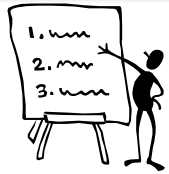
Data Management

MongoDB

Prof. Alejandro Zunino

Prof. Alfredo Teyseyre

Contents



No SQL

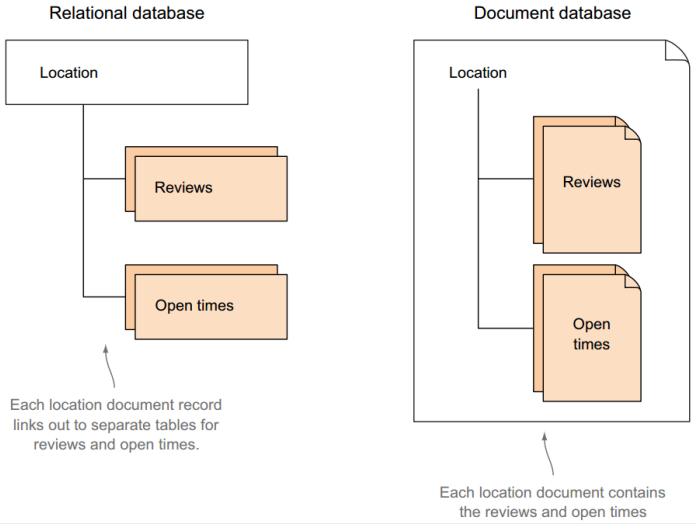
A variety of storage solutions were designed

- Key-value
- Graph database
- Document-oriented
- Column family

Document storage solutions that were designed for better availability, simple querying, and horizontal scaling

Document Storage

Storing your data as holistic documents will allow faster read operations since your application won't have to rebuild the objects with every read



Document Storage

Although there are a few document-oriented databases, none are as popular as MongoDB

- Faster read operations
- Model changes
  - Schemaless
    - different properties for different objects
  - e-commerce application: furniture
    - large number of empty columns (relational)

Features

- Document-Oriented storage
- Full Index Support
- Replication & High Availability
- Auto-Sharding
- Querying
- Fast In-Place Updates
- Map/Reduce

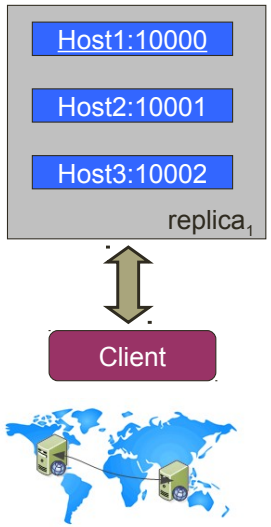
Agile

Scalable

Replica Sets

To provide data redundancy and improved availability, MongoDB uses an architecture called replica set

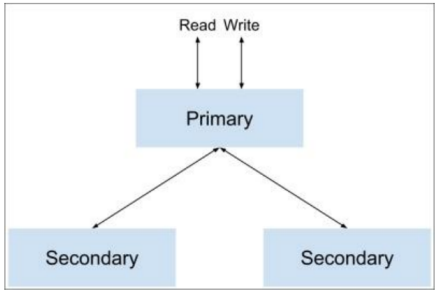
- Redundancy and Fail over
- Zero downtime for upgrades and maintenance
- Master-slave replication
  - Strong Consistency
  - Delayed Consistency
- Geospatial features



Replica Sets

To provide data redundancy and improved availability, MongoDB uses an architecture called replica set

