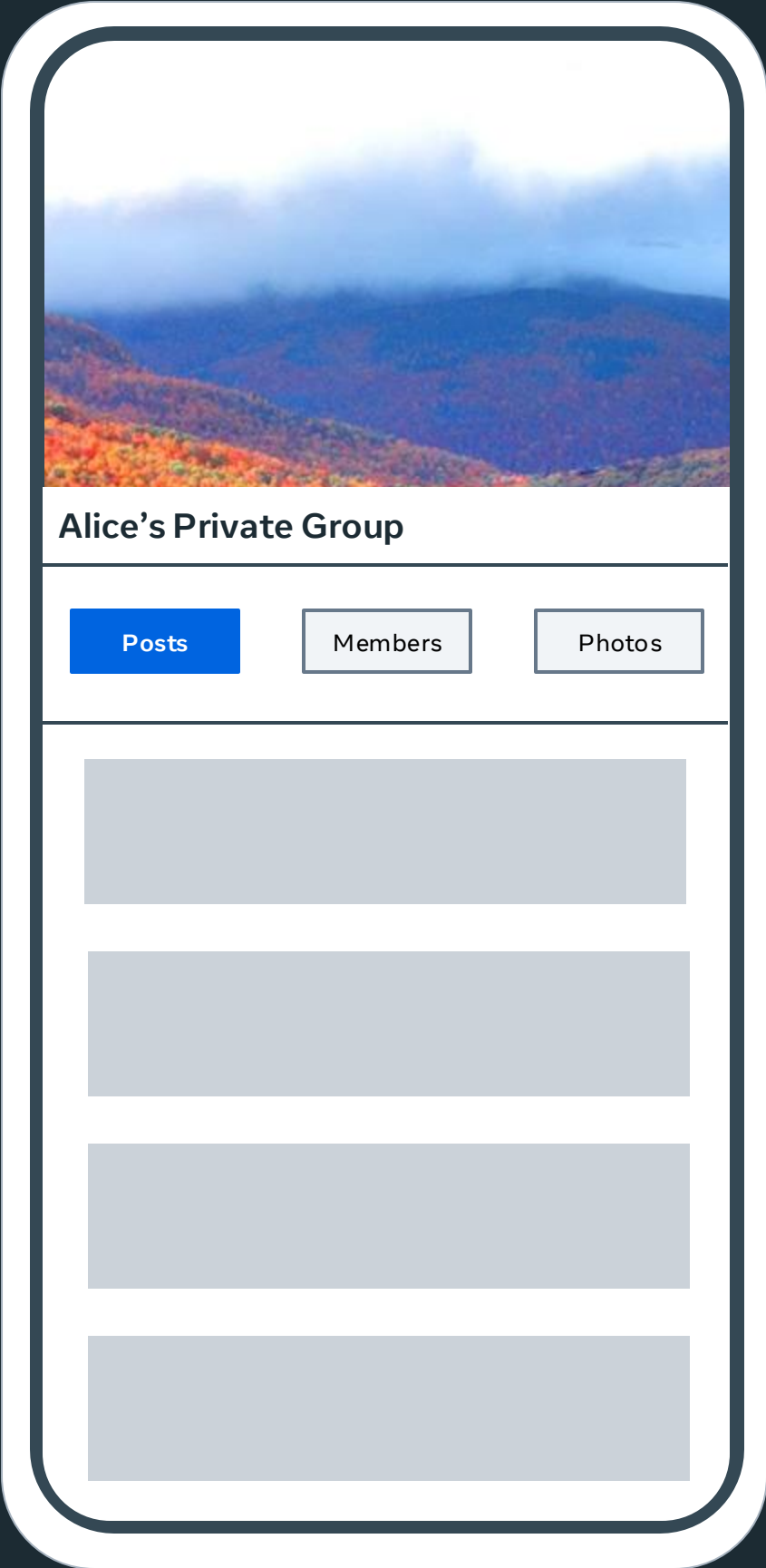
# Problem



# Problem



Alice



Load "Alice's Private Group"

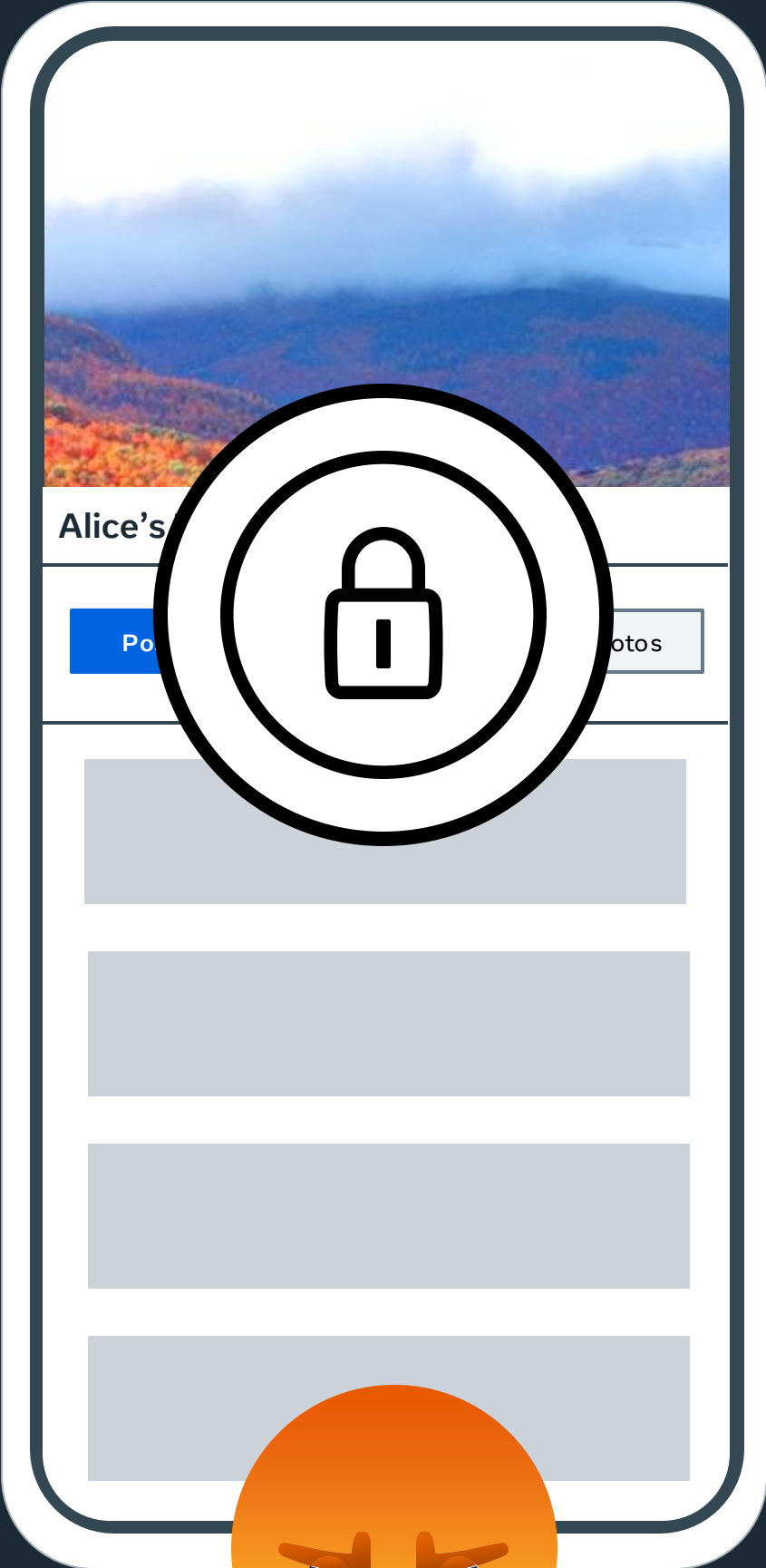


Bob

# Problem



Alice



Load "Alice's Private Group"

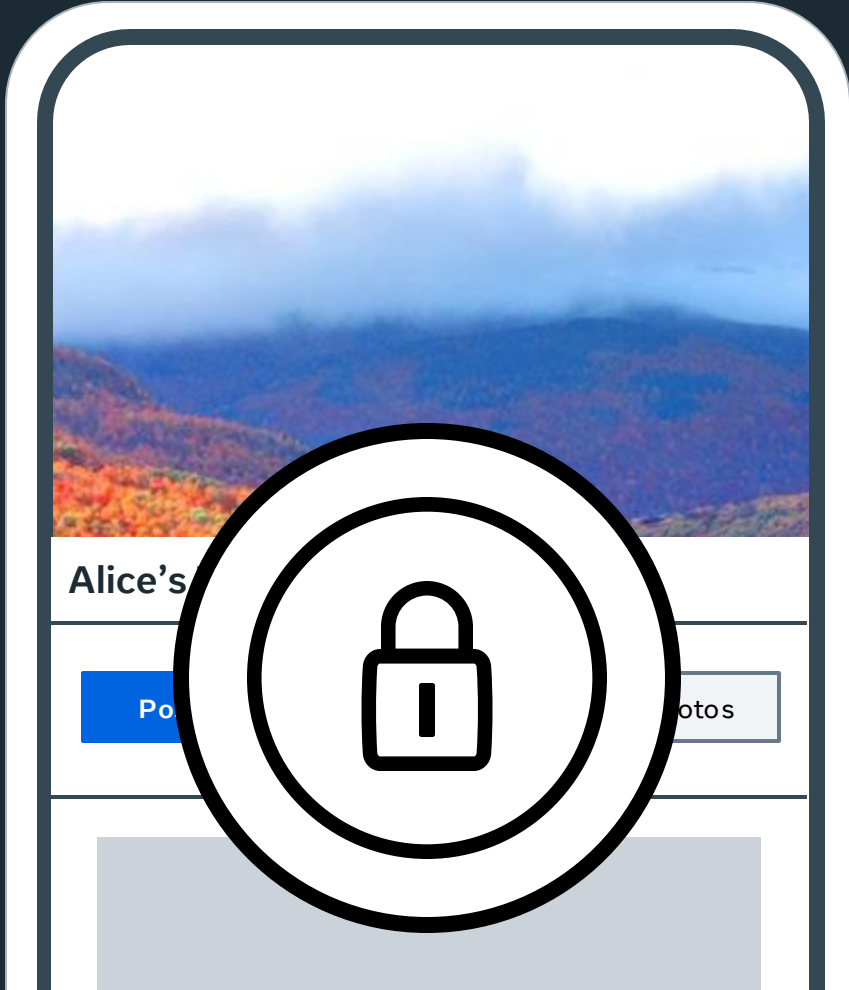


Bob

# Problem



Alice



Load "Alice's Private Group"

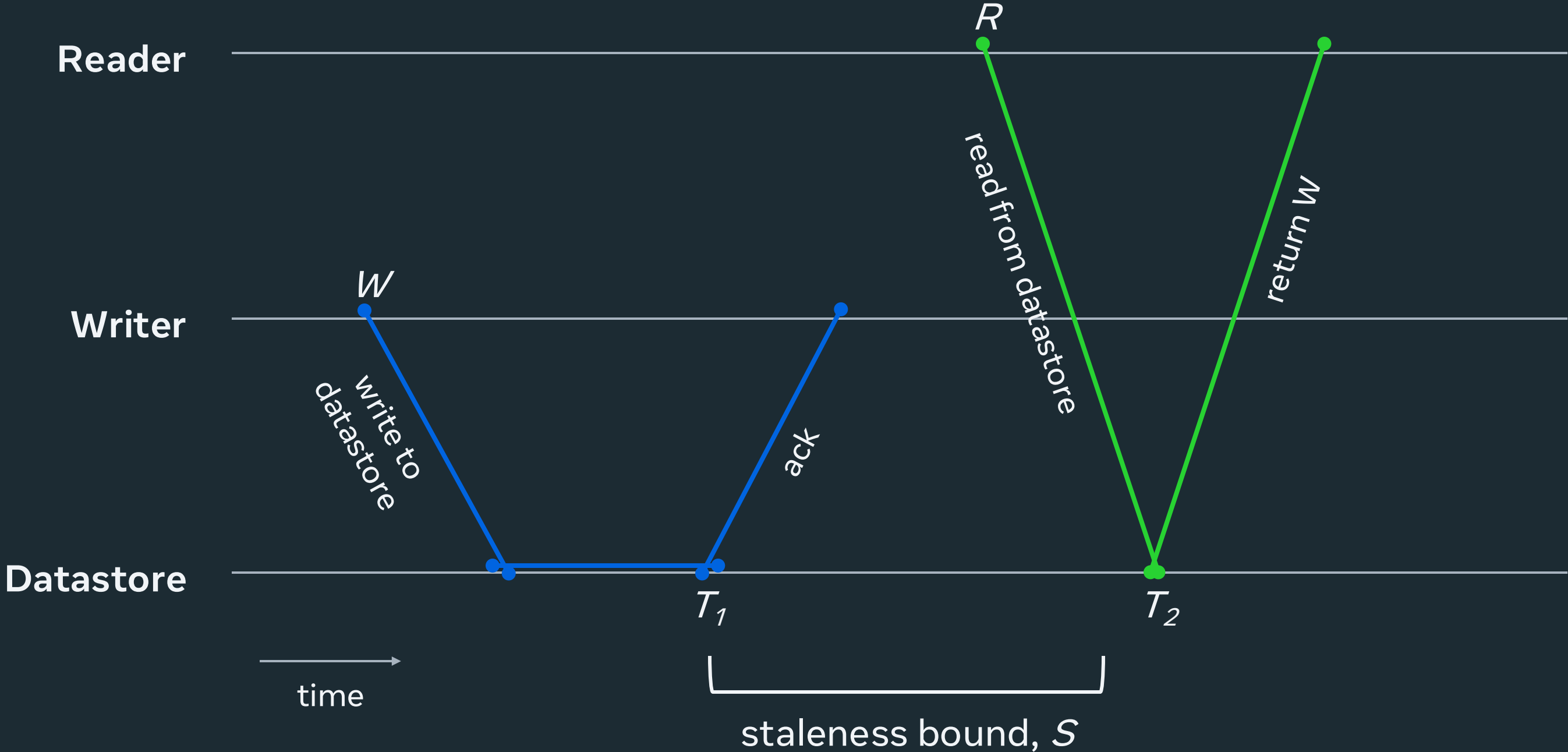


Bob

Can we bound how long it takes writes to be visible while still achieving the scalability of eventual consistency?



