MongoDB

# Introduction to MongoDB

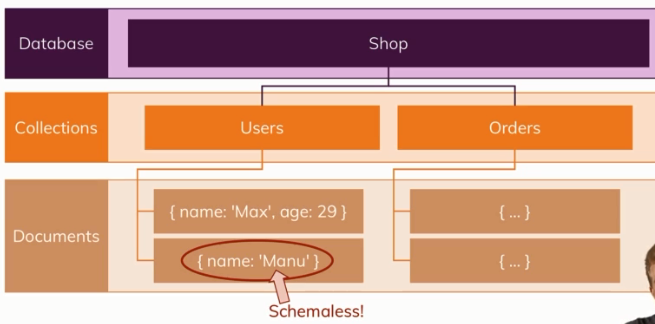
## Introduction

* MongoDB is an exciting technology, an exciting database solution and even more than just a database solution.
* You can use it in basically any project you're building. If that's a website, a mobile app, a desktop app or if you're working as a data scientist, you're probably working with some kind of data, you need to store data, you need to fetch data and you need to do all of that **fast and efficiently** without worrying too much about database configuration, setup and data structures and mongodb is an awesome solution for that.

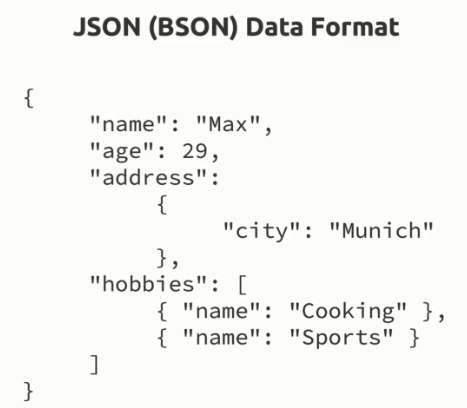
## What is MongoDB

* MongoDB is a database solution.
* The name mongodb is stemming from the word **humongous** because this database is built to store lots and lots of data and not just from a data size perspective but also in a sense that you can store lots of data and you can then work with it efficiently which of course also is super important.

## How MongoDB is different than other Databases / How it works



* mongodb is most importantly a database server that allows you to then run different databases on it.
* Relational DBs (MySQL, Oracle) have tables, in mongodb we have so-called collections.
* We can have multiple databases and multiple collections per database.
* Now inside of a collection, you have so-called documents which look like JSON objects and indeed this is basically how you store data in mongodb.
* The documents inside a collection are **schemaless**, in the sense that the second document may not look like the first one and this is a flexibility mongodb gives you.
* It is really all about flexibility, where SQL based databases are very strict about the data you have to store in there, mongodb is more flexible, you can store totally different data in one and the same collection and therefore, your database can grow with your application and your application needs.
* We can store nested data within a document (embedded document). This allows you to create complex relations between data and store them in one and the same document which makes working with it and fetching it super efficient.
* The data format used to store documents in mongodb is called **BSON (Binary JSON)**. Behind the scenes on the server, mongodb converts your JSON data to a **binary** version of it, which can basically be stored and can be queried more efficiently. But we as developers use the JSON data to interact with MongoDB.



* The whole theme of mongodb really is the **flexibility and the optimization for usability** and that is really what sets mongodb apart from other database solutions and which makes it so awesome and so efficient from a performance perspective too because you can query data in the format you need it instead of running complex restructurings on the server.

## The Key MongoDB Characteristics

* Mongodb is a so-called **NoSQL** solution because it's basically following an opposite concept or philosophy than all the SQL based databases do.

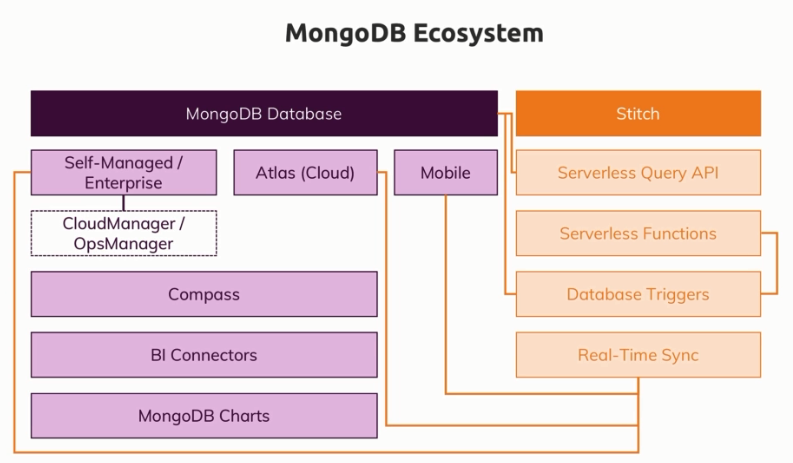
### No Schema

* Instead of normalizing data which means storing it, distribute it across multiple tables where every table has a clear schema and then using a lot of relations, mongodb goes for storing data together in a document and it also doesn't force a schema on you.
* So we got no schemas in mongodb. If we have multiple documents in one and the same collection, they can have different structures.
* This can of course lead to messy data but it's still your responsibility as a developer to work with clean data and to implement a solution that works but on the other hand, it gives you a **lot of flexibility** and flexibility is always good.
* You can use mongodb for apps that might still evolve, where the exact data requirements are just not set yet, you can get started and you can always add data with more information in a collection in the same collection at a later point of time.

