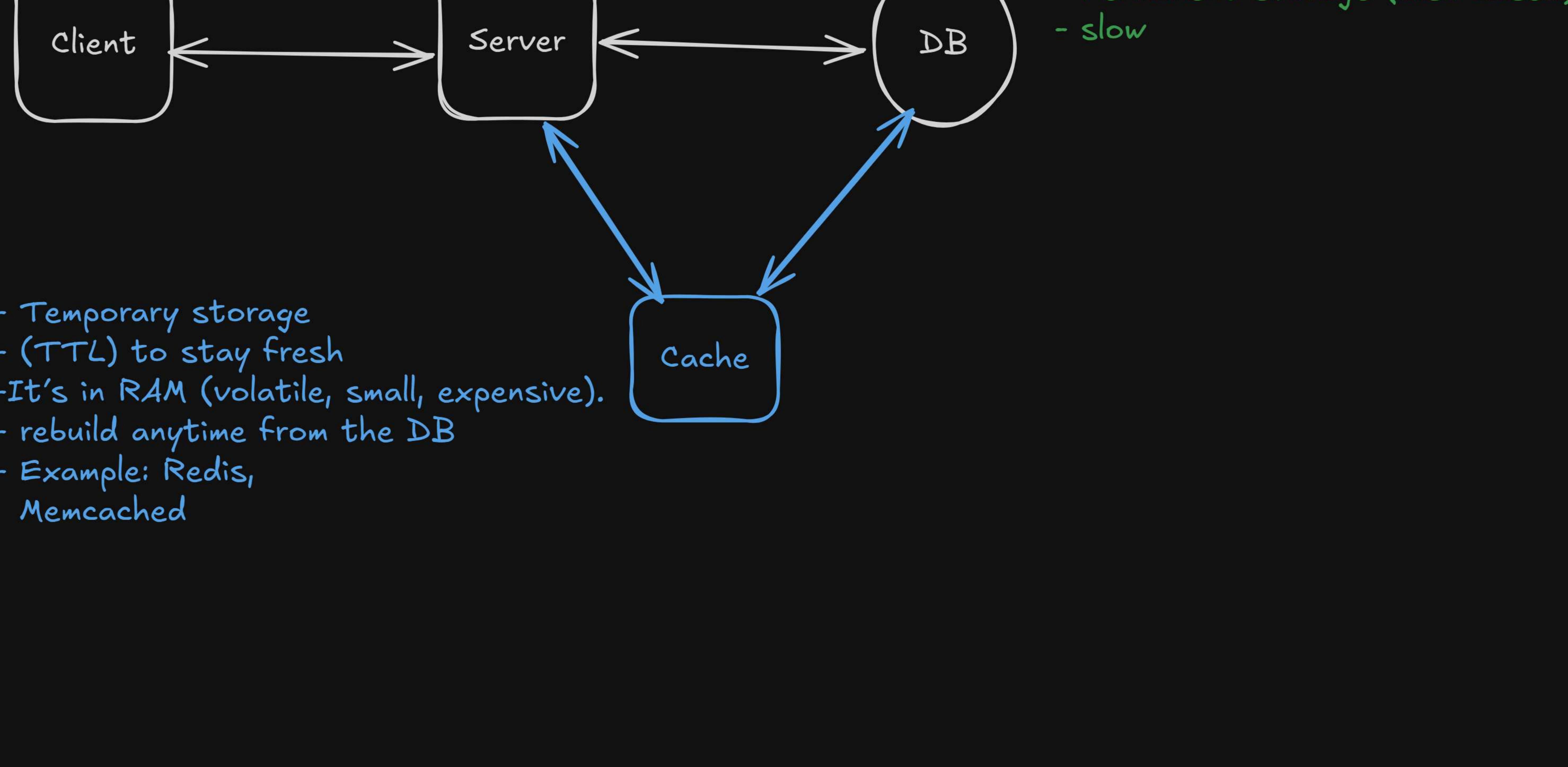


Caching

- storing frequently accessed data or data which takes a lot of time to compute in a faster storage (RAM)



Cache Invalidation

- How to update?

