MongoDB Exercise

Link to github:
https://github.com/MiaSimone/Database_Assignments/tree/master/AssignmentDB3/my-react-app

a) **What is sharding in mongoDB?**

If we have a large amount of data then we can divide it onto several machines. This will make it faster to use CRUD operations on the collection.

b) **What are the different components, required to implement sharding?**

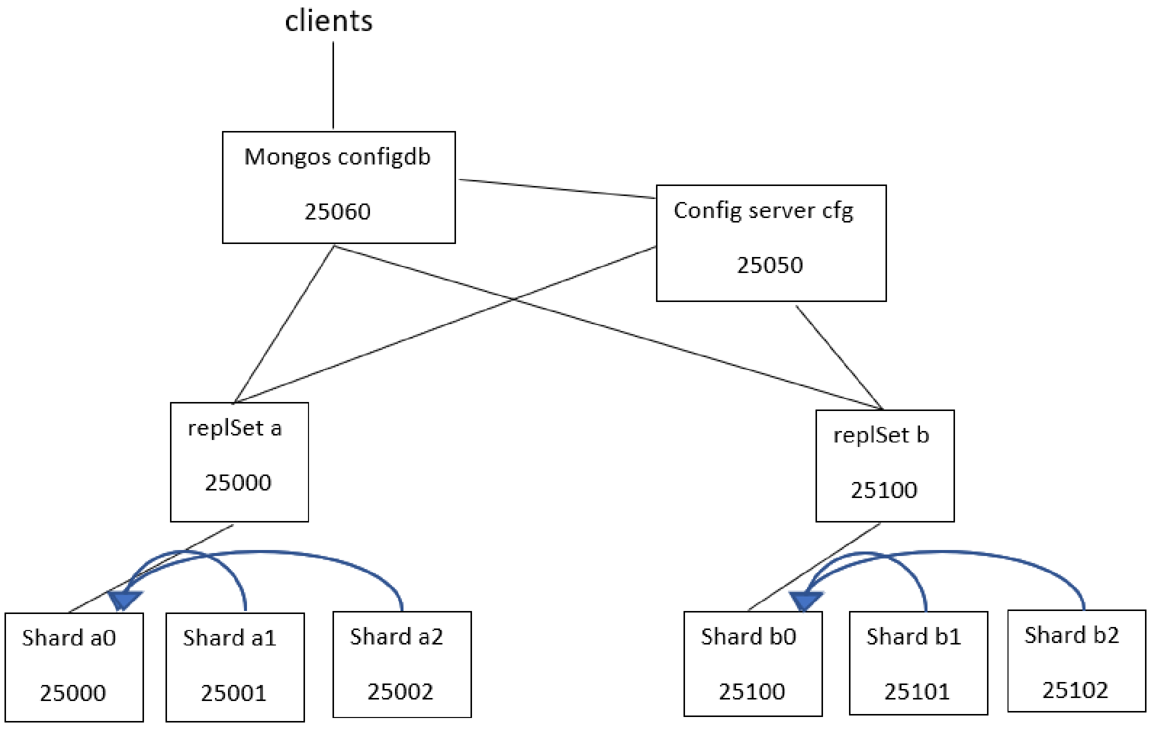
We need a component to be the configuration which the shards will run on. It will have its own port.

Of course we also need some shards. In our setup we created 6 shards and connected them 3 and 3 in replSets. These shards get added onto the config server.

The last component we need is the config db, created from the config server, but runs on another port than the config server.

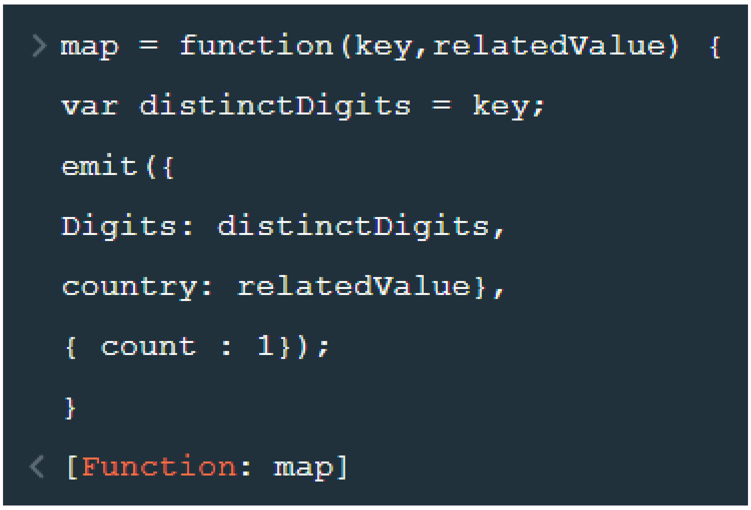
c) **Explain architecture of sharding in mongoDB?**

The architecture of sharding in mongoDB, is where there is a config file that holds the routing to the different replica files and the replica files have different shards. It can be only one shard and it can also be multiply shard. Every shard holds different data to process the data faster and more efficiently.