### No / Less Relations

* With mongodb, you also work with less relations.
* With embedded documents, one core thing of mongodb indeed is that you have less less collections which you connect but instead that you store data together and this is where the efficiency is derived from.
* Since data is stored together, when your application is fetching data, it doesn't need to reach out to collection A, merge it with collection B, merge it with collection C, instead it goes to collection A then mongodb has a very efficient querying mechanism behind the scenes so that it can go through all the data very fast when looking for a specific document, so this will be super fast and then it finds that document and it's done, it doesn't need to do any merging most of the time.
* So this is really where the **speed, the performance and flexibility** comes from.
* And this hopefully speaks for itself that this can be useful when building applications and this is also why NoSQL solutions and amongst them most of all mongodb is super popular **for read write heavy applications**, applications with a lot of workload, applications that store a lot of data, let's say smart devices which send some sensor data every second, for such applications but also for building an online shop or a blog, mongodb is an amazing solution due to the performance and the flexibility it gives you.
* If there is any relational data, that has to be merged manually (kind of).

## Understanding MongoDB Ecosystem

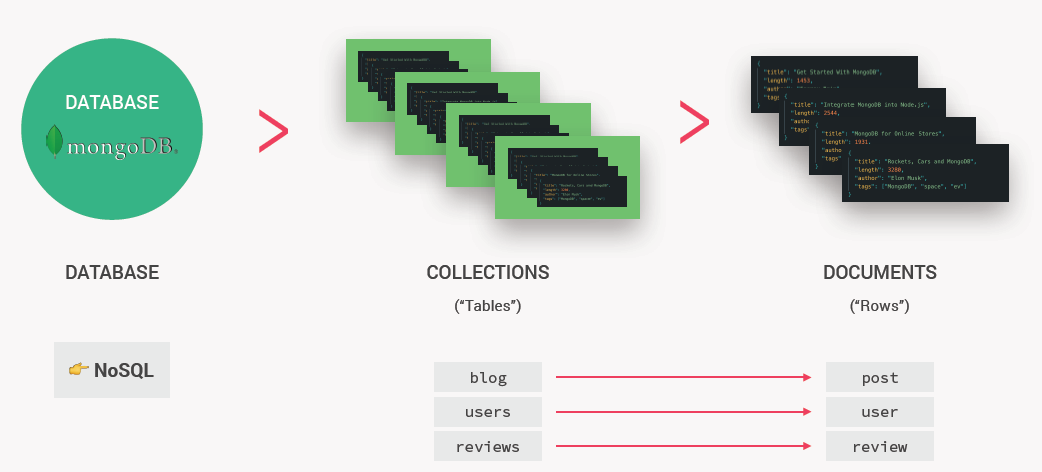


* The company behind the mongodb database solution is also called mongodb and actually, they have a couple of other products too.
* MongoDB Ecosystem has below systems/features.
* Obviously we got the **mongodb database** which is the core asset, the core feature.

There you have a **self-managed and an enterprise solution**.

* Then they have some tools (**CloudManager/OpsManager**) that are related to managing that database that is more of a system admin, database admin task.
* And they have a **cloud solution (Atlas),** which is an amazing solution for you as a developer, because all the managing a system administrator would have to do is done for you there and you can focus on the data and your logic there.
* They also have a **mobile solution** which basically means you can install mongodb on a mobile device to store data there directly and even work **without an internet connection**.
* They got a graphical user interface **(compass)** which allows you to connect to your database and look into it with a nice user interface.
* And we also get some other tools like BI connectors or Mongodb charts if you're more on the data science area. This is not really something where you will write a lot of code. These are more tools that allow you to connect different analytics tools if you need that.
* Now this is the entire database world and this is of course their core feature hence they have a lot of products there as you can see.
* Now they also got a relatively new offering which is called **stitch** and stitch is basically their **serverless backend solution**. So this is a bit decoupled from the database, still strongly connected but also it offers you more than just data storage. Stitch for example gives you a serverless query API. This is essentially a toolset, a tool which you can use to efficiently query your database directly from inside your client side apps, so from inside your react app or anything like that.
* You also have **serverless functions** which allows you to execute code in the cloud on demand. So now this is totally unrelated to the database, you can execute any javascript code in the cloud there and if you know something like AWS Lambda or Google cloud functions, this is your equivalent.
* We also have **database triggers** which is basically a service that allows you to listen to events in a database, like document was inserted and then execute a function in response to that.
* There is a feature called **real time sync** which basically is built to synchronize a database in a cloud with that mobile offline supporting database you might have on mobile devices.