# Bounded Staleness



# Why Bounded Staleness?



Matches user expectations  
of the system

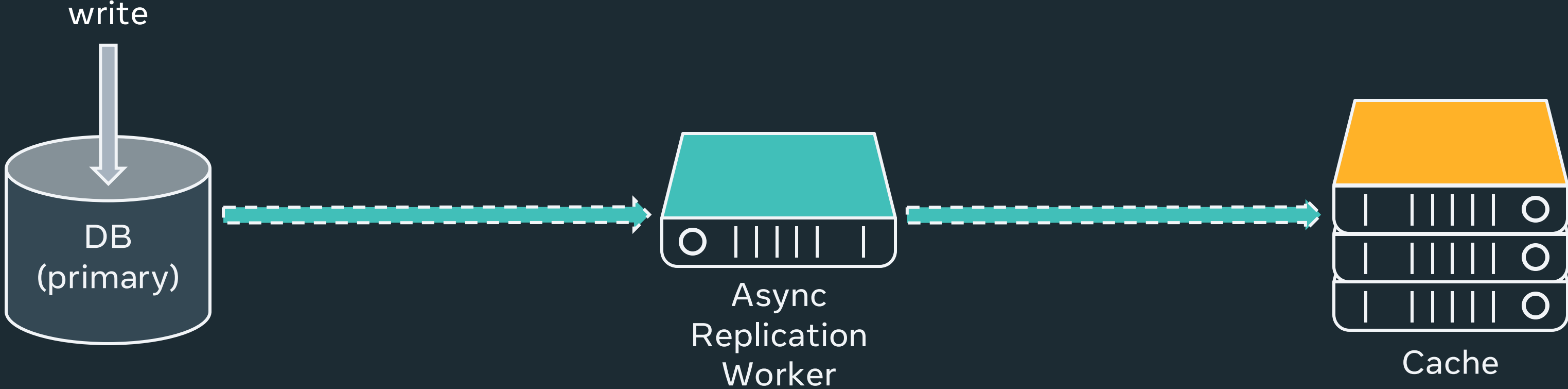


Lightweight enough to be  
applied to all cache reads

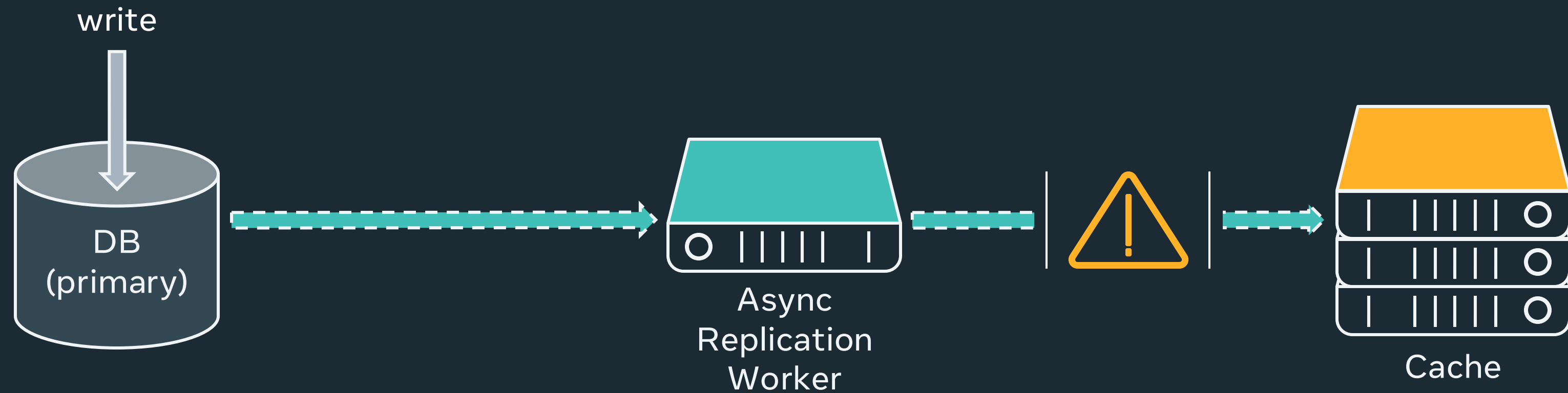


Applies to all writes, regardless  
of where they occur or who is  
doing them

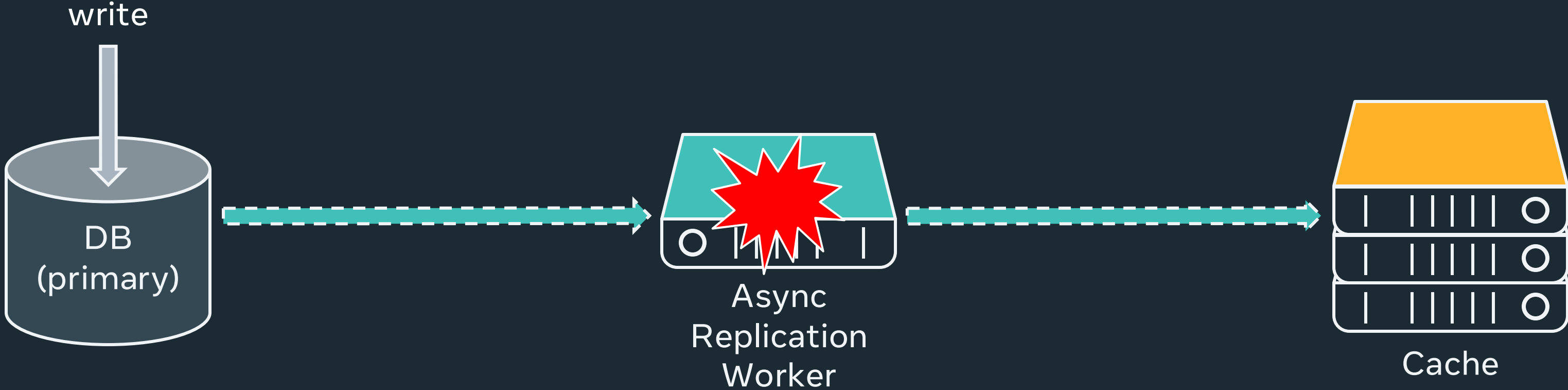
# Challenges



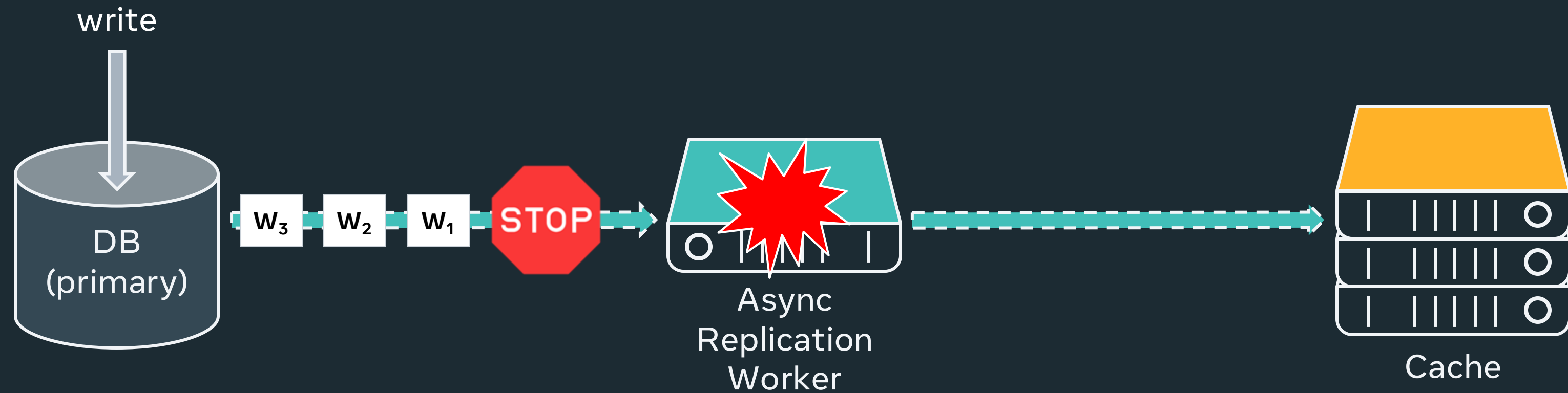