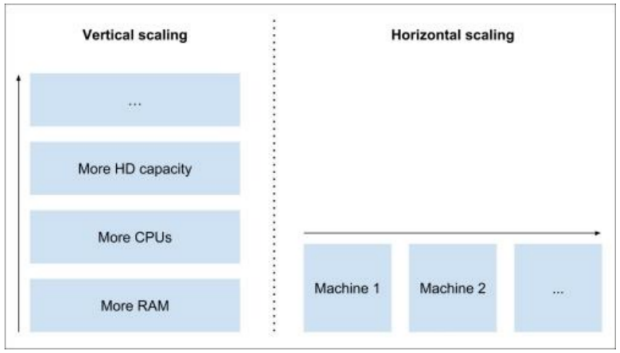
- One service is used as the primary and the other services are called secondaries
- All of the set instances support read operations
- Only the primary instance is in charge of write operations
  - When a write operation occurs, the primary will inform the secondaries about the changes
- Replica set is its automatic failover



Scaling

Scaling is a common problem with a growing web application

- Vertical scaling
  - ✓ Easy
  - ✗ More Expensive
  - ✗ Limited
  - ✗ Cloud hosting limit
- Horizontal scaling
  - ✓ Cheaper
  - ✓ Better scalability
  - ✗ More Complicated

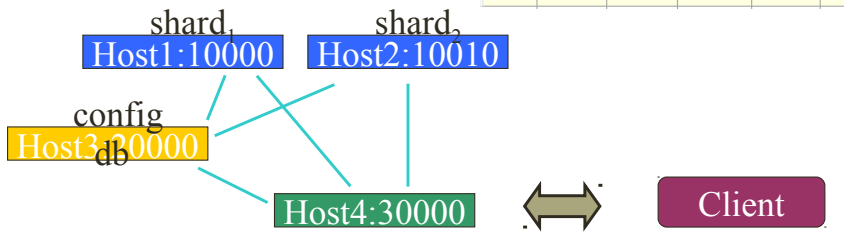


Sharding

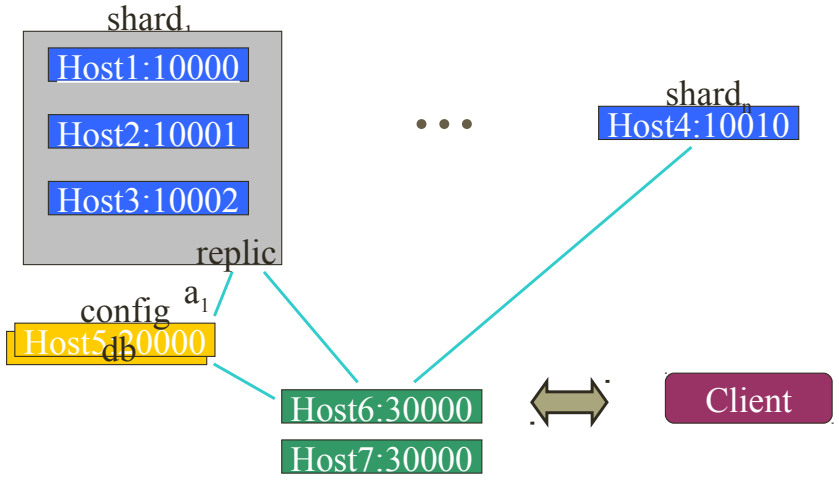
Sharding is the process of splitting the data between different machines, or shards

- Partition your data
- Scale write throughput
- Increase capacity
- Auto-balancing

id	company	customer	article	currency	price
4250250	020	073000	5994537812	00	142,50
4250251	020	073000	5994537852	00	141,12
4250252	020	073000	5994537854	00	105,99
4250253	020	073000	5994537856	00	109,52
4250254	020	073000	5994537862	00	131,49
4250255	020	073000	5994567308	00	29,86
4250256	020	073000	5994567422	00	57,13
4250257	020	073000	5994567428	00	68,59
4250258	020	073000	5994605089	00	51,09
4250259	020	073000	5994607975	00	93,93
4250260	020	073000	5994701005	00	74,22