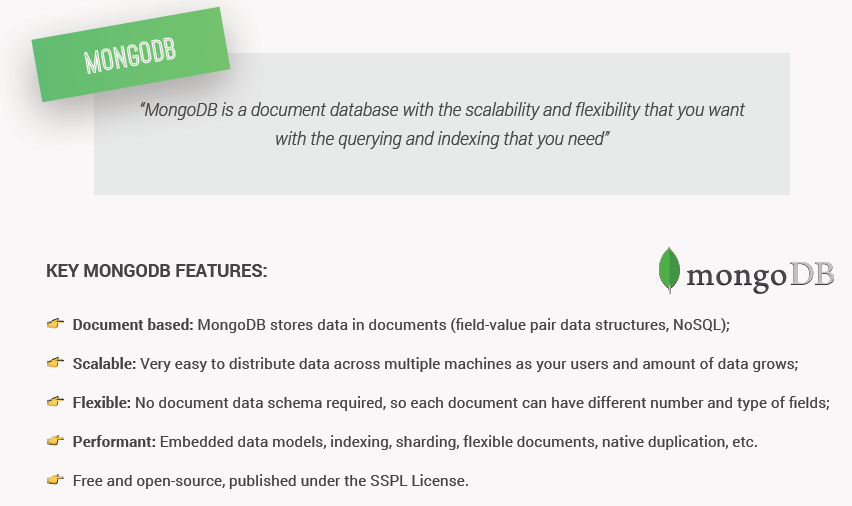
## Overview of MongoDB



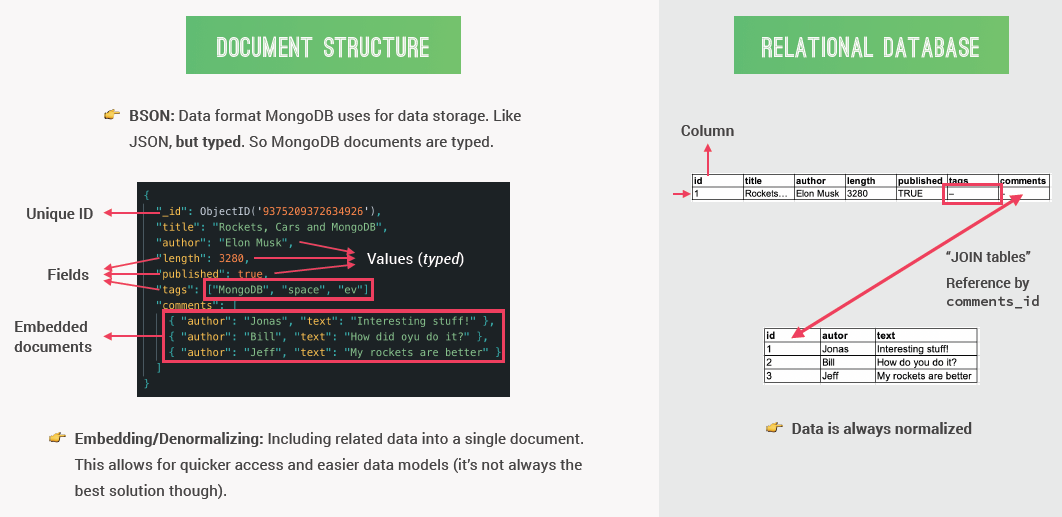
* In MongoDB (or just “mongo”), each database can contain one or more collections. A collection is like a table of data.
* Each collection can contain one or more data structures called documents. A document is like a row in a table. So each document contains the data about one single entity, for example, one blog post or one user or one review, etc.
* The collection is like the parent structure that contains all these entities. For example, a blog collection for all posts, users collection or a reviews collection.
* The document has a data format like JSON.

## Features of mongoDB

* MongoDB is a document database with the **scalability** and **flexibility** that you want, and with the **querying** and **indexing** that you need.
* MongoDB is a free and open-source database.
* MongoDB is a great database system to build many types of modern, scalable and flexible web applications.
* Mongo is probably the most used database with Node.js.



## Documents, BSON and Embedding



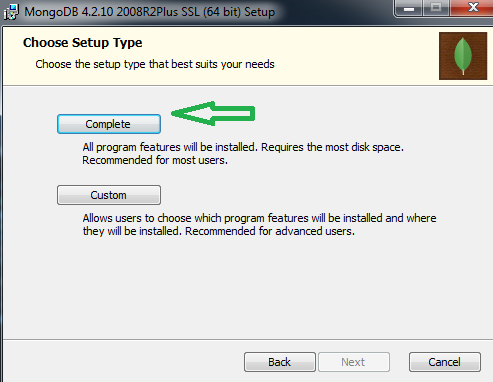
* MongoDB uses a data format similar to JSON for data storage called BSON. It looks basically the same as JSON, but it's **typed**, meaning that all values will have a data type such as string, Boolean, date, integer, double, object or more.
* These BSON documents will also have fields and data is stored in key value pairs. Field is like column in relational DB.
* We can have **multiple values to same field** in the form of arrays (see ‘tags’ field above), which is something complicated to do with relational DB.
* Another extremely important feature in MongoDB is the concept of **embedded documents**. The process of embedding, or de-normalizing as we can also call it, is basically to include so to embed some related data all into one single document. E.g. all the comments above.
* Embedded document makes a database more performant in some situations because this way, it can be easier to read all the data that we need all at once.
* The opposite of embedding or de-normalizing is normalizing, and that's how the data is always modeled in relational databases.
* The maximum size for each document is currently **16 MB**, but this might increase in the future.
* Each document contains a **unique ID**, which acts as a primary key of that document. It's automatically generated with the ObjectID data type each time there is a new document. So we don't have to worry about it.

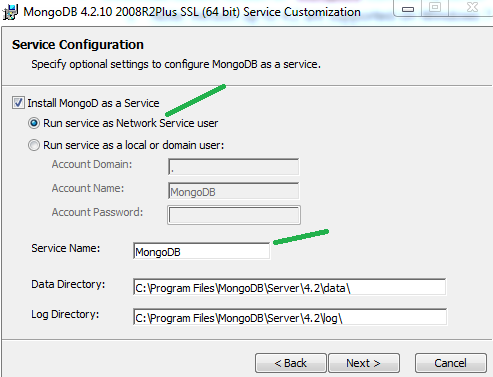
## Setting up MongoDB on Windows

Here we will install the MongoDB server and optionally MongoDB Compass (UI tool), then will create a directory to store the mongodb databases, then will see how to start and connect to the mondogb server and finally adding the mongodb in System Path Variable so that it can be run from anywhere.

