

September 9, 2025

[]:

partitioning
availability
performance
parallelism
manageability
reduce cost (horizontal scaling cheaper than vertical scaling)

Sharding

What's the difference from partitioning?

technique to do horizontal partitioning ?

we also have to introduce a routing layer / load balancer

bad when your data is spread over different db instances (analytical type of queries, scatter-gather problem)

Sharding is Horizontal scaling

1 How to decide which data go in which shard

1. hash function
 - for an ecommerce website, we have to find orders for a person, so customer id is a good basis to do our sharding in comparison of order id
2. we can store customer id and shard id

2 Advantages

1. scalability more data more queries
2. Fault tolerance

3 Disadvantages

non uniformity of data => reshading

Queries may be across shards

sql db dont support it by default, you have to implement your own logic at the application layer
Sql and vertically scalable, nosql are horizontally scalable

4 ShardKeys

male female
first character of last name

5 tinder

geographic location
gaurav sen