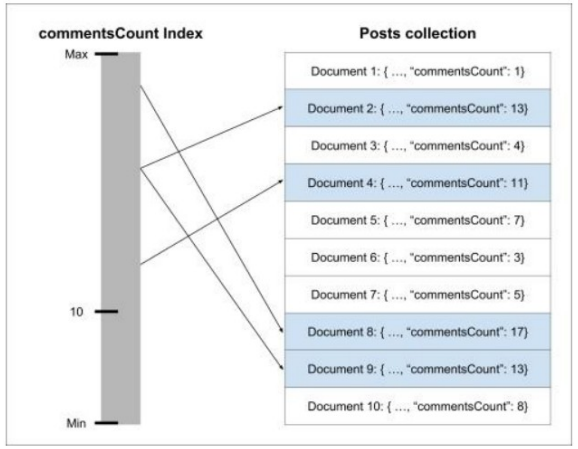
Mixed



MongoDB indexing

Indexes are a unique data structure that enables the database engine to efficiently resolve queries

```
{
"_id":
ObjectId("52d02240e4b01d67d71
ad577"),
"title": "First Blog Post",
"comments": [
],
"commentsCount": 12
}
```

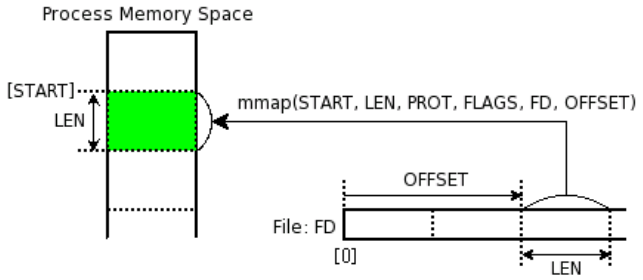



```
db.posts.find({ commentsCount: { $gt: 10 } });
```

Memory Mapped Files

- ***"A memory-mapped file is a segment of virtual memory which has been assigned a direct byte-for-byte correlation with some portion of a file or file-like resource."***

mmap()



Document Store

RDBMS		MongoDB
Database	⇨	Database
Table, View	⇨	Collection
Row	⇨	Document (JSON, BSON)
Column	⇨	Field
Index	⇨	Index
Join	⇨	Embedded Document
Foreign Key		Reference
Partition		Shard

```
> db.user.findOne({age:39})
{
        "_id" :
ObjectId("5114e0bd42…"),
        "first" : "John",
        "last" : "Doe",
        "age" : 39,
        "interests" : [
                "Reading",
                "Mountain Biking ]
        "favorites": {
                "color": "Blue",
                "sport": "Soccer"}
}
```


CRUD

- Create
  - db.collection.insert(<document>)
  - db.collection.save(<document>)
  - db.collection.update(<query>, <update>, { upsert: true })
- Read
  - db.collection.find(<query>, <projection>)
  - db.collection.findOne(<query>, <projection>)
- Update
  - db.collection.update(<query>, <update>, <options>)
- Delete
  - db.collection.remove(<query>, <justOne>)

CRUD Example

```
> db.user.insert({
        first: "John",
        last : "Doe",
        age: 39
})
```

```
> db.user.find ()
{
        "_id" : ObjectId("51…"),
        "first" : "John",
        "last" : "Doe",
        "age" : 39
}
```

```
> db.user.update(
        {"_id" : ObjectId("51…")},
        {
            $set: {
                age: 40,
                salary: 7000}
        }
)
```

```
> db.user.remove({
        "first": "John"
})
```

MongoDB shell

- https://docs.mongodb.org/getting-started/shell/

Restaurant sample document

```
<{
 "address": {
    "building": "1007",
    "coord": [ -73.856077, 40.848447 ],
    "street": "Morris Park Ave",
    "zipcode": "10462"
 },
 "borough": "Bronx",
 "cuisine": "Bakery",
 "grades": [
    { "date": { "$date": 1393804800000 }, "grade": "A", "score": 2 },
    { "date": { "$date": 1378857600000 }, "grade": "A", "score": 6 },
    { "date": { "$date": 1358985600000 }, "grade": "A", "score": 10 },
    { "date": { "$date": 1322006400000 }, "grade": "A", "score": 9 }
 ],
 "name": "Morris Park Bake Shop",
 "restaurant_id": "30075445"
}
```

```
mongoimport --authenticationDatabase admin --username root --password bitnami --db test --collection restaurants --drop --file primer-dataset.json
```