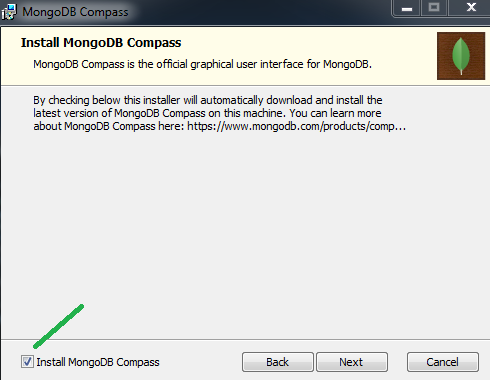
### Installing MongoDB

* Go to <https://www.mongodb.com/>
* Download current version of “MongoDB Community Server” installable (.msi) for Windows.
  + Note: Version up to 4.2 are supported on Windows 7





* + You can either select ‘Install MongoDB Compass’ so that the Compass tool will also be installed. Or you can deselect this option and separately install MongoDB Compass from the mongodb website.



### Create folder to store the MongoDB database

* Now once the MongoDB is installed, we need to create a directory in which MongoDB will store our data.
  + Go to the disk where MongoDB is installed.
  + If let’s say it is C: drive, then create folder ‘data’ under C drive and another folder ‘db’ under the ‘data’ folder like this –   
    C:\data  
    C:\data\db – This is the place where MongoDB will store our databases.
  + By **default**, MongoDB expects the data and data\db folders directly under root (For windows, it should be directly under C: drive where MongoDB is installed)
  + If you don’t have these default folders at this exact same path(C:\data\db), the mongodb server will fail to start.
  + If you want to provide a different path, e.g. say C:\Users\Sameer\data\db, then while starting the MongoDB Server, we need to specify this path as –

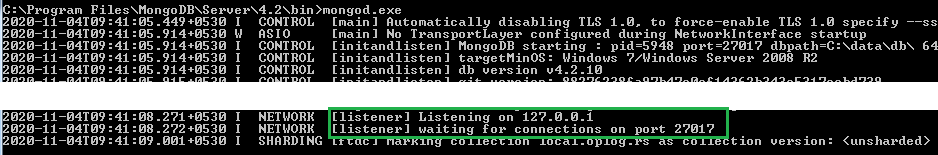
mongod --dbpath “C:\Users\Sameer\data\db”

* Few files at MongoDB installed location - C:\Program Files\MongoDB\Server\4.2\bin
  + mongod.exe – To start the MongoDB server.
  + mongo.exe – This is MongoDB Shell through which we can interact with our MongoDB databases and perform DB operations.
  + mongoimport.exe – used for importing data.

### Starting the MongoDB server and Connecting using MondoDB Shell

* Go to C:\Program Files\MongoDB\Server\4.2\bin and run mongod.exe to start the **server**.

C:\Program Files\MongoDB\Server\4.2\bin>**mongod**

* After you run this file, you will see an message in the end – “[listener] waiting for connections on port 27017”  
  
* MongoDB dy default runs on port 27017. If you want to change this, you need to start the server with ‘port’ flag like this –

>**mongod** --port 27018

* So we've basically started a server and now we need a shell to connect to the server to be able to manipulate our databases (Create, Delete, Query, etc.).
* Note:
  + On windows, you may not have to start the mongodb server (mongod) manually because on windows it is running as a “Service” so it is anyways running in the background. However if you get error while connecting to Shell (below), then you can manually start the server.
  + To stop the background MongoDB server which is running as a service, execute –

net stop MongoDB

To start the MongoDB server as a service in background, execute this –

net start MongoDB



* To run the **Shell**, we need to run the mongo.exe file

C:\Program Files\MongoDB\Server\4.2\bin>**mongo**

* After running this command, you can see that we've automatically connected to the default port (27017). And you will see the shell to interact with the MongoDB database.



* If you have used a different port to run MongoDB server, then you have to specify same port explicitly while running the Shell. E.g. >**mongo –-port 27018**
* To ‘double’ test the connection, we can run the db command and you should see output as ‘test’ which is the default database we get.

> db

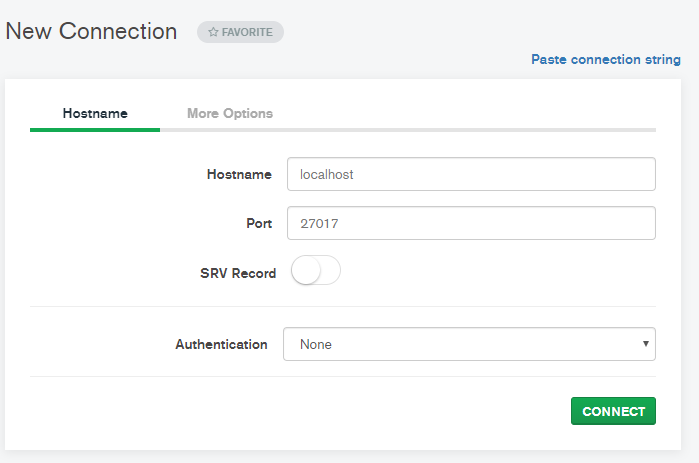
Test

### Set System Environment Variables

* In order to run the Mongo server from any directory, we need to set path for the mongodb in System Environment variables.
* Add path ‘C:\Program Files\MongoDB\Server\4.2\bin’ to the ‘Path’ System Environment Variables on your machine.