Some data in cache (Profile pic)

Time-to-Live (TTL / Expiry-based)

- set expiry time
- Simple
- Data may be stale So not perfect
- Use case: Product catalog, profile pictures

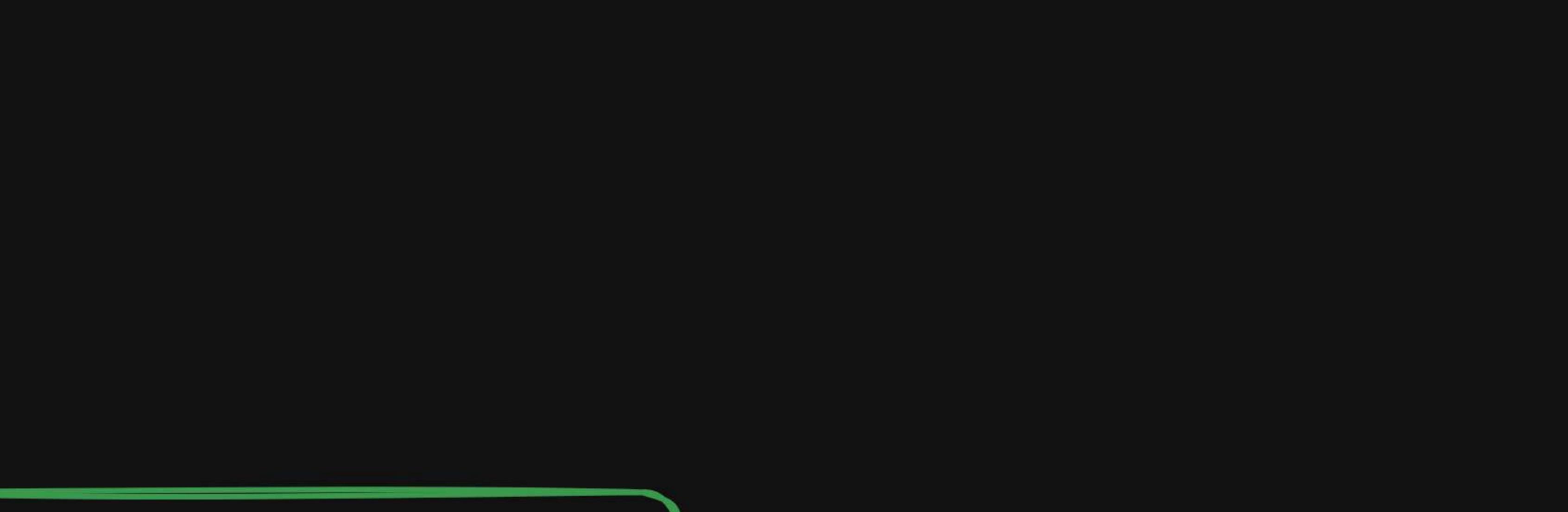
Write-through

- data written at the same time to both DB and cache.
- Consistency guaranteed
- Slower writes
- Use case: account balances.



