Caching

Caching is used to speed up a system. This is done by storing the results of computationally long operations / network call in a piece of hardware / software where the data recovery would be faster without much latency.

Caching can be done both at the client level (Eg: Network calls) as well as in server level

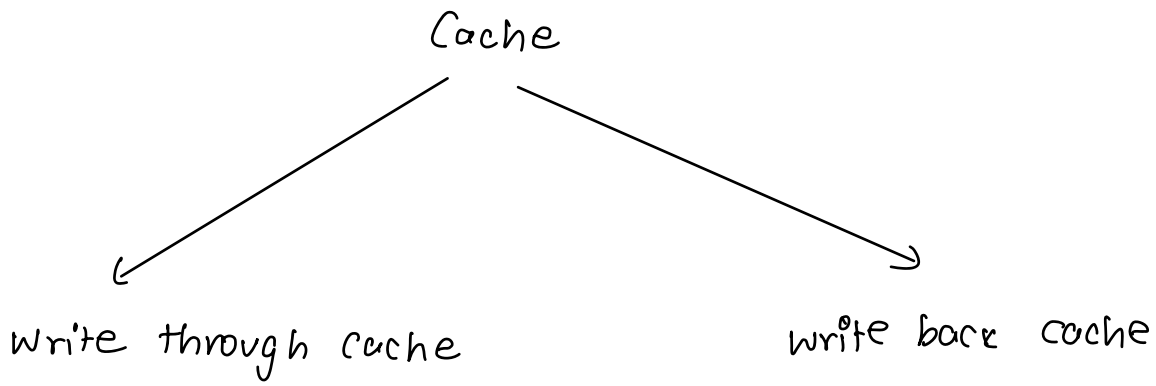
Caching in server level

Assume there is a Instagram celebrity and gets searched a lot. Instead of searching for that celebrity in the DB everytime we can cache the data thus reducing the load on the DB and also provide much faster search results.

Redis → popularly used cache

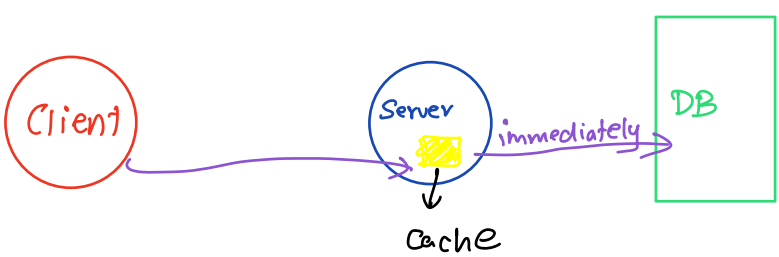
Redis is an open source, in-memory database (IMDB)

(IMDB stores the data in a computers memory → RAM instead of the disk to produce quicker response)



Write Through Cache:

In a write through cache changes made will affect both the cache and the database at the same time



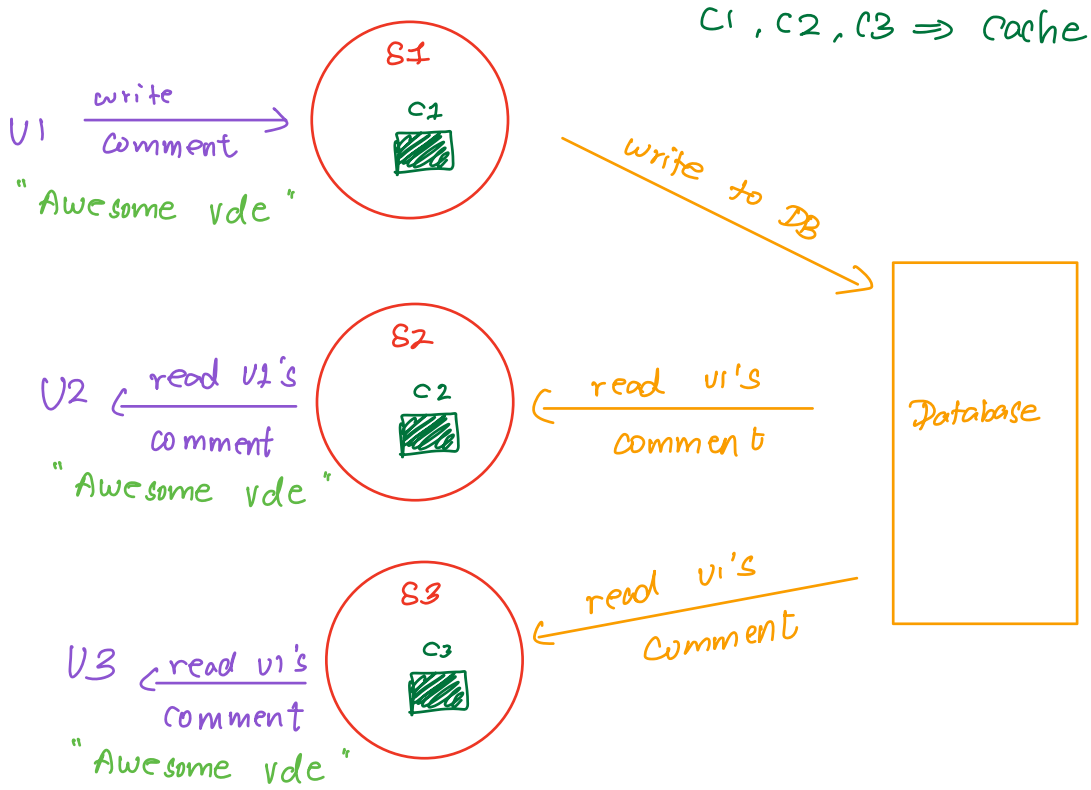
Write Back Cache :