## MongoDB Compass

* Instead of using the terminal to work with MongoDB, we can also use a graphical user interface app called MongoDB Compass.
* For MongoDB Compass installation, refer – [Installing MongoDB](#_Installing_MongoDB)
* In order to create a new connection to our local database in Compass, make sure you have the Mongo server running in the background.
* Open MongoDB Compass -> New Connection -> click Connect. Because by default it has the local connection settings (Host: localhost and Port: 27017)



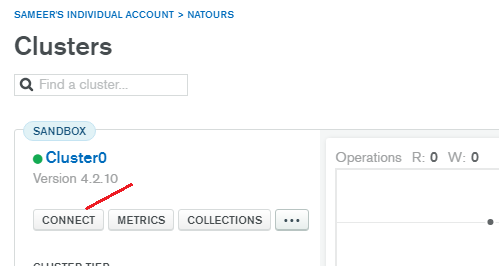
* Compass is just a graphical user interface for doing the exact same stuff that we can do with terminal/Shell.
* In Compass, we can really do all kinds of stuff like aggregations, define some schemas or analyze schemas, take look at indexes, performances and all that good stuff.

## Creating Remote Database Hosted on MongoDB Atlas

* We can either use local database (hosted on our machine) or use a remote database hosted on a service called Atlas, which is actually owned by the same company that involves MongoDB.
* Go to MongoDB website ( <https://www.mongodb.com> ) and select product MongoDB Atlas or you can use this link directly <https://www.mongodb.com/cloud/atlas>
* Atlas is a so-called database as a service provider which takes all the pain of managing and scaling databases away from us.
  + So that already is a huge advantage for us but it's also **extremely useful to always have our data basically in the Cloud**, because this way we can develop our application from everywhere.
  + And, even more importantly, we don't have to export data from the local database and then upload it to a hosted database, once we are ready to deploy our application.
* Create a cluster in the link given above.
* An Atlas cluster is basically like an instance of our database.
* Once a cluster is created, we now have a blank empty database, ready to connect to our own development computer.

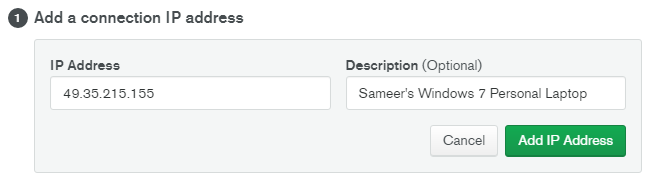
## Connecting to the Remote Hosted Database

* We can connect our remote hosted database with our Compass app and also with the Mongo shell.
* In the MongoDB Atlas app, click on ‘Connect’ under the cluster.

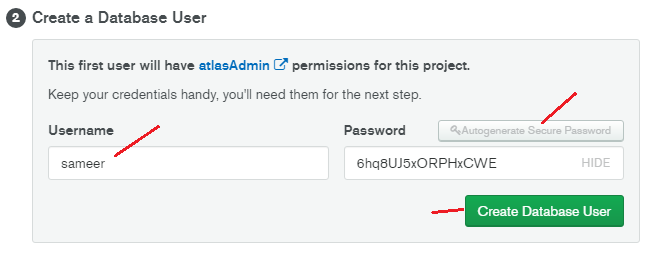


* First, we need to add our current IP address, so that our computer is actually able to connect to this cluster.





* Then, we need to create a database user.



Copy the password and add it into our Node application’s config.env file.

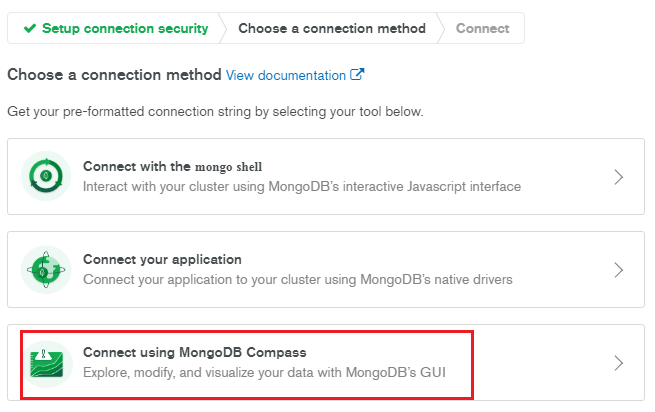
DATABASE\_PASSWORD=6hq8UJ5xORPHxCWE

Later, when we will then connect our application to the database, we will then use this environment variable to create that connection.

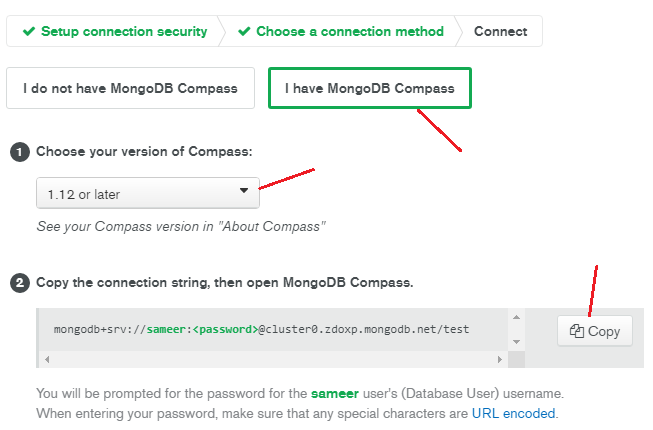
* Next, we need to choose a connection method.



