MongoDB Sharding

The application’s database read / write needs to be scaled horizontally (i.e. a single MongoDB instance is unable to manage the write operations, or the dataset is larger than a single MongoDB’s HDD / storage capacity, or size of active working set exceeds capacity of the RAM on the VM).

It’s a means of distributing data across multiple servers for storage. To shard (i.e. adding more servers to a database – combining to make a logical database) will allow a data set to be split and distributed across multiple shards / independent databases. This, in turn, allows for a larger write capacity (i.e. because it’s distributed across multiple databases / shards).

Shard is a single MongoDB instance, shard is a replica set that provides for redundancy and availability (based on a collection basis)

The distribution of the collection data is based on a Shard Key (it’s a field in every document)

Choosing the Shard Key

1. Schema of your Data
2. The way data is queried and written to the DB

Shard Key characteristics

1. Easily Divisible
2. High degree of Randomness
3. Target a Single Shard
4. Use a Compound Shard Key

Range based Sharding

* Data is divided based on ranges of the value of the shard key

Hash based Sharding

* Hash of a field’s value
* Data is evenly distributed, and random distribution across chunks and shards