# Challenges



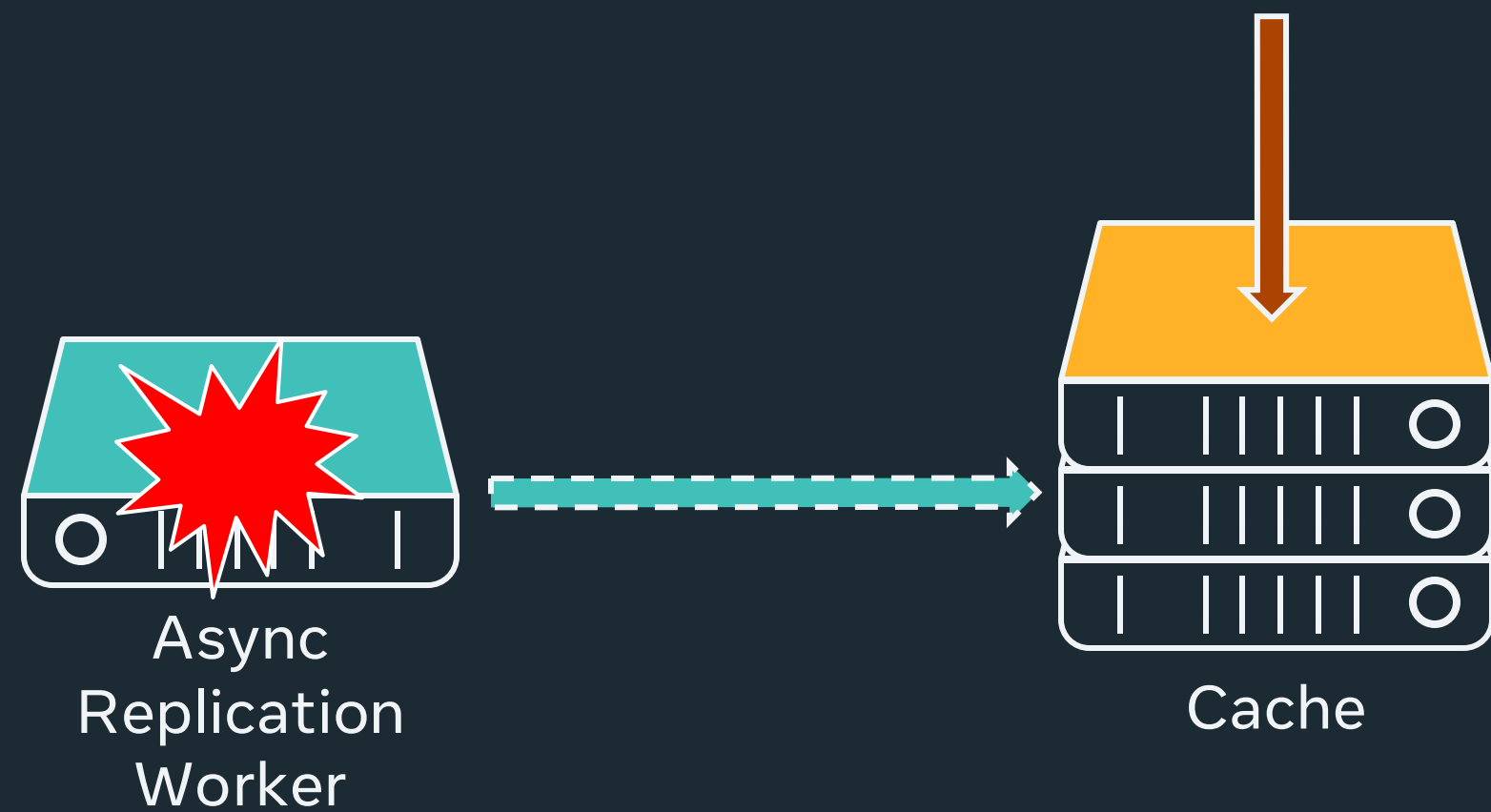
# Challenges



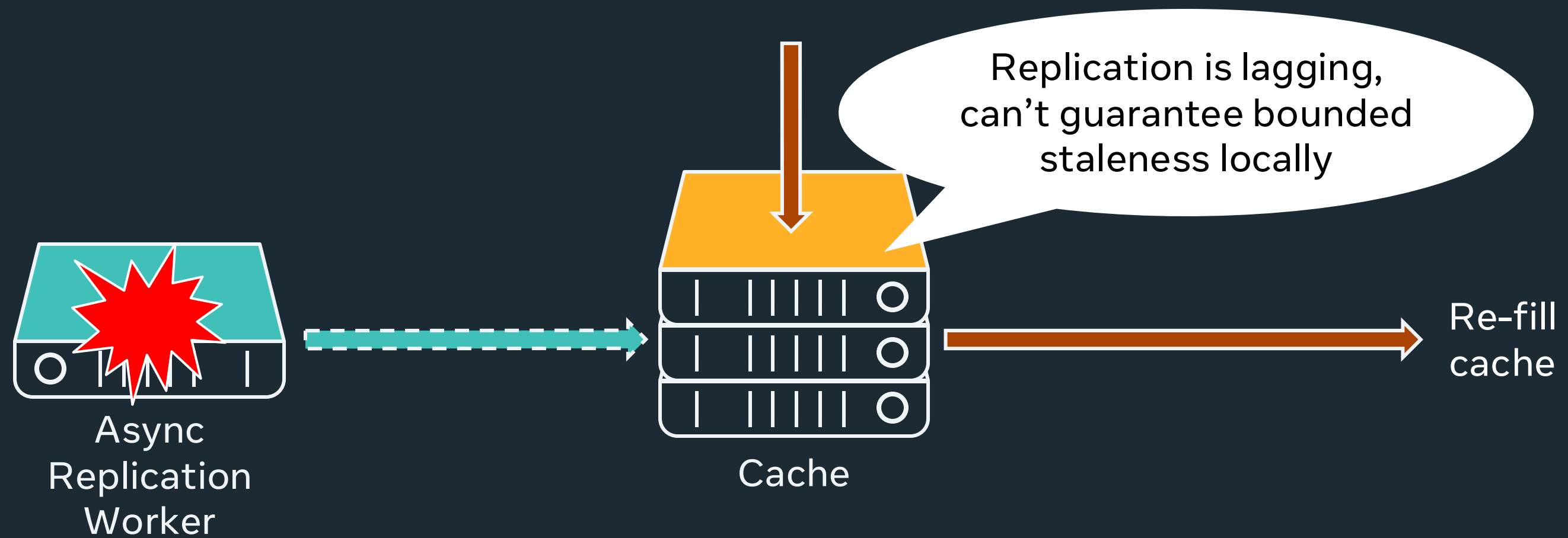
# Challenges



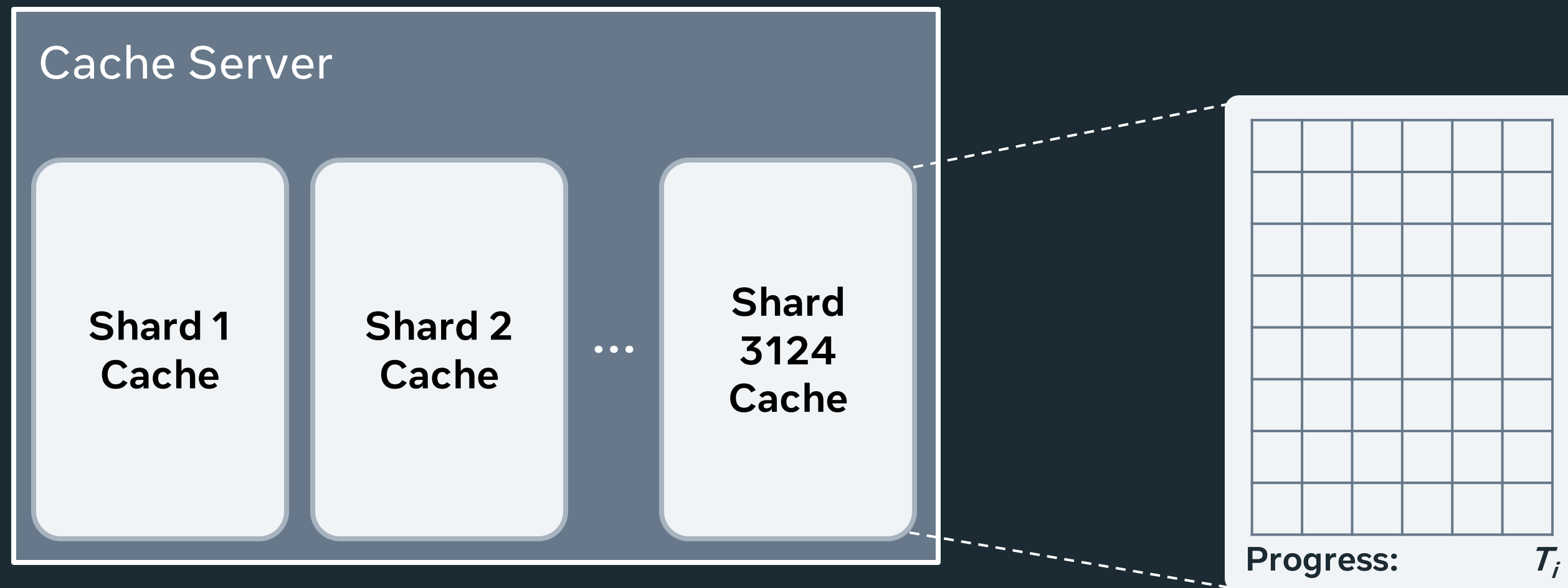
# Challenges



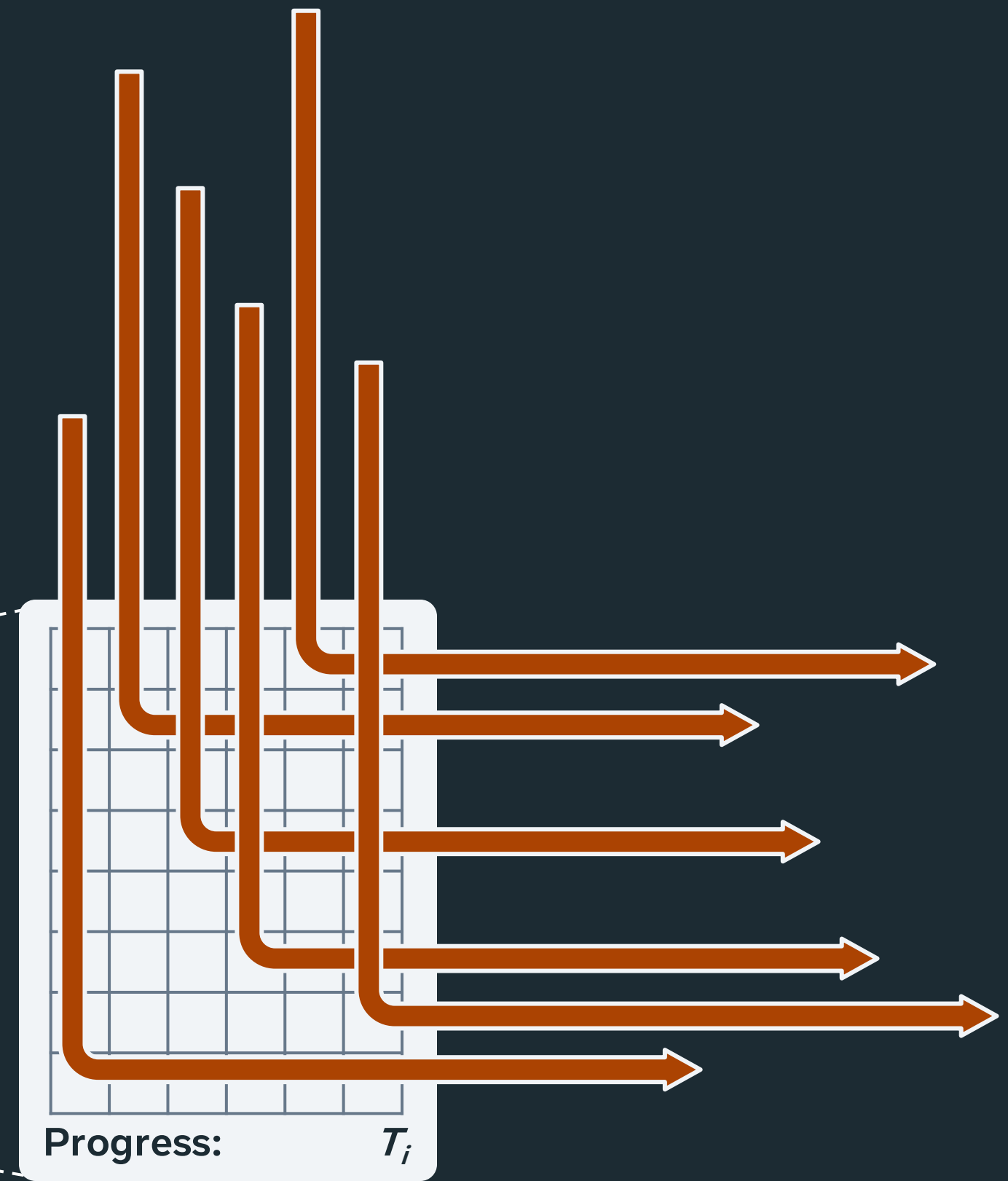
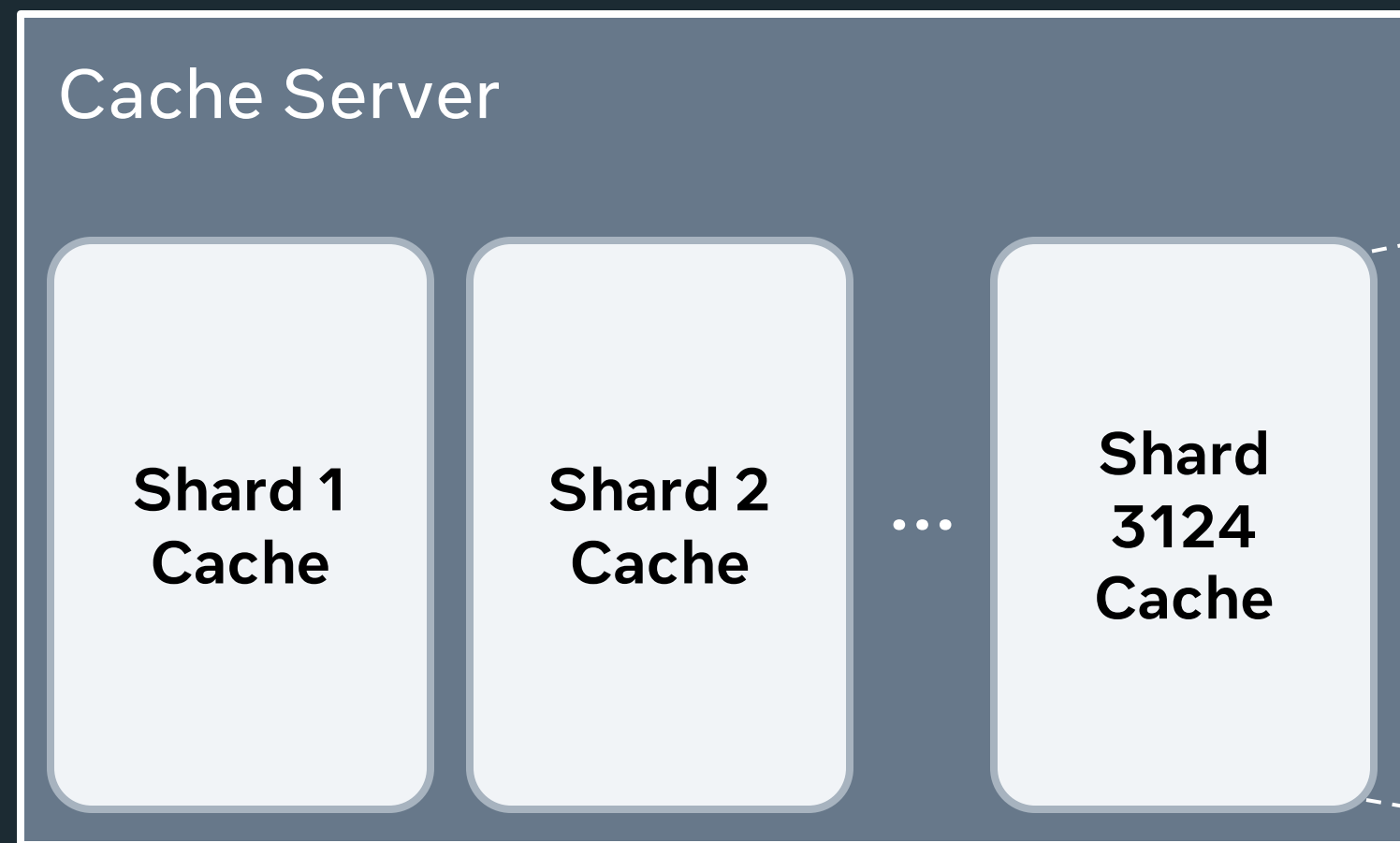
# Challenges



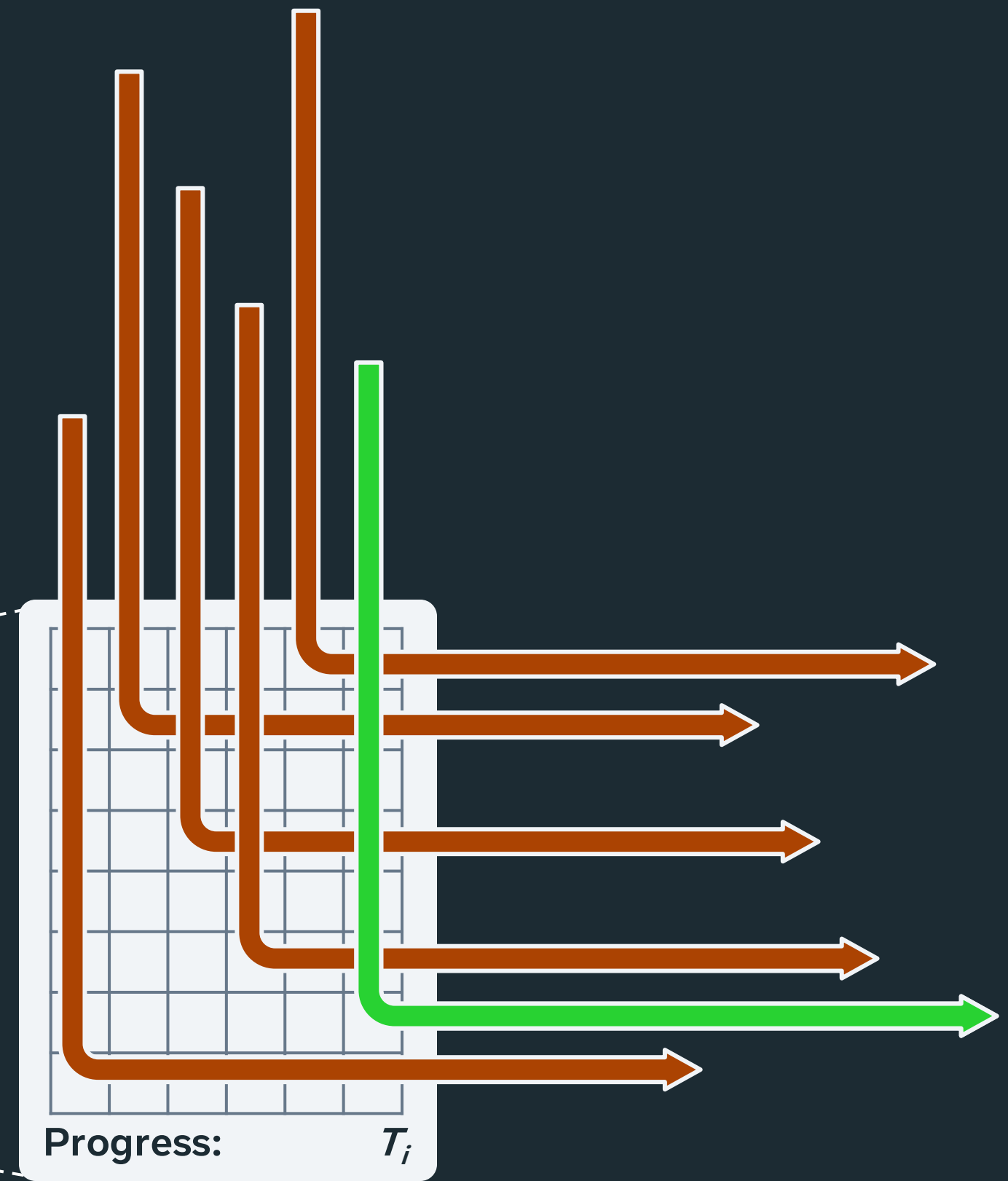
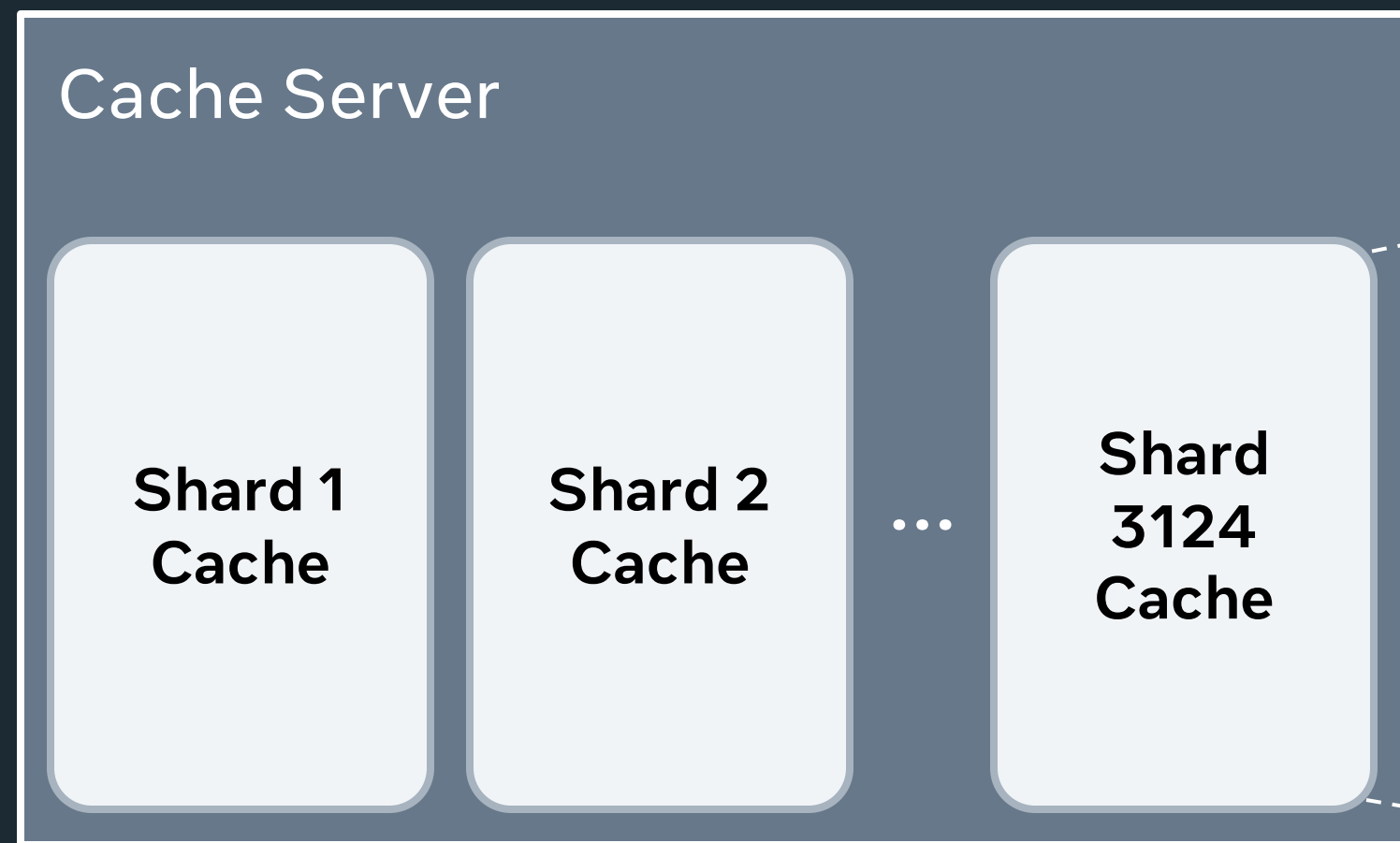
# Challenges