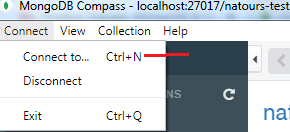
Select ‘Connect using MongoDB Compass’

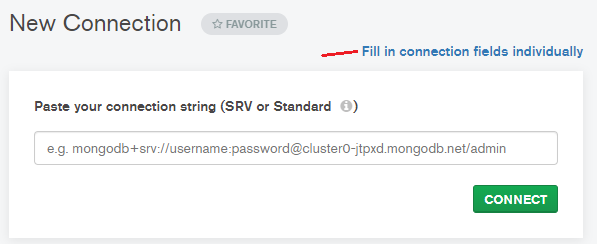


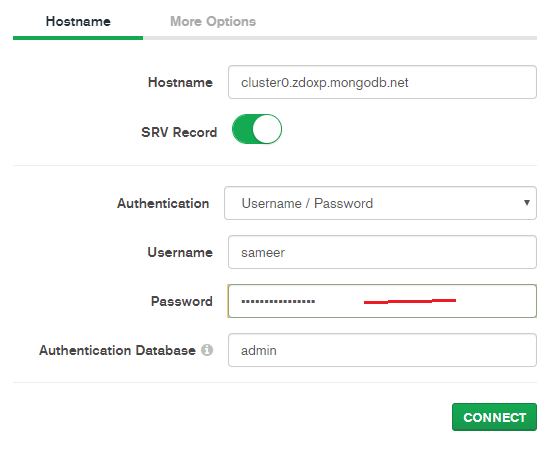
Select ‘I have MongoDB Compass’ (if not, first install it). Then select installed Compass version and then copy the connection string so that we can use our Remote database from our Compass app.