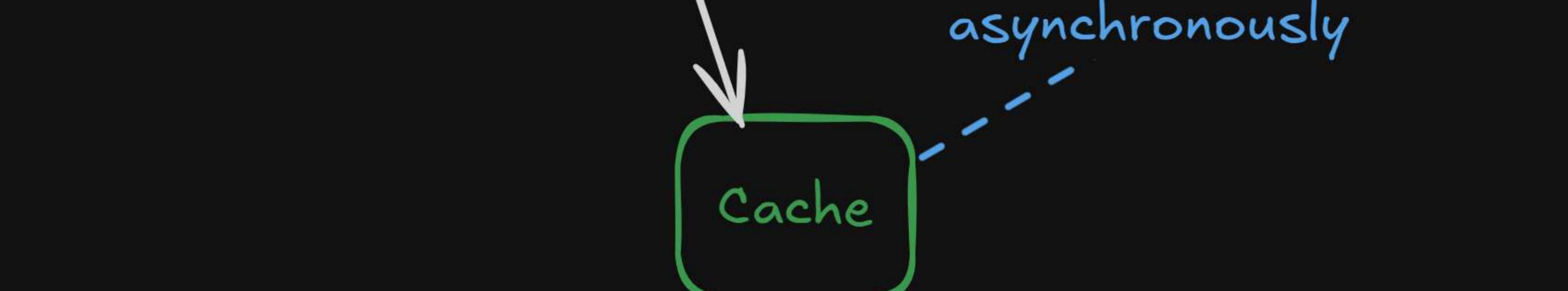
Write-around

- Writes to DB only, not cache

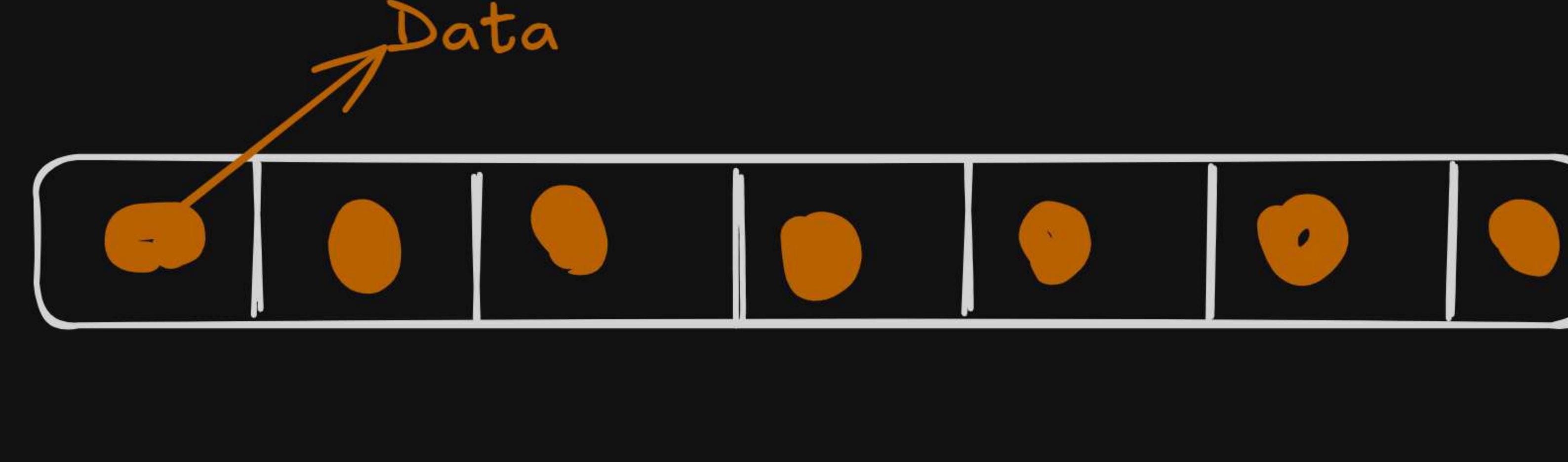


Write-back (Write-behind)

- Writes to cache first
- asynchronously updates DB later
- fast writes
- Data loss risk if cache crashes before DB sync.
- Use case: Logging systems

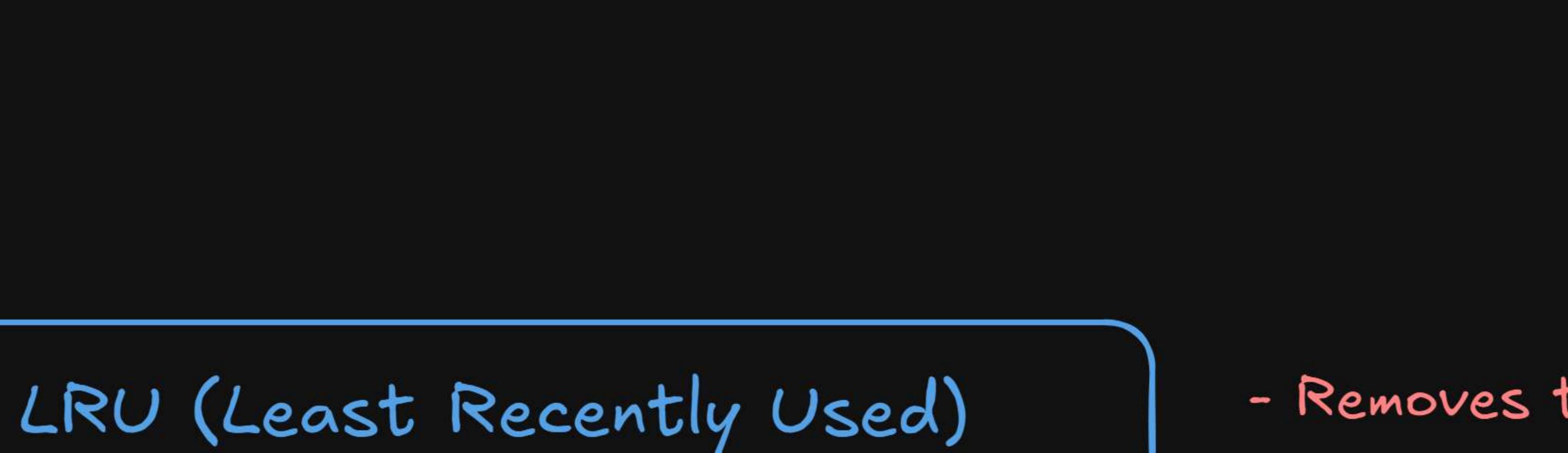
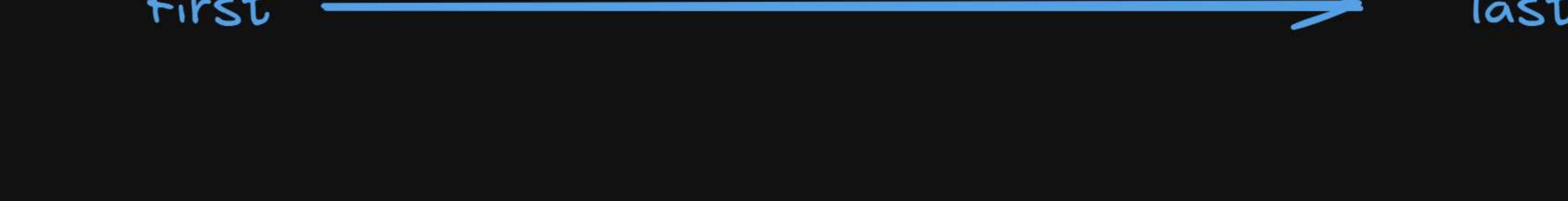


