Modern Web Development

MongoDB, NoSQL, mongoose

What is MongoDB

- It is document-oriented, NoSQL database.
- The concepts of tables and row are replaced with collections and documents.
- It has support for indexing, replication, sharding and data aggregation/processing.
- The documents do not have a predefined structure.
- It is best suited for Big Data, real-time web apps and Analytics.

What is NoSQL

- “‘non-sql” , no structured query language, no-relational database.
- No schema.
- Simplicity of design.
- Simpler horizontal scaling.
- More flexibility.
- Depending on the specific problem some operations are faster.

NoSQL vs SQL

- Typically in an SQL database the data is modeled in tables and rows.
- Enitities have relationships between them, a row in table A is linked with a row in table B.
- Using joins to retrieve the data needed for some operation.

- Data is modeled with collection and documents.
- Tipically in a NoSQL database all the data needed is available in the same collection and can be retrieved with a single query.

Collections and documents

- A collections is a store for documents.
- A document is an entity.
- Document ~= Row (entry)
- Collection ~= Table

Embedded relationships

The document contains all necessary information:

```
{
  title: "50 Tips and Tricks for MongoDB Developer",
  author: "Kristina Chodorow",
  published_date: ISODate("2011-05-06"),
  pages: 68,
  language: "English",
  comments: [{} ...]
  publisher: {
          name: "O'Reilly Media",
          founded: 1980,
          location: "CA"
        }
}
```

Reference Relationship

Separate documents data modeling:

```
{
  title: "50 Tips and Tricks for MongoDB Developer",
  author: "Kristina Chodorow",
  published_date: ISODate("2011-05-06"),
  pages: 68,
  language: "English",
  comments: [ "43242", "43245"]
  publisher: "43242"

}
```

Aggregation framework

- Mechanism for data processing.
- Processes data from multiple documents and return the computed result.
- Modeled on the concept of data processing pipelines.
- The pipeline consists of multiple stages.
- Stage ex: $match, $group, $project
- Only works on one collection by design.

Map Reduce

- Mechanism for data processing.
- Invented at Google by Jeff Dean.
- Works on multiple collections.
- 3 stages: query, map, reduce.
- Map and reduce are actual JS functions that run inside mongo.
- Is slower than aggregation pipeline but more flexible.

Indexing

Indexing greatly improves read operations at the expense of write operations.

Indexes are special data structures that store a small portion of the collection’s data in an easy to traverse form.

The indexes are used to limit the documents that are inspected.

Trade-offs have to be made in order to choose the better technical solution.

Replication

- A replica set is a group of mongod processes that maintain the same data set.
- Provides redundancy.
- High availability.

Sharding

- Sharding is a method for distributing data across multiple machines.
- Horizontal scaling such that the costs of the systems that hold and process the information is kept low.
- Each machine handles a subset of data.

Mongoose

- Simplifies mongo operations.
- Model abstraction.
- Validation and casting is done via schema.
- Business logic boilerplate can be done via middleware.
- It ads SCHEMA, why in NoSQL????

Schema

- Mongoose schemas are flexible.
- Sometimes you need structure in the data models.
- Validation is possible via a Schema.
- Fields that are not defined in the schema are not saved
- Fields that are defined but not mandatory may or may not appear in the document.

Sources

MongoDb manual: https://docs.mongodb.com/manual/

Mongoose documentation: http://mongoosejs.com/docs/

Wikipedia: https://en.wikipedia.org/

Further reading

https://www.tutorialspoint.com/mongodb/index.htm

https://developer.mozilla.org/en-US/docs/Learn/Server-side/Express_Nodejs/mongoose