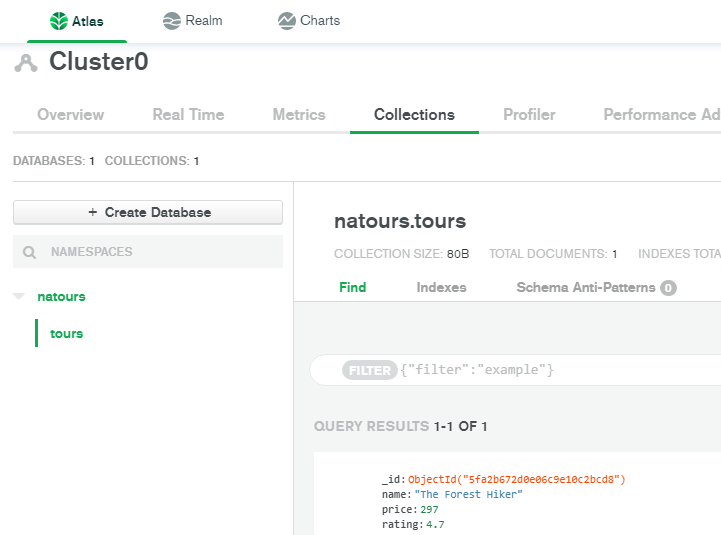
* Open MongoDB Compass and navigate to

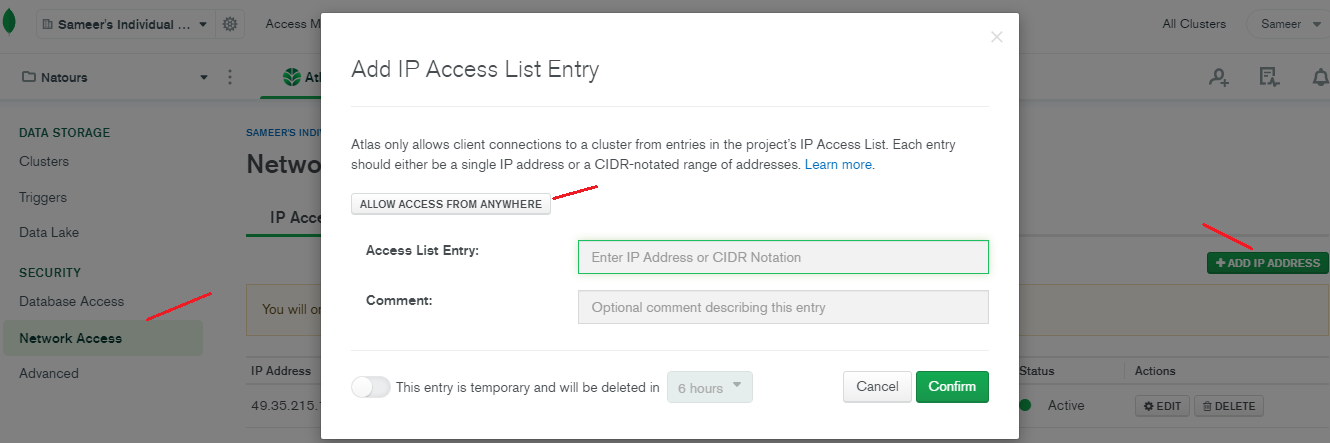


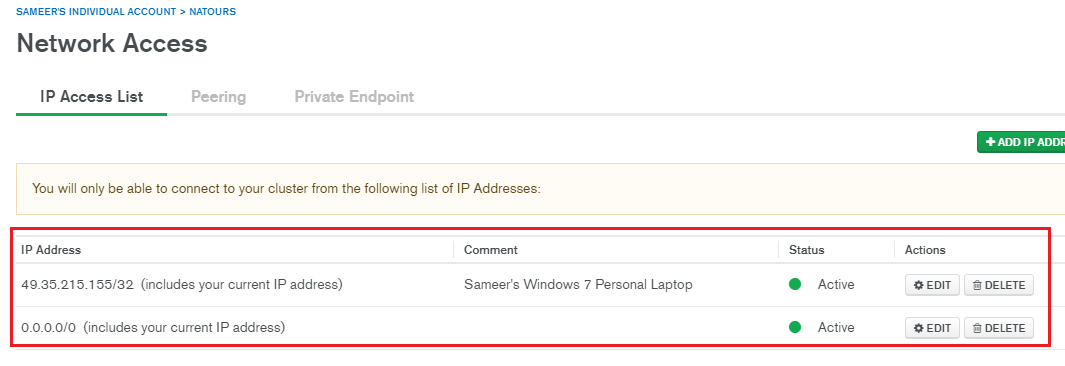
* Paste the copied connection string to below textbox and click on ‘Fill in connection fields individually’. There enter the database password which had created in previous step. And click on Connect.  
  



* Now we are connected to the Remote Database. From here on, we can the usual business of creating database, collections, documents, etc.
* Any DBs/collections/documents created here will be created on the Remote Host.
* To see these in Atlas, go to Atlas website and go under Clusters.

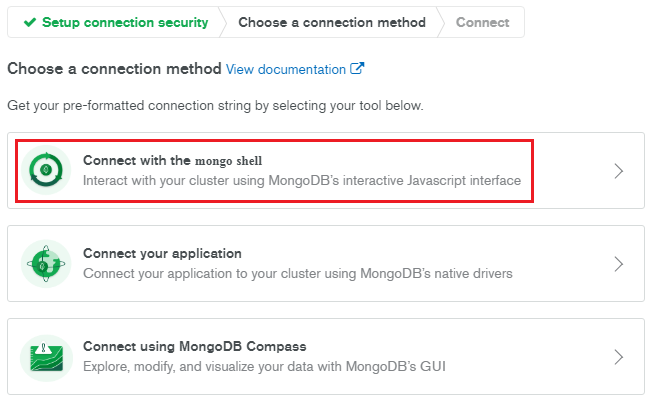


* Now another thing that we can and should do, is to **allow access** from everywhere to this cluster. In the beginning, we whitelisted our IP in order to grant access to our current computer to this cluster. But if you happen to switch computers during development, you might need to whitelist the IP of that computer as well, because otherwise you might not be able to connect. If we're not really dealing with sensitive data (test projects), we can simply whitelist every single IP in the world and allow access from everywhere. (of course, we will always still need our username and our password)
* For this, do this –   
  
* We can update this setting later on, if we want

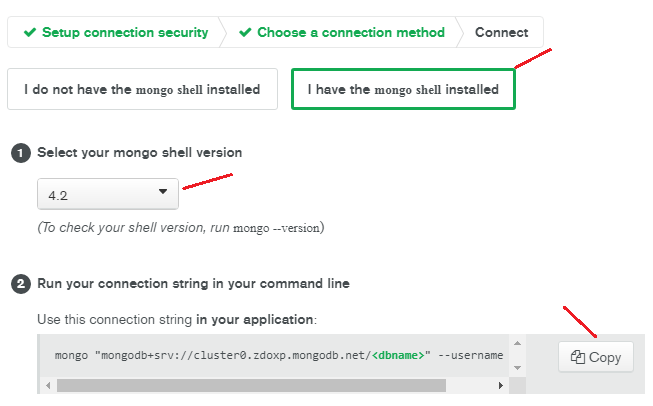


## Connecting Mongo Shell to the Remote Cluster

* Go to Cluster -> Connect -> Select ‘Connect using the mongo shell’



* Select appropriate option and copy the connection string



* Open New command prompt and paste the copied text. This text is the complete command to connect to our cluster.
* Now replace the text <dbname> with your actual database name .e.g natours. And hit enter.

E.g.

>mongo "mongodb+srv://cluster0.zdoxp.mongodb.net/**natours**" --username sameer

* Then it will ask for the password. Copy the database password which we created above and paste it and hit enter again.
* You should be connected now. To confirm it, enter command show dbs. You should see your DB.

MongoDB Enterprise atlas-kcxepu-shard-0:PRIMARY> show dbs

admin 0.000GB

local 1.252GB

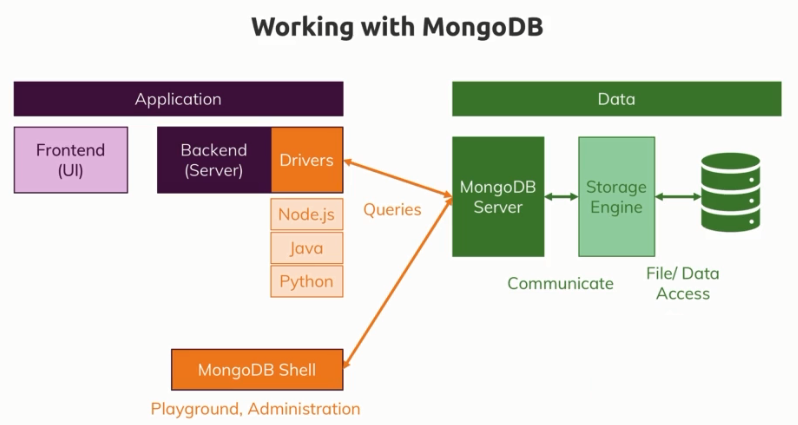
natours 0.000GB

* Now we are all set to go ahead.