# Challenges

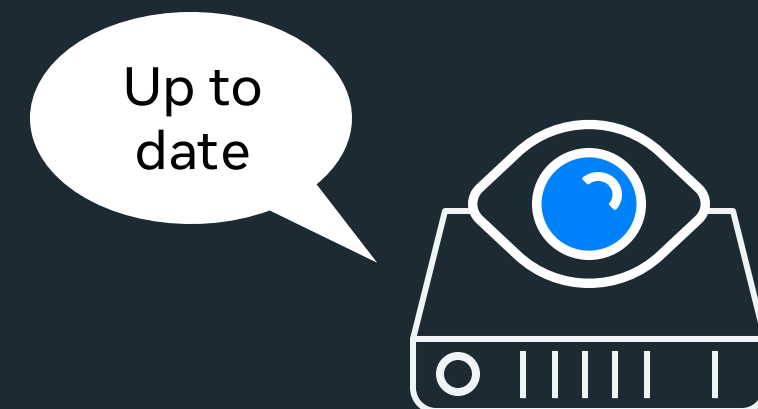
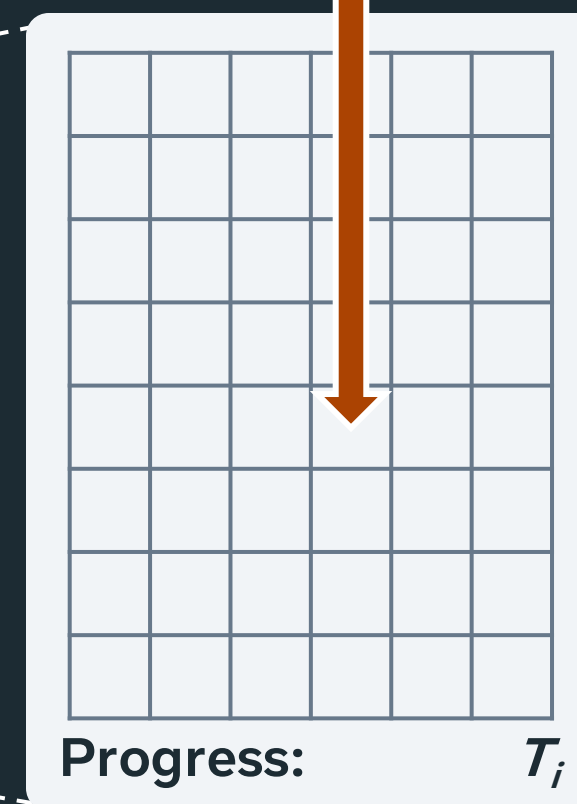
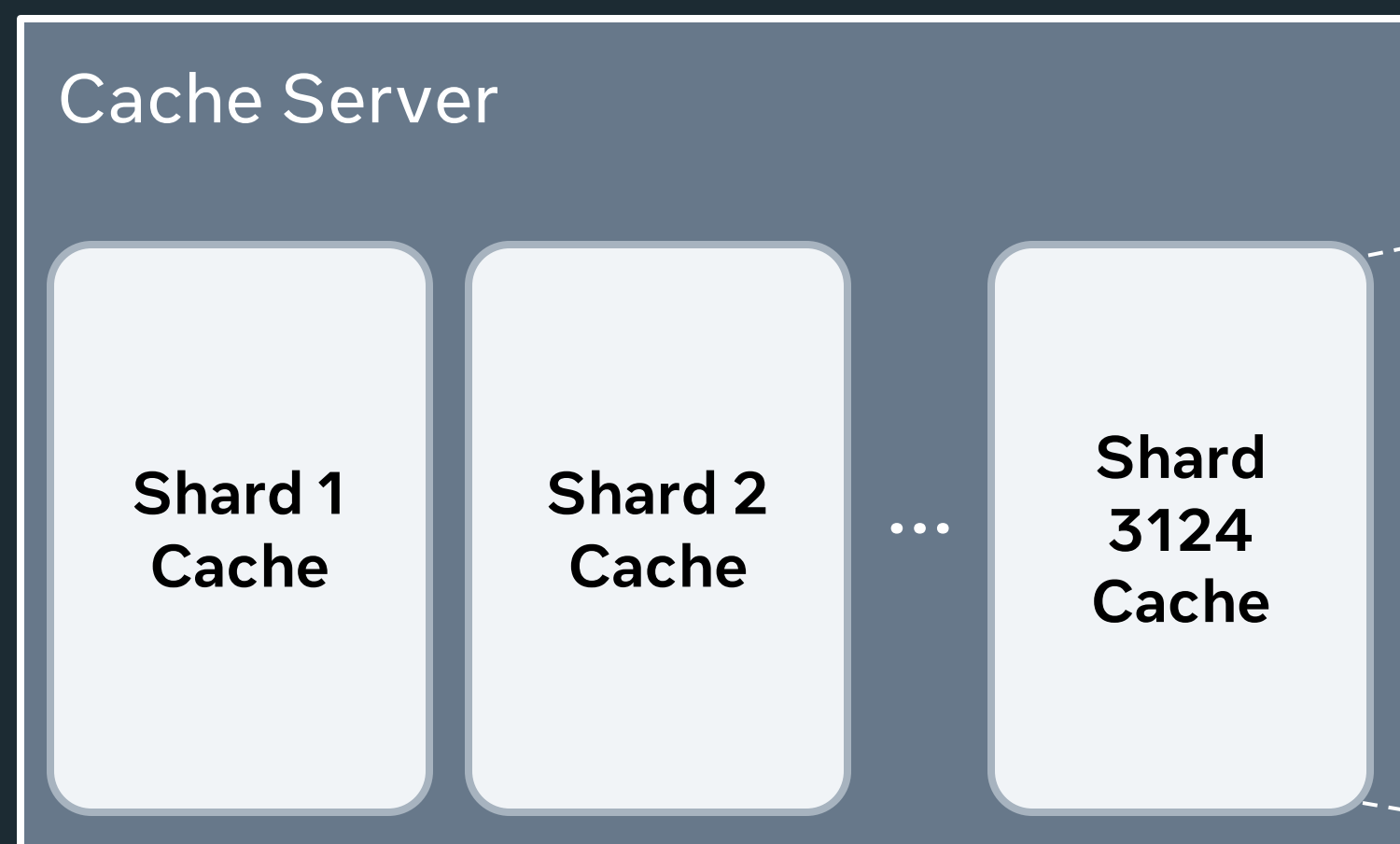


