Chapter 3: Sharding

Lab - Documents in Chunks

Back to the Question

First, connect to the `mongos` as the `m103-admin` user:

```
mongo -u m103-admin -p mypass123 --authenticationDatabase
```

COPY

Use `sh.status()` to confirm that the `products` collection is sharded on `sku`.

Because the values of our shard key are normal integers, it is possible to visually identify which chunk our document belongs in. However, `sh.status()` does not provide the chunk ID necessary.

Query the `config.chunks` collection to identify which chunk contains the document in question. First, ensure you are filtering only those chunks belonging to the `m103.products` namespace:

COPY

```
db.getSiblingDB("config").chunks.find(
  {
    "ns" : "m103.products"
  }
)
```

The result of this query is *all* chunks associated to the sharded `products` collection. We can visually identify which chunk our document belongs in by looking at the `min.sku` and `max.sku` fields, which define the *inclusive* minimum and *exclusive* maximum range of shard key values that are associated to the chunk.

However, for larger datasets, there may be many hundreds or thousands of chunks, making visual identification time consuming or unrealistic. Instead, we can perform a query against the `config.chunks` database to identify the chunk where `min <= sku < max`

COPY

```
db.getSiblingDB("config").chunks.find(
   {
      "ns" : "m103.products",
      $expr: {
         $and : [
          {$gte : [ 21572585, "$min.sku"]},
          {$lt : [21572585, "$max.sku"]}
         ]
```

```
        }
    }
)
```

`$expr` allows us to use aggregation operators and syntax in normal queries. Without `$expr`, we would not be able to use the `$min.sku` and `$max.sku` variable expressions to represent the value of `min` or `max` for any given product. The query returns the single chunk where the document resides:

COPY

```
{
  "_id" : "m103.products-sku_19765188",
  "ns" : "m103.products",
  "min" : {
    "sku" : 19765188
  },
  "max" : {
    "sku" : 22935319
  },
  "shard" : "shard2",
  "lastmod" : Timestamp(1, 4),
  "lastmodEpoch" : ObjectId("5a6103511d9376be96849296")
}
```

The `_id` field's value is the chunk ID needed for this lab.

Proceed to next section