mailto:shaohua.zhang@antra.com

De-normalization: De-normalization can retrieved data faster than normalization, but less organized than the normalization. How de-normalization works is to put part connected data together into another table. Example, if we have a hotel, we might need a table that have the room number, room information (bedroom, bathroom, level, price …), customers' information (name, birthday, gender, id …). This would be in a normalized way. But let's say if a manager what only check room number and room's level and price. It will need much more time to get all the information in the normalized way(Joining is very expansive for some large databases using). By the using de-normalization, we can put those related data into a table, although it would cause fewer redundancy, but faster retrieve the data. There is a trade of between normalization and de-normalization.

Sharding: Distribute a large amount data to many machines with almost same amount data. Distribute process can cause faster computation.

Replica: A replica is copy the data, the goal is to have the same data set between primary and the coped.

Example can be I have a 3 table in my database : A, B, and C, I can have 3 node for sharding and replica

For sharding, I can 3 sharding node, 1 is A,B, 1 is B and C, and 1 is C and A for evenly Distribute. If one died, the other 2 will maintain the data so there will be no interruption on the database

For replica, both of my primary node and replica will have 3 table. If one died, the replica node will become the primary node.