# Challenges

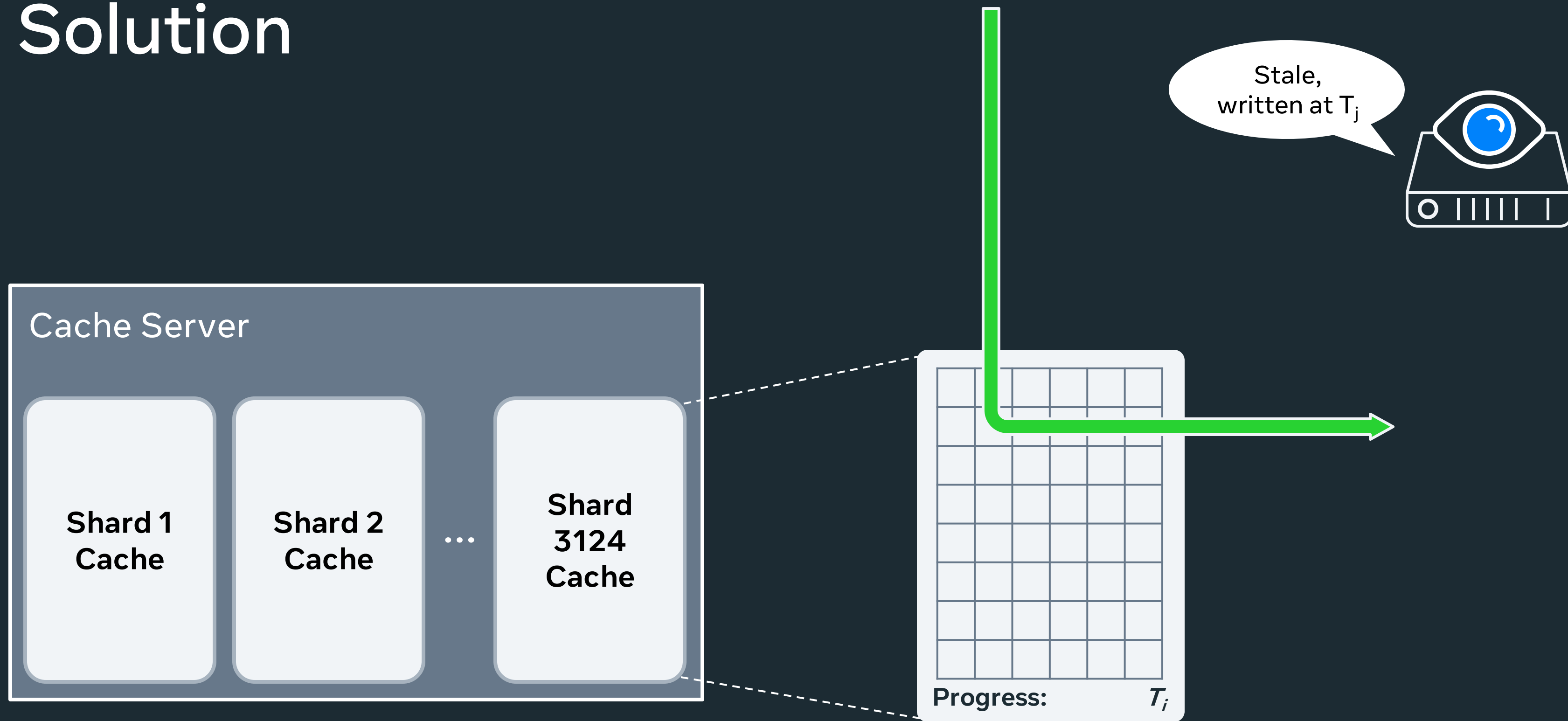




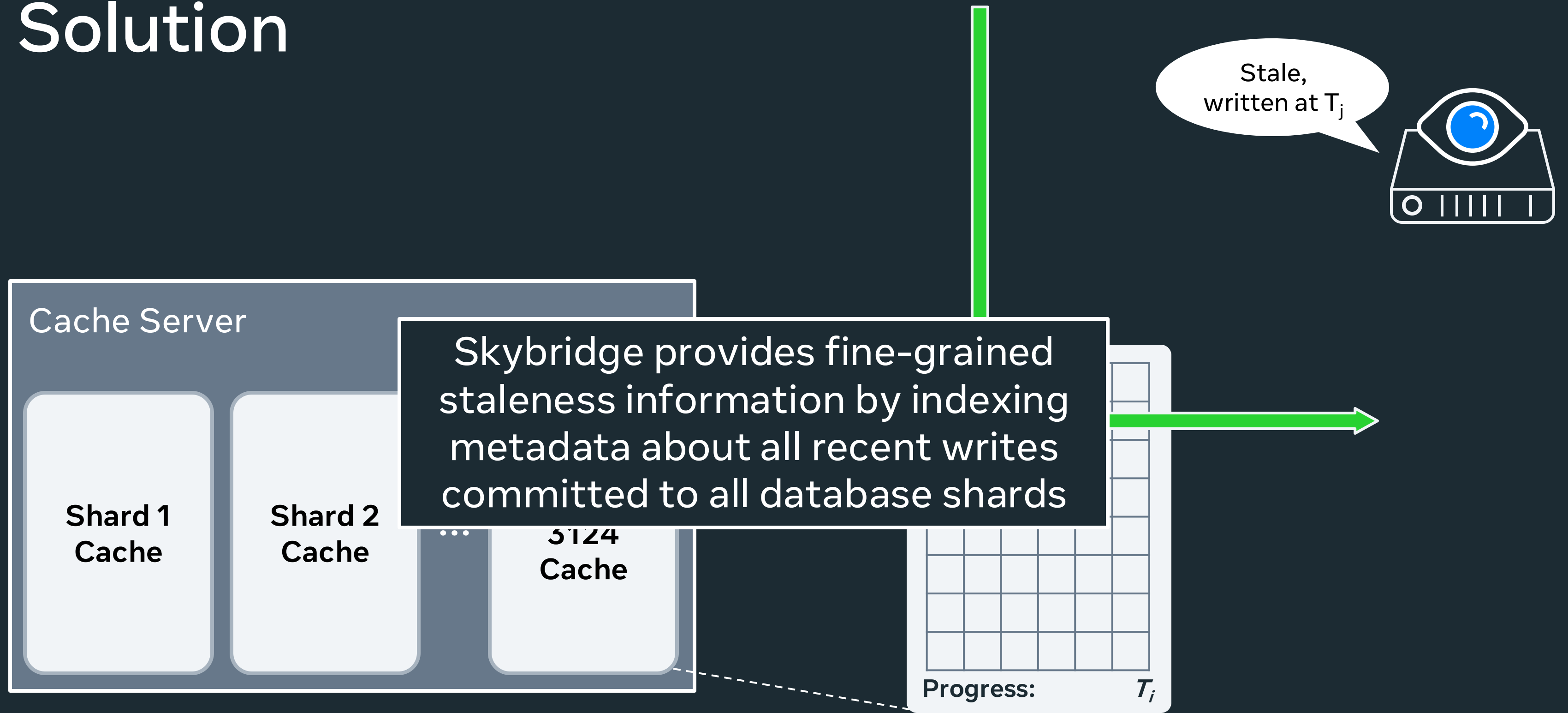
# Solution



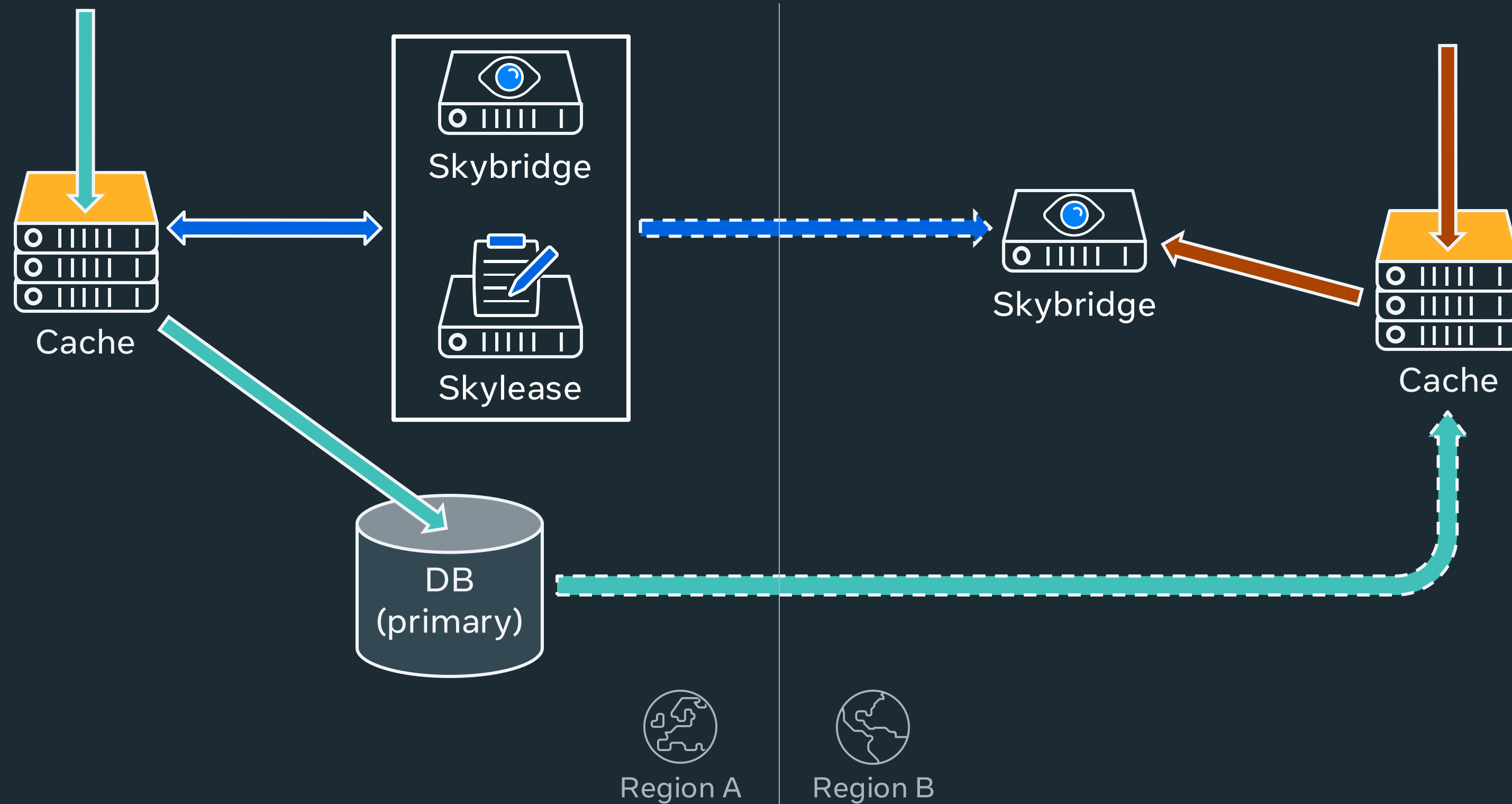
# Solution



# Solution



# Skybridge Architecture



# Replication with Gap Detection



Only replicate write metadata

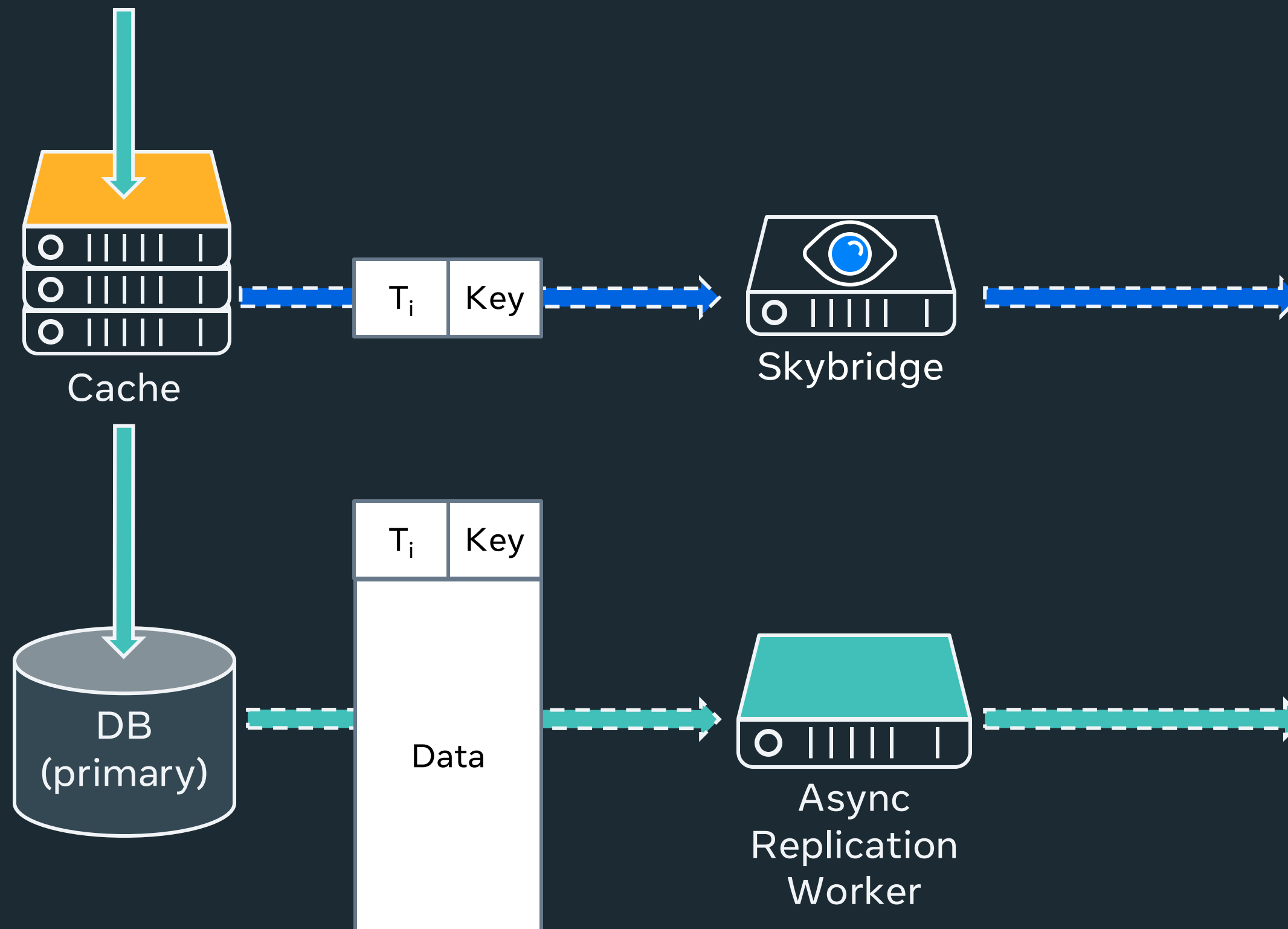


Allow known data loss

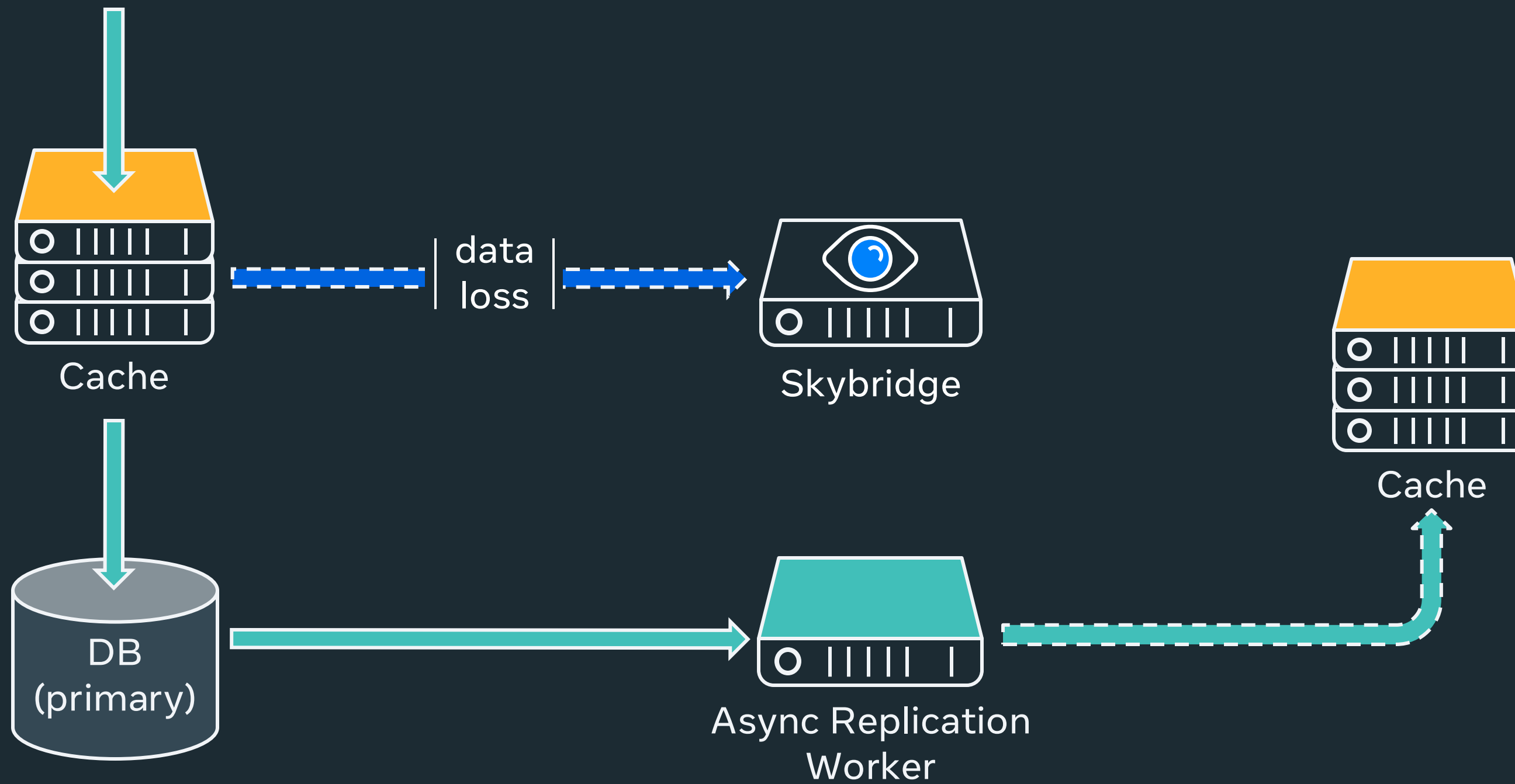


Out-of-order replication

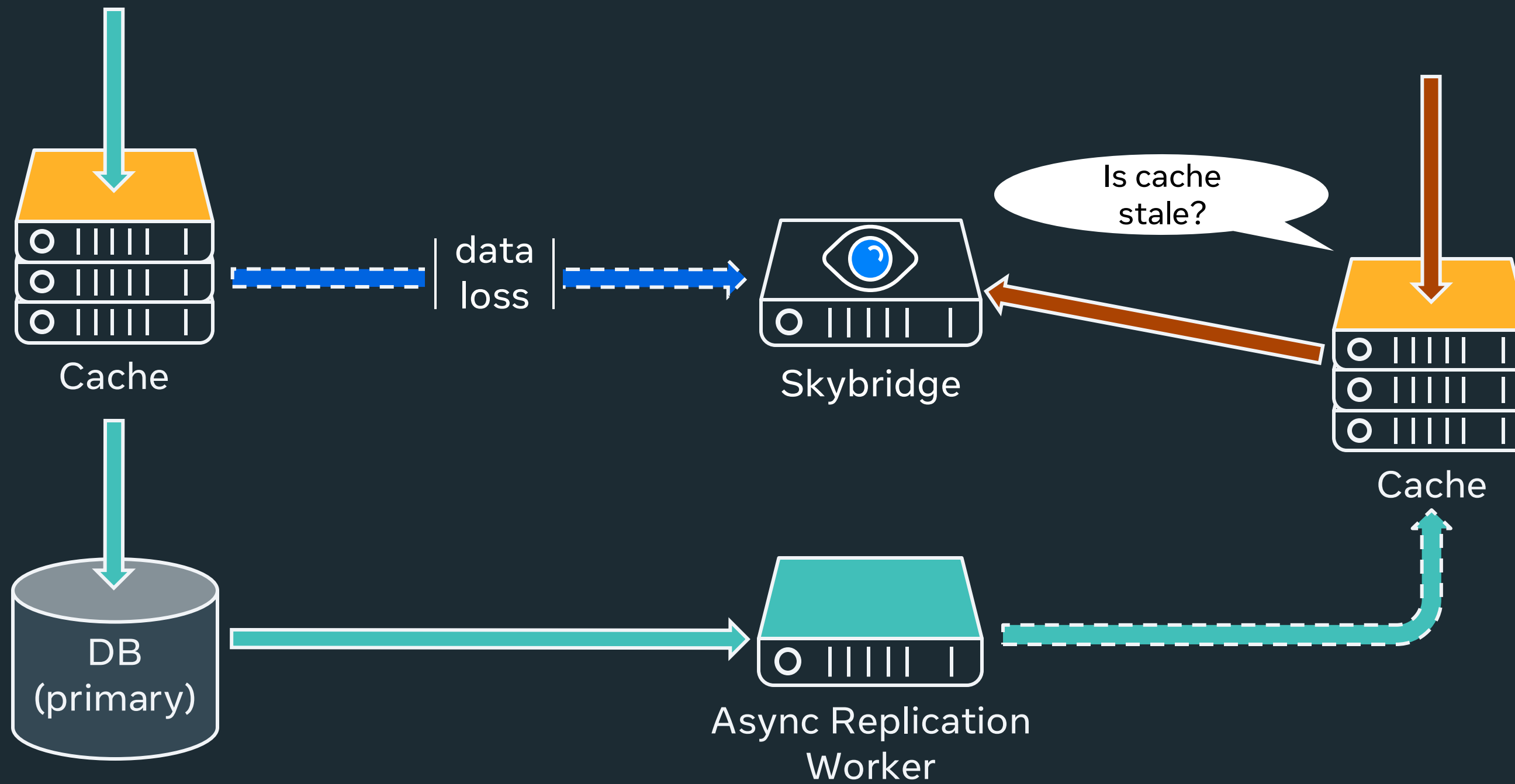