MongoDB shell

Insert

```
mongo admin --username root --password bitnami
use test
db.createUser({ user: "Alice",pwd: "Moon1234", roles: [ "readWrite", "dbAdmin" ]} )
mongo  --username Alice --password Moon1234
db.restaurants.insert(
  {
    "address" : {
      "street" : "2 Avenue",
      "zipcode" : "10075",
      "building" : "1480",
      "coord" : [ -73.9557413, 40.7720266 ]
    },
    "borough" : "Manhattan",
    "cuisine" : "Italian",
    "grades" : [
      {
        "date" : ISODate("2014-10-01T00:00:00Z"),
        "grade" : "A",
        "score" : 11
      },
      {
        "date" : ISODate("2014-01-16T00:00:00Z"),
        "grade" : "B",
        "score" : 17
      }
    ],
    "name" : "Vella",
    "restaurant_id" : "41704620"
  }
)
```


MongoDB shell

Queries

```
db.restaurants.find()

                    //top level field
db.restaurants.find( { "borough": "Manhattan" } )

                    // Field in an Embedded Document
db.restaurants.find( { "address.zipcode": "10075" } )

                    // Field in an Array
db.restaurants.find( { "grades.grade": "B" } )

                    //greater condition
db.restaurants.find( { "grades.score": { $gt: 30 } } )

                    // logical and
db.restaurants.find( { "cuisine": "Italian", "address.zipcode": "10075" } )

                    //logical or
db.restaurants.find(
   { $or: [ { "cuisine": "Italian" }, { "address.zipcode": "10075" } ] }

)
                    // sort query results
db.restaurants.find().sort( { "borough": 1, "address.zipcode": 1 } )
```


MongoDB shell

Updates

```
                    //top level field – single document
db.restaurants.update(
   { "name" : "Juni" },
   {
    $set: { "cuisine": "American (New)" },
    $currentDate: { "lastModified": true }
   }
)
                    // Embedded Field¶
db.restaurants.update(
  { "restaurant_id" : "41156888" },
  { $set: { "address.street": "East 31st Street" } }
)

                    // multiple documents
db.restaurants.update(
  { "address.zipcode": "10016", cuisine: "Other" },
  {
   $set: { cuisine: "Category To Be Determined" },
   $currentDate: { "lastModified": true }
  },
  { multi: true}
)
```

MongoDB shell

Removes

```
                    //Remove All Documents That Match a Condition
db.restaurants.remove( { "borough": "Manhattan" } )

                    // Use the justOne
db.restaurants.remove( { "borough": "Queens" }, { justOne: true } )

                    //Drop a Collection
db.restaurants.drop()
```


MongoDB shell

Data Aggregation

```
//groups the documents in the restaurants collection by the borough field and
uses the $sum accumulator to count the documents for each group.


db.restaurants.aggregate(
  [
    { $group: { "_id": "$borough", "count": { $sum: 1 } } }
  ]
);
```

Result

```
{ "_id" : "Staten Island", "count" : 969 }
{ "_id" : "Brooklyn", "count" : 6086 }
{ "_id" : "Manhattan", "count" : 10259 }
{ "_id" : "Queens", "count" : 5656 }
{ "_id" : "Bronx", "count" : 2338 }
{ "_id" : "Missing", "count" : 51 }

/
```


MongoDB shell

- https://docs.mongodb.org/getting-started/shell/

MongoDB NodeJS

Declare MongoClient variable and other variables

```
var MongoClient = require('mongodb').MongoClient;
var assert = require('assert');
```