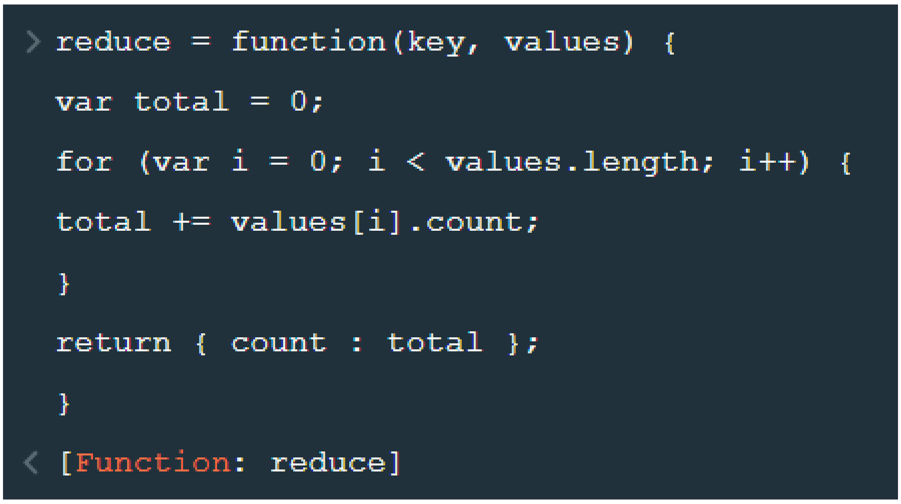


The image above is how the architecture is in our setup. (The numbers fx 25000 is the ports they run on).

d) **Provide implementation of map and reduce function**

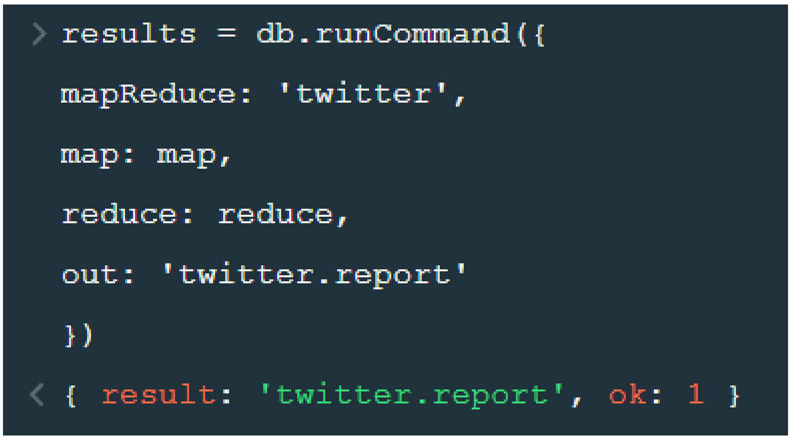


```
> map = function(key,relatedValue) {
  var distinctDigits = key;
  emit({
  Digits: distinctDigits,
  country: relatedValue},
  { count : 1});
  }
< [Function: map]
```

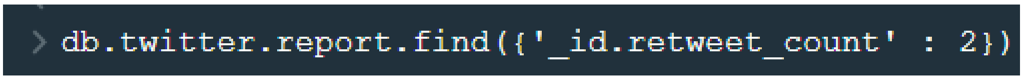


```
> reduce = function(key, values) {
  var total = 0;
  for (var i = 0; i < values.length; i++) {
  total += values[i].count;
  }
  return { count : total };
  }
< [Function: reduce]
```