# Only replicate write metadata



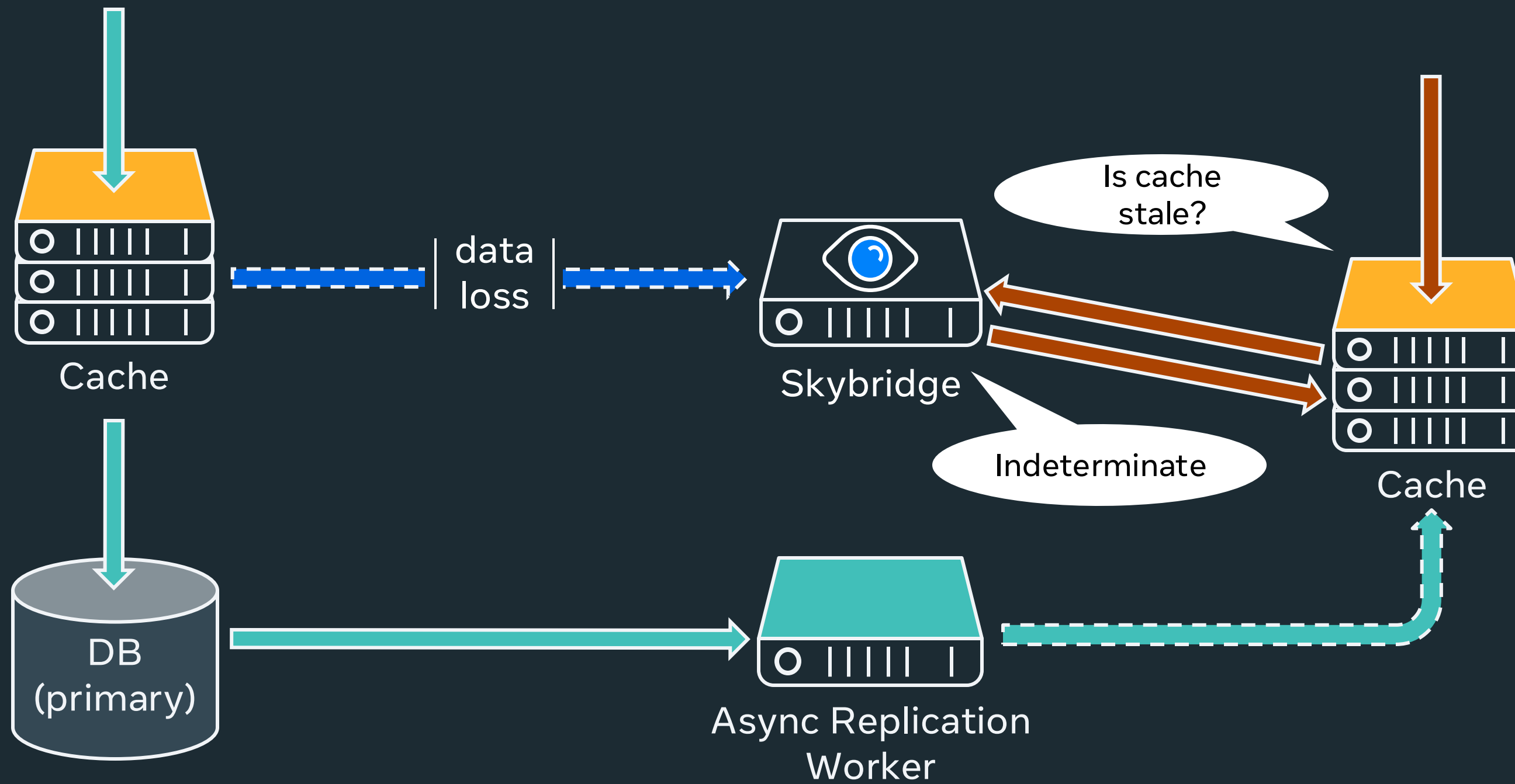
# Allow known data loss



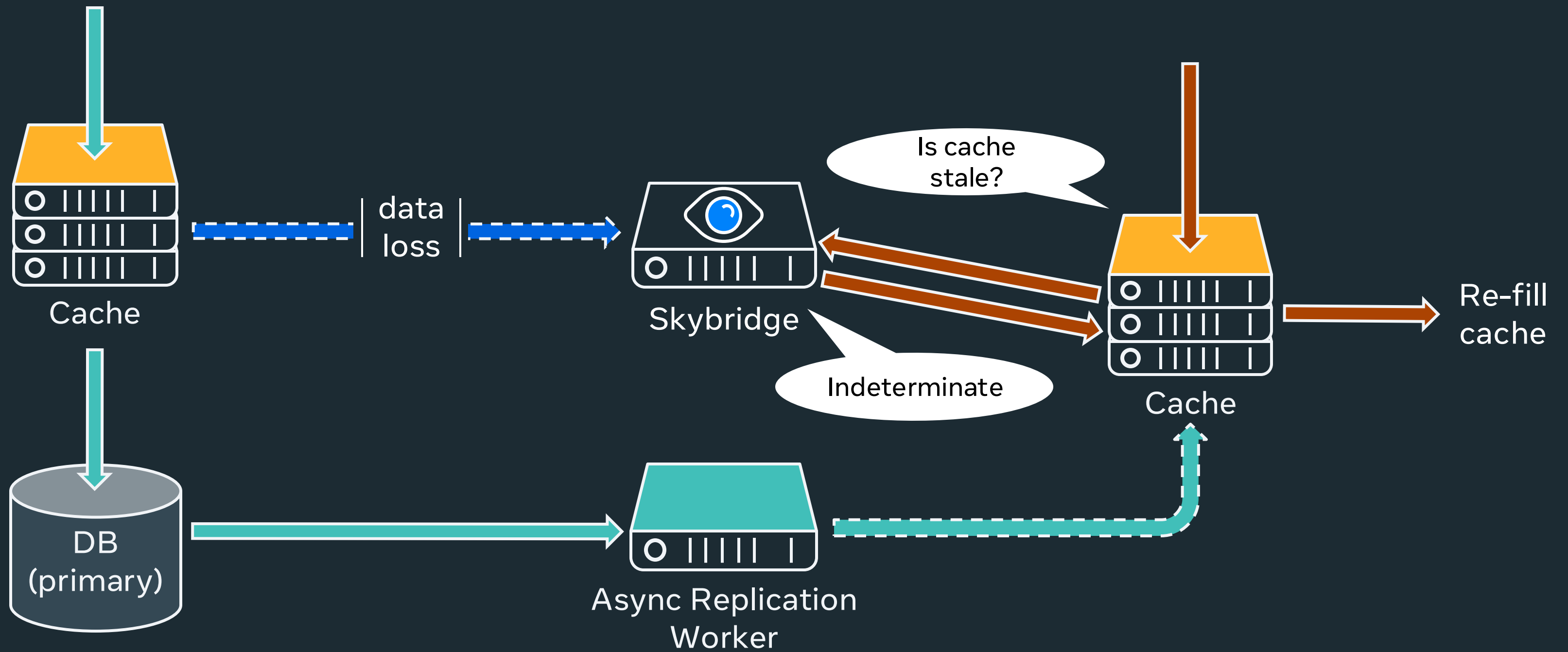
# Allow known data loss



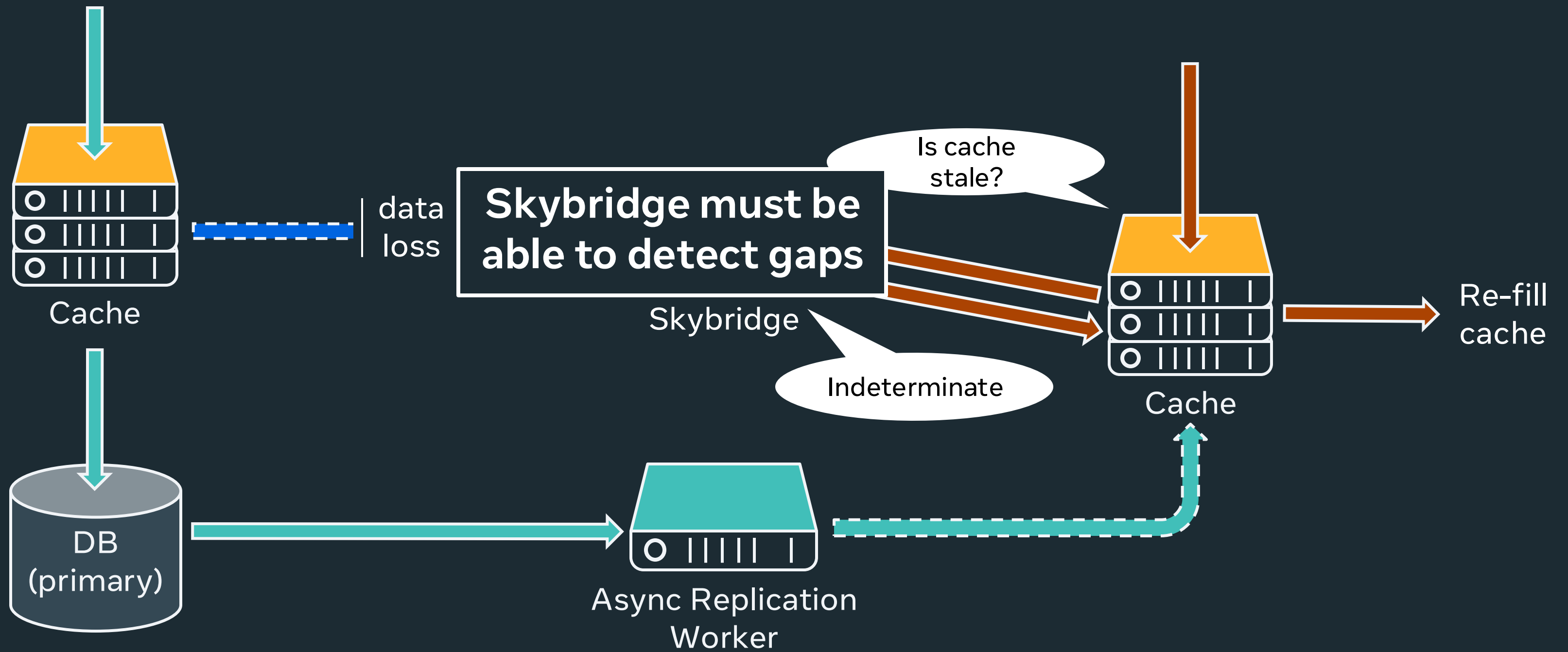
# Allow known data loss



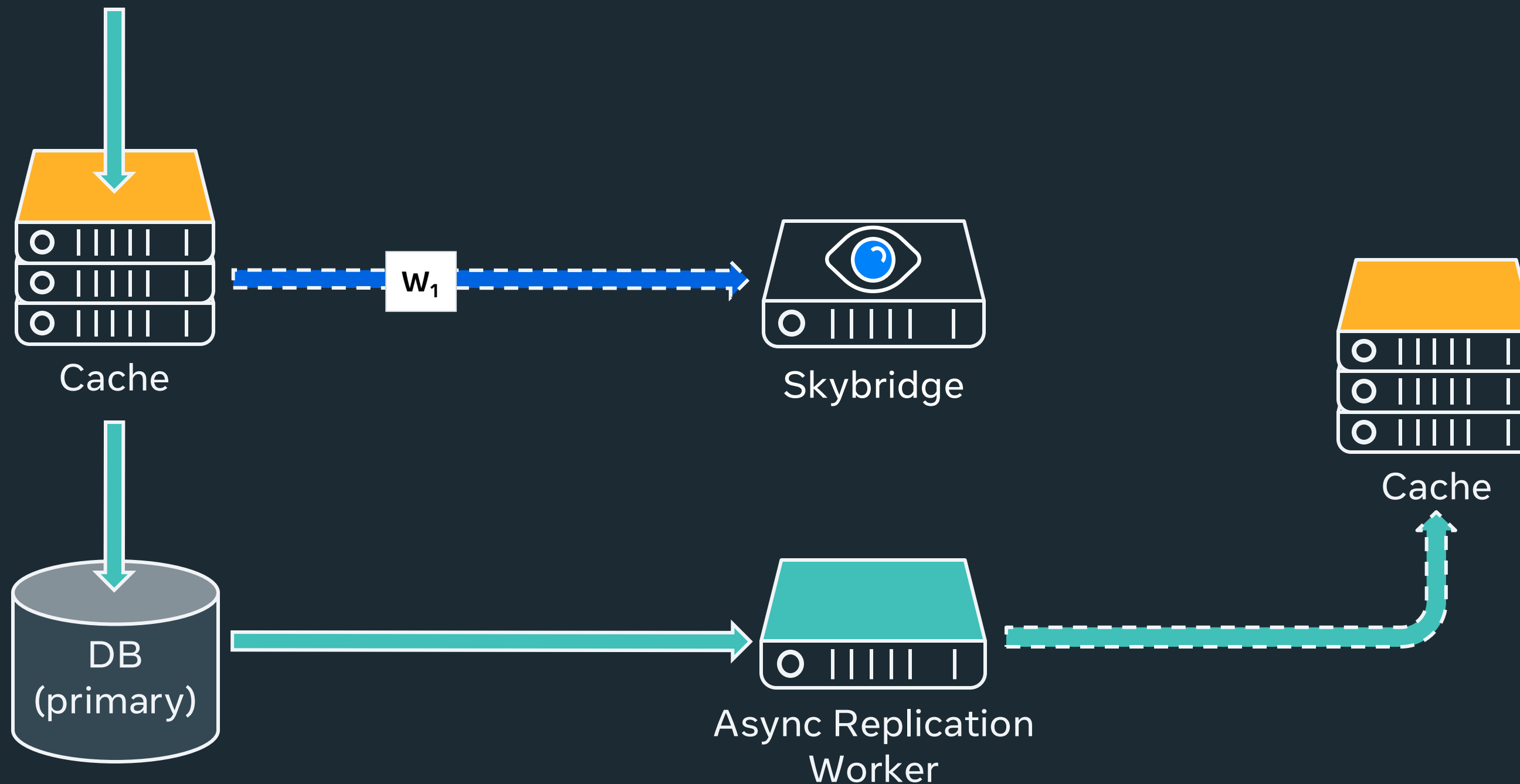
# Allow known data loss



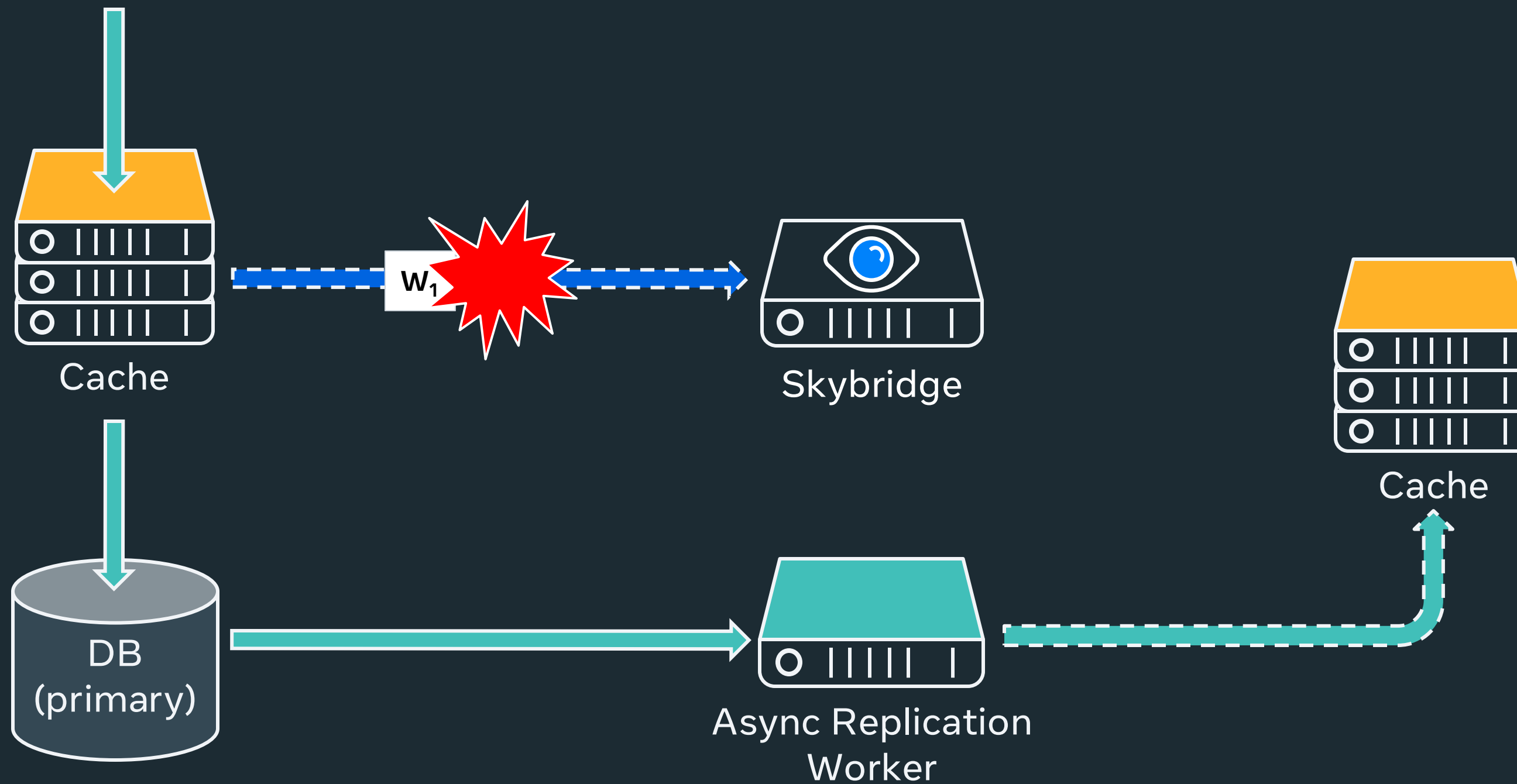
# Allow known data loss



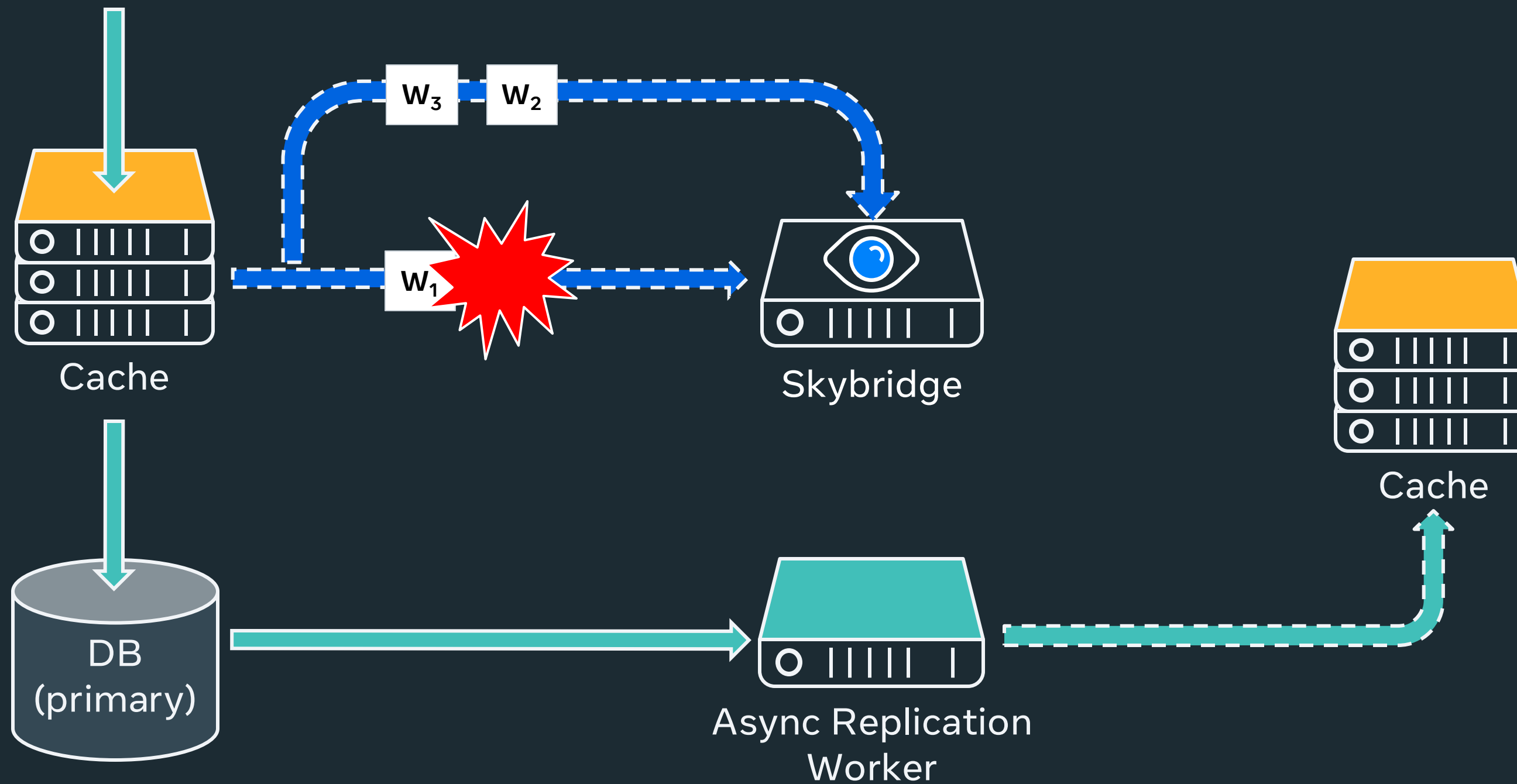