## Shell vs Driver

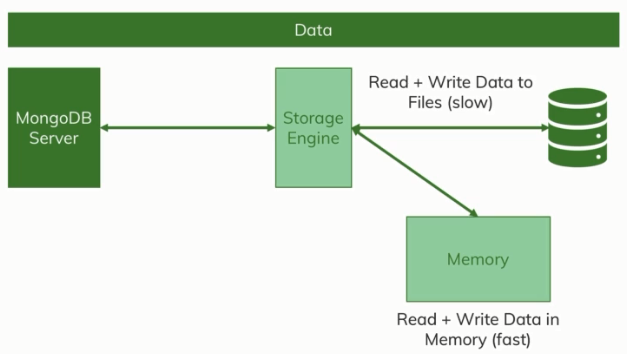
* Use Shell, to see all the commands and the different ways of using these commands because the shell is a great neutral ground for working with mongodb.
* Drivers are packages that you install/use for the different programming languages your app might be written in, and these drivers are then your bridges between your programming language and the mongodb server.
* And in these drivers, you use the same commands as you use in the shell, just adjusted to the syntax of the language you are working with of course.

## Working with MongoDB



* Now the mongodb server will actually not directly write the data into files but talk to a so-called **storage engine** which you could replace with your favorite storage engine but the default one called **Wired Tiger** is actually an awesome storage engine which allows you to efficiently work with your data, store it efficiently and so on.

### A closer look



* At that data layer – with the server, the storage engine and the file system, we actually have to differentiate between writing and reading from files which is a bit slower and writing and reading from memory which is faster.
* The storage engine actually does both, it loads a chunk of data into memory and manages that, such that the data you often use is in memory if possible.
* It also writes data in memory at first so that this is really fast but then of course it also goes ahead and stores data in the database files.
* In general, you need to be aware that you always talk to the mongodb server and behind the scenes, the server talks to the storage engine which manages your data and stores it in files in the end but also in memory in between so that you can work with the data in a very fast way. So this is how the mongodb server works behind the scenes.

# Understanding the Basics and CRUD Operations

## Understanding Databases, Collections and Documents

* In a MongoDB world, you have one or more **databases** on your database server and each database can hold one or more **collections**, a collection would be a table in a SQL database. Now in that collection, you have these **documents**, multiple documents per collection and the documents are really the data pieces you're storing in your database.
* Now important, when working with MongoDB, the databases, the collections and the documents are all automatically created for you, they are **created implicitly** when you start working with them, when you start storing data. This is a cool feature which makes getting started super simple.

## Default Databases

* When you run show dbs command, you see admin, config and local as databases.
* These default databases here simply exist to store configuration for this database server or for example the admin database will allow you to create users and roles and how people can use and interact with the database.

## Creating a Local Database

* We will create a local database using the **Mongo Shell** (mongo.exe) which is the terminal application.
* To clear the terminal – Ctrl + L
* To create a database, we use the "use" command inside the Mongo Shell, and then the name of the database that we want to create. This command here is also used to switch to an already existing database, but if we try to basically **switch to** a database that doesn't yet exist, it will then **create** a new one (actually it will create db if we add at least one document inside a collection).
* E.g. Here it created the database and switched to it.

> use natours-test

switched to db natours-test

* The data in mongodb is always a Document and the documents are part of a Collection. So we first need to create or use a collection in order to operate on documents.
* ‘db’ stands for current database.
* To create a document inside a collection, we use insertOne() or insertMany() operators like this

> db.tours.insertOne({ name: "The Forest Hiker", price: 297, rating: 4.7 })

{

"acknowledged" : true,

"insertedId" : ObjectId("5fa234442873acd05244dc3b")

}

* + In the insertOne() method we pass BSON string representing our document.
  + Use double quotes instead of single quotes.
  + Doubles quotes for key is optional.
  + Here ‘tours’ is a collection name. If it doesn’t exist, it will be created.
  + As you can see from the output of the command, mongodb automatically created unique ID property for our document.
* To query/see the documents inside a collection, use command db.<collection\_name>.find()

E.g.

> db.tours.find()

{ "\_id" : ObjectId("5fa234442873acd05244dc3b"), "name" : "The Forest Hiker", "price" : 297, "rating" : 4.7 }

* As you can see, mongodb works with the Javascript syntax and JSON, which is one of the main reasons why MongoDB is so popular for Node JS applications.
* To see all the databases in mongodb, use command show dbs.

> show dbs

admin 0.000GB

config 0.000GB

local 0.000GB

natours-test 0.000GB

* To see collections inside a database, use show collections.

> show collections

tours

* To exit/quit the mongo shell, use quit() operator.

## ObjectId Object

* ObjectId object is simply a special type of data provided by mongodb which is a unique ID that also allows you to sort your documents because it will also have some **timestamp** data in there. So this id also can be used for **sorting** because it's guaranteed that if you would add another document, that would be treated as a more recent document than the already inserted document.
* Each document needs a unique id in the field “\_id”. If not given explicitly, mongodb creates automatically in the form of ObjectId() and it also maintains **insertion order**.

## JSON vs BSON

* Mongodb does not use JSOB but BSON which stands for **Binary JSON** for storing data in your database.
* Now you don't really need to care too much about that because all you (developers) will write and see will be JSON code and this is also what we see when we retrieve data from mongodb. But behind the scenes, mongodb actually uses BSON data.
* This conversion (JSON to BSON) is done by the **mongodb drivers**. So this basically takes your JSON code and stores it in binary data and this is