SYSTEM DESIGN

Cheat Sheet

www.skillslash.com

Geo Sharding

Geo sharding is nothing but database sharding based on geo location.

Use-case

1. It is used by ride hailing apps such as Uber, Grab to store live locations of drivers and passengers.
2. It is also used by tinder and other dating sites to provide recommendations based on location.

How to do?

1. Divides the geo locations into cells (for e.g. 1 cell = 100x100 miles square).
2. You get the live location (coordinates lat, long) of user and pass to S2 library, which returns the cell_id in which that location falls into.
3. Use this cell_id as a key to partition the data.

Consistent Hashing

Consistent hashing is the algorithm used by -

1. Most of the load balancers to distribute the load across multiple services.
2. Most of the database horizontal partitioning to distribute the data across multiple databases.

The idea of consistent hashing is to place all the services (or databases) across the consistent hash ring and distribute the load based on some hashing algorithm. This gives us the flexibility to remove or add the service (or database) to the ring without disturbing the whole cluster with minimum or no data loss.

total database in cluster = 5 [DB_1, DB_2, DB_3, DB_4, DB_5]
consistent hashing = 5 databases makes a consistent hash ring

DB_1
↗ ↘
DB_5 DB_2
↖ ↙
DB_4 ⟵ DB_3

- If DB_2 goes down, Next database in the ring i.e. DB_3 is going to take additional load (future writes). If replication is not implemented then existing data of DB_2 will be lost but other database will still work.
- If DB_6 is added newly between DB_5 and DB_1 then it share the 50% load of the next database in the ring i.e. DB_1.

File Storage Options (Image/Video)

Scalability

E-Commerce like Amazon, Social Media like Instagram, Facebook, Twitter, Streaming Provider like Netflix requires to store a huge amount of Images and Videos.
Images and Videos are stored a BLOB (Binary Large Object) and Database Storage is not a good option for this kind of storage. Why?

1. You are not going to query on BLOB Storage
2. You do not need any ACID guarantees for BLOB storage as such provided by RDBMS
3. Costly to store in Database

Cheaper and scalable options to store such files are Distributed File Systems (DFS) such as Sharepoint, Amazon S3.
Details about the image/video such as its DFS URL, metadata can be save in Database, referred by Image ID.

ID	URL	METADATA	IMAGE_THUMBNAIL/VIDEO_COVER_PHOTO

It is also a smart idea to save thumbnail of image, or cover photo of the video in database, size of which should not exceed few KBs. Initially the thumbnail or cover photo can be sent to the client along with the URL of image/video. It is a good experience for the user to see something while image/video is downloading/streaming from DFS.

Availability

You may want to use Content Delivery Network (CDN) which distribute the same images/videos to different locations geographically near to the client locations.

SKILLSLASH

Caching

You need to use Caching at some point of time in your system design for faster read.

Few famous Caching solutions are -

1. Redis(recommended)
2. Memcached
3. Hazelcast

FOLLOW SKILLSLASH

FOR MORE SUCH CONTENT

www.skillslash.com