In a write back cache changes made will affect only the cache and the system will asynchronously update the database after some interval (say the DB will be updated every 5 mins). The problem with this method is, if the data in the cache gets corrupted before it is updated in the DB then there will be a permanent loss of data

Stale Cache : => `Cache data != data in DB`

The data in the cache can become stale, if the main source of truth for the data (ie the database) gets updated and cache doesn't. In such a case we have to evict the cache.

Eg : Youtube Comment Section



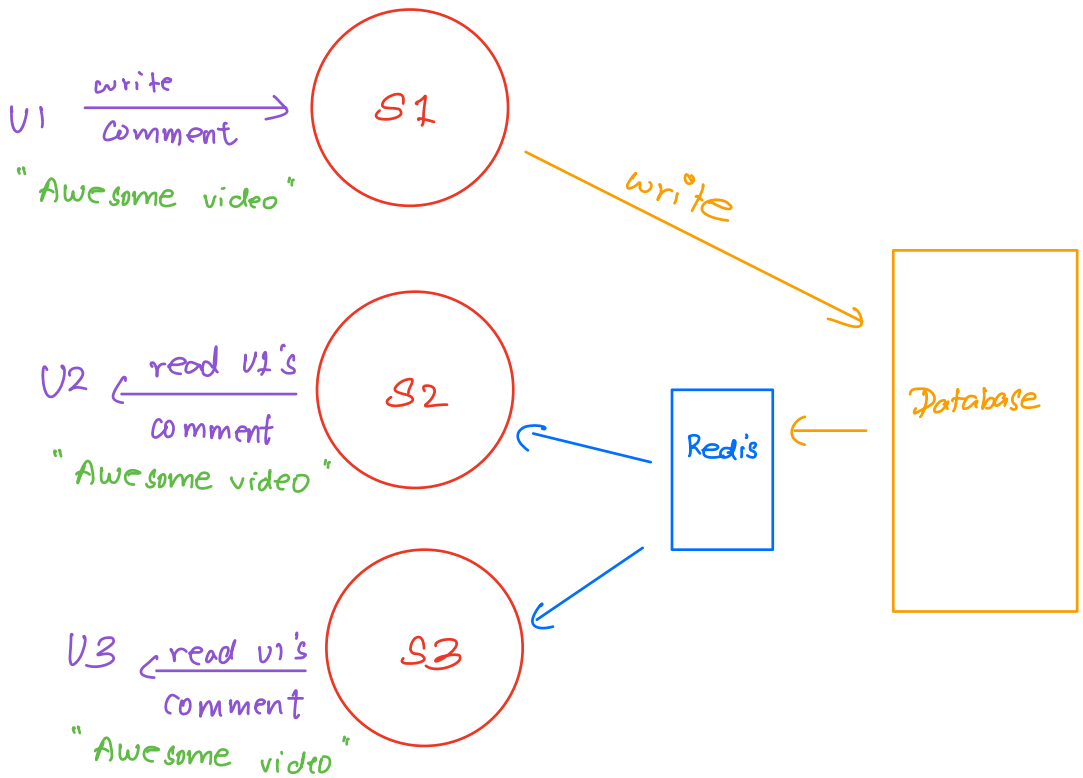
User-1 edits his comment to "Awesome video". His edits are affected in C1 and in DB. But since C2 & C3 (cache-2, cache-3) has V1's comment, the edited data from the DB will not be shown to V2 & V3 instead the outdated cached data "Awesome vde" from C2 & C3 will be shown to V2 & V3.

So the data in C2 and C3 has become stale and needs to be evicted.

Solution : Youtube Comment

Instead of having the cache on

each and every server have a common cache such as Redis.



Cache Eviction Policy

The policy by which values gets evicted or removed from cache

Eg: LRU (least Recently Used)

LFU (least Frequently Used)

FIFO (First In First Out)

Content Delivery Network

A CDN is a third party service that acts like a cache for servers. Sometimes the web application can be slow for users in a particular region if the servers are located in another region. A CDN has servers all around the world meaning the latency of an CDN's server will be better than our server.

A CDN's server is referred to as "Points of Presence" => PoPs

Eg: Cloudflare, Google cloud CDN.