e) **Provide execution command for running MapReduce**

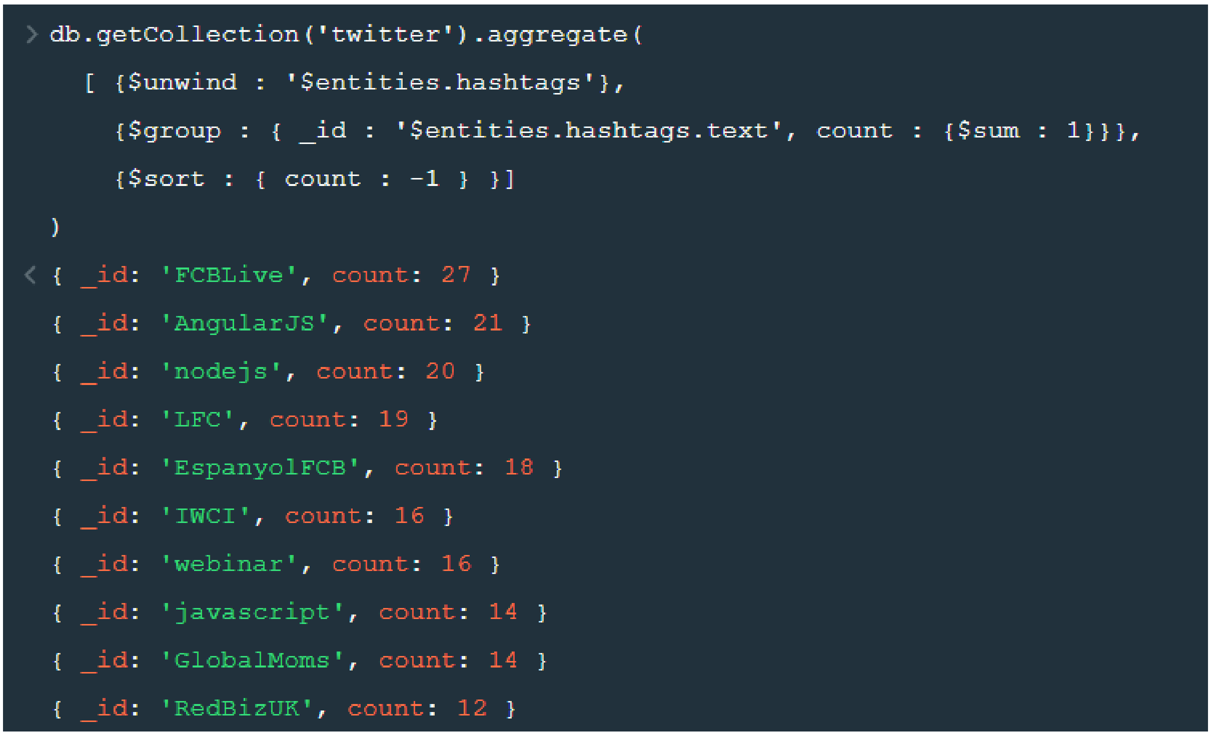


```
> results = db.runCommand({
  mapReduce: 'twitter',
  map: map,
  reduce: reduce,
  out: 'twitter.report'
  })
< { result: 'twitter.report', ok: 1 }
```



```
> db.twitter.report.find({'_id.retweet_count' : 2})
```

f) Provide top 10 recorded out of the sorted result. (hint: use sort on the result returned by MapReduce)



```
> db.getCollection('twitter').aggregate(
    [ {$unwind : '$entities.hashtags'},
      {$group : { _id : '$entities.hashtags.text', count : {$sum : 1}}},
      {$sort : { count : -1 } }]
  )
< { _id: 'FCBLive', count: 27 }
  { _id: 'AngularJS', count: 21 }
  { _id: 'nodejs', count: 20 }
  { _id: 'LFC', count: 19 }
  { _id: 'EspanyolFCB', count: 18 }
  { _id: 'IWCI', count: 16 }
  { _id: 'webinar', count: 16 }
  { _id: 'javascript', count: 14 }
  { _id: 'GlobalMoms', count: 14 }
  { _id: 'RedBizUK', count: 12 }
```

The 10 most used hashtags are

1. FCBLive
2. AngularJS
3. nodejs
4. LFC
5. EspanyolFCB
6. IWCI
7. webinar
8. javascript
9. GlobalMoms
10. RedBizUK

Sharding:

Jacob er config, Mia og Christian er shards

Jacob:

- *mkdir cfg*
- Starts mongo server: *mongod --configsvr --dbpath "%HOMEPATH%\Desktop\cfg" --port 25050 --replSet cfg*
- *mongo --port 25050*
- *rs.initiate()*
- *exit*

Mia:

- *mkdir a0 a1 a2*
- *mongod --shardsvr --replSet a --dbpath "%HOMEPATH%\Documents\mongo\a0" --port 25000*
- *mongod --shardsvr --replSet a --dbpath "%HOMEPATH%\Documents\mongo\a1" --port 25001*
- *mongod --shardsvr --replSet a --dbpath "%HOMEPATH%\Documents\mongo\a2" --port 25002*
- *mongo --port 25000*
- *rs.initiate()*
- *rs.add("localhost:25001")*
- *rs.add("localhost:25002")*
- *exit*

Christian:

- *mkdir b0 b1 b2*
- *mongod --shardsvr --replSet b --dbpath "%HOMEPATH%\Documents\mongo\b0" --port 25100*
- *mongod --shardsvr --replSet b --dbpath "%HOMEPATH%\Documents\mongo\b1" --port 25101*
- *mongod --shardsvr --replSet b --dbpath "%HOMEPATH%\Documents\mongo\b2" --port 25102*
- *mongo --port 25100*
- *rs.initiate()*
- *rs.add("localhost:25101")*
- *rs.add("localhost:25102")*
- *exit*

Jacob (Først når ovenstående er færdigt):

- *mongos --configdb "cfg/localhost:25050" --port 25060*
- *mongo --port 25060*
- *sh.addShard("a/localhost:25000")*
- *sh.addShard("b/localhost:25100")*
- *exit*

Man skal lave en db på port 25060 hvor twitter datasættet ligger. Mongos finder selv ud af på hvilke shards dataen skal ligge.

To run the program:

1. Clone repo https://github.com/MiaSimone/Database_Assignments
2. cd into AssignmentDB3
3. cd into my-react-app
4. In terminal for backend:
   - cd backend
   - npx nodemon
5. In terminal for frontend:
   - From /backend -> cd ..
   - npm start