Cache eviction (what to remove when cache gets full?).



Cache Eviction Policies

FIFO (First-In-First-Out)



LRU (Least Recently Used)

- Removes the item which hasn't been used for the longest time



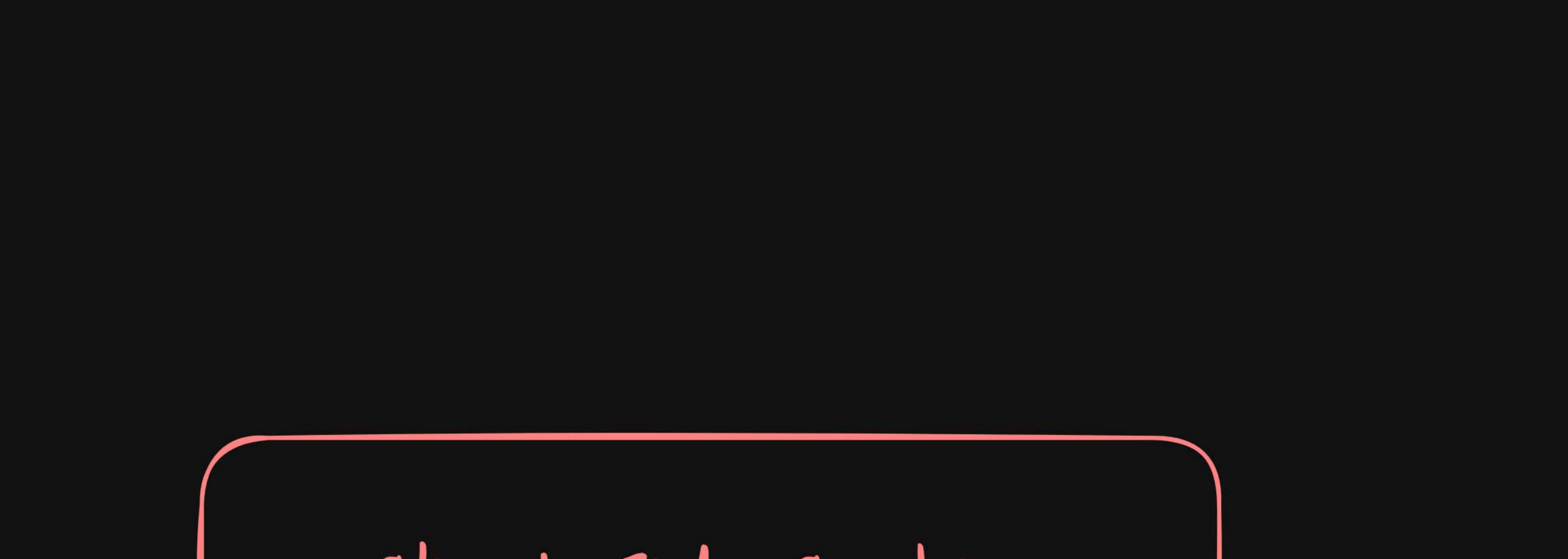
LFU (Least Frequently Used)



CDN (Content Delivery Network)

- network of geographically distributed servers that cache and deliver content (like images, videos, JS, CSS, API Responses) to users from the nearest server.

- Popular CDNs: Akamai, Cloudflare, AWS CloudFront



India

USA

Client-Side Caching

- Caching that happens on the client (browser, mobile app, desktop app) rather than on the server.

Browser / HTTP Caching

LocalStorage

App / Mobile Cache

- for read-heavy apps, offline support, and reducing latency.

- combine with server-side caching and CDNs.