# Out-of-order replication



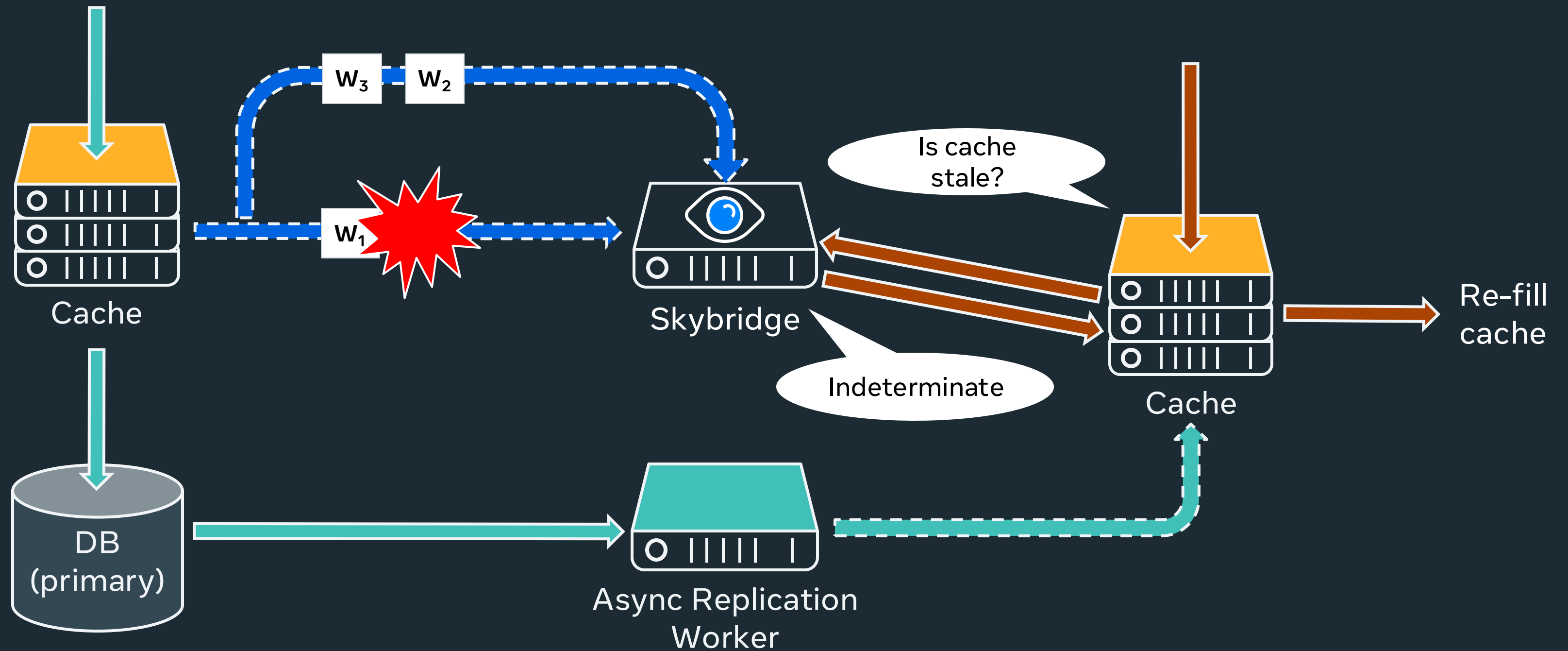
# Out-of-order replication



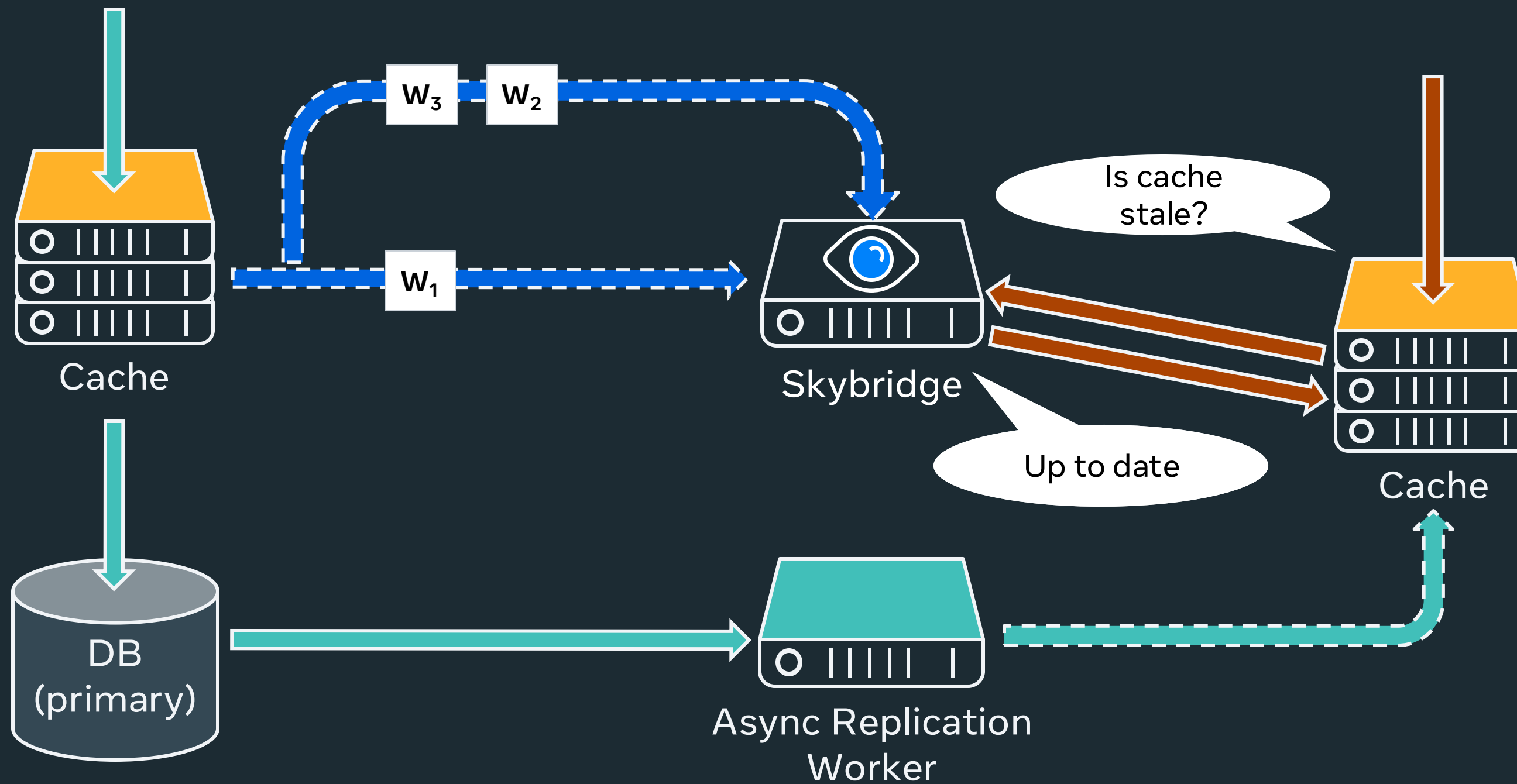
# Out-of-order replication



# Out-of-order replication



# Out-of-order replication



# Implementation

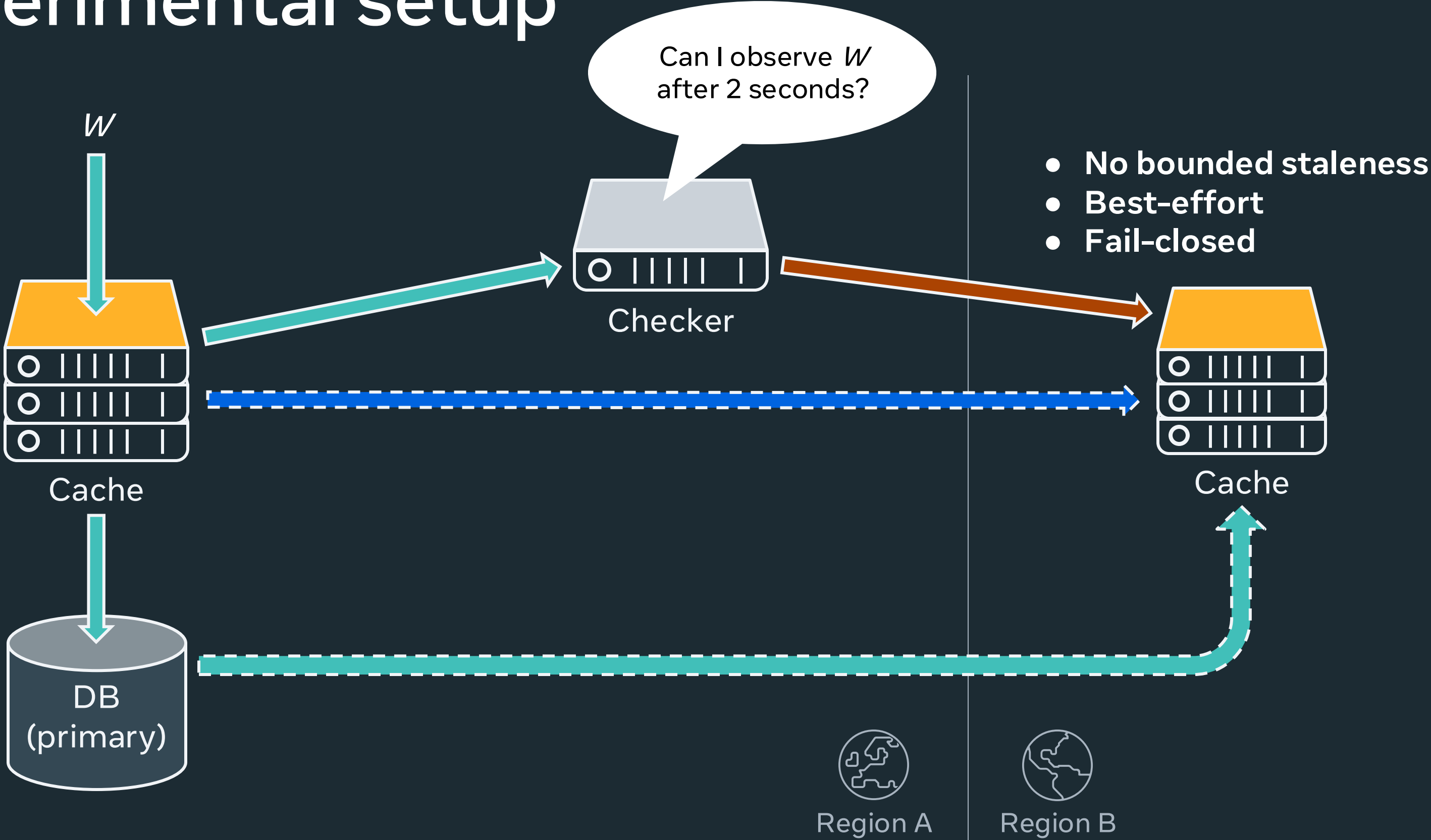


Skybridge provides best-effort bounded staleness by default, with a fail-closed mode for users that require stronger consistency.

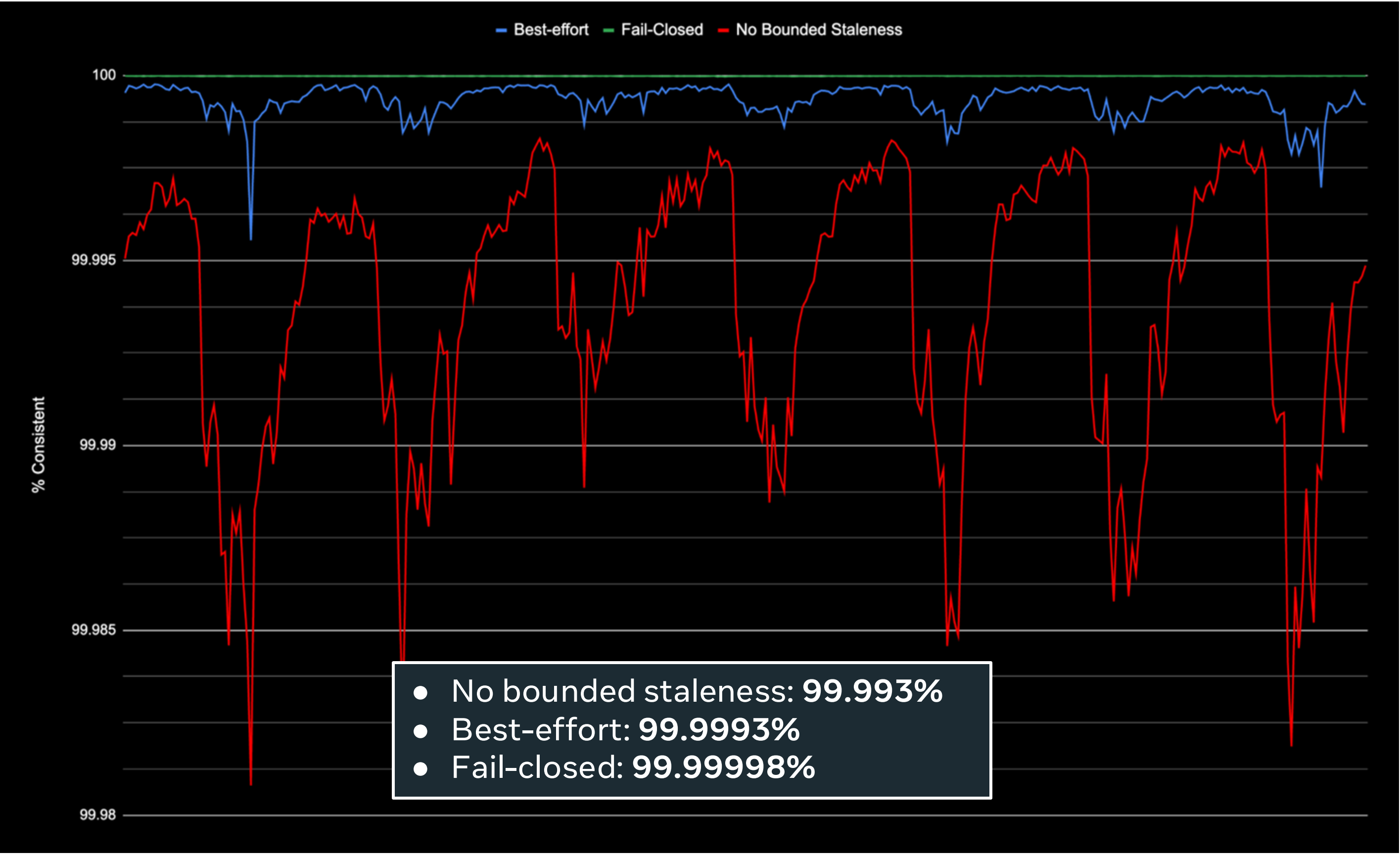