Connect using the MongoClient to a running mongod instance by specifying the MongoDB uri.

```
var url = 'mongodb://localhost:27017/test';
MongoClient.connect(url, function(err, db) {
  assert.equal(null, err);
  console.log("Connected correctly to server.");
  db.close();
});
```

MongoDB NodeJS

- https://docs.mongodb.org/getting-started/node/
- Tarea: Definir un documento propio y ejercitar operaciones CRUD
  - Enviar archivo a alfredo.teyseyre@gmail.com

Mongoose

Mongoose provides a straight-forward, schema-based solution to model your application data

- Features
  - Schemas to model your entities
  - Built-in type casting
  - Predefined validation along with custom validations
  - Virtual attributes
  - Query building
  - Business logic hooks

```
var var mongoose = require('mongoose');
mongoose.connect('mongodb://localhost/test');

var Cat = mongoose.model('Cat', { name: String });

var kitty = new Cat({ name: 'Zildjian' });
kitty.save(function (err) {
  if (err) {
    console.log(err);
  } else {
    console.log('meow');
  }
});
```


Mongoose

Defining your schema

```
var mongoose = require('mongoose');
var Schema = mongoose.Schema;

var blogSchema = new Schema({
  title:  String,
  author: String,
  body:   String,
  comments: [{ body: String, date: Date }],
  date: { type: Date, default: Date.now },
  hidden: Boolean,
  meta: {
    votes: Number,
    favs:  Number
  }
});
```

```
var Blog = mongoose.model('Blog', blogSchema);
```


Mongoose

Instance methods

```
// define a schema
var animalSchema = new Schema({ name: String, type: String });

// assign a function to the "methods" object of our animalSchema
animalSchema.methods.findSimilarTypes = function (cb) {
  return this.model('Animal').find({ type: this.type }, cb);
}
```

Now all of our animal instances have a findSimilarTypes method available to it.

```
var Animal = mongoose.model('Animal', animalSchema);
var dog = new Animal({ type: 'dog' });

dog.findSimilarTypes(function (err, dogs) {
  console.log(dogs); // woof
});
```

Mongoose

Statics

```
// assign a function to the "statics" object of our animalSchema
animalSchema.statics.findByName = function (name, cb) {
  return this.find({ name: new RegExp(name, 'i') }, cb);
}
```

```
var Animal = mongoose.model('Animal', animalSchema);
Animal.findByName('fido', function (err, animals) {
  console.log(animals);
});
```


Mongoose

Indexes

```
var animalSchema = new Schema({
  name: String,
  type: String,
  tags: { type: [String], index: true } // field level
});

animalSchema.index({ name: 1, type: -1 }); // schema level
```


Mongoose

Virtuals

```
// define a schema
var personSchema = new Schema({
  name: {
    first: String,
    last: String
  }
});

// compile our model
var Person = mongoose.model('Person', personSchema);

// create a document
var bad = new Person({
    name: { first: 'Walter', last: 'White' }
});
```

```
console.log(bad.name.first + ' ' + bad.name.last); // Walter White
```

```
personSchema.virtual('name.full').get(function () {
  return this.name.first + ' ' + this.name.last;
});
console.log('%s is insane', bad.name.full); // Walter White is insane
```


Mongoose

Queries

```
var Person = mongoose.model('Person', yourSchema);

// find each person with a last name matching 'Ghost', selecting the `name`
and `occupation` fields
Person.findOne({ 'name.last': 'Ghost' }, 'name occupation', function (err,
person) {
  if (err) return handleError(err);
  console.log('%s %s is a %s.', person.name.first, person.name.last,
person.occupation) // Space Ghost is a talk show host.
})
```

```
// named john and at least 18
MyModel.find({ name: 'john', age: { $gte: 18 }});
```

Mongoose

- http://mongoosejs.com/docs/guide.html
- Tarea: Modelar con Mongoose el mismo ejemplo planteado para NodeJS
  - Enviar archivo a alfredo.teyseyre@gmail.com