Skybridge targets 2-second bounded staleness

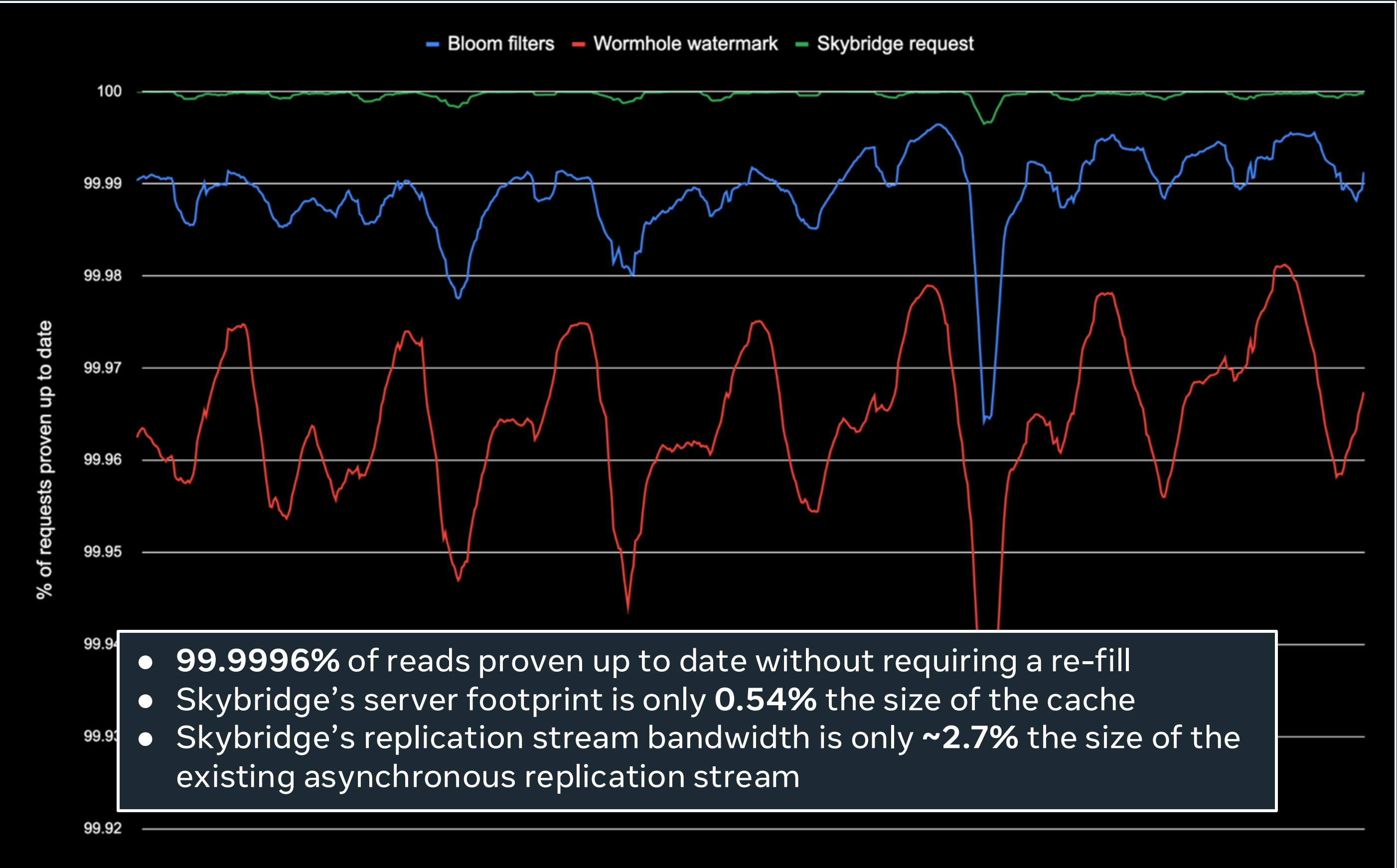
# Experimental setup



# Evaluation



# Evaluation



# Conclusion



Skybridge provides fine-grained information about which cache items are stale



Skybridge is effective, boosting 2-second cache consistency to 99.99998%



Skybridge is lightweight, enabling cache hosts to apply bounded staleness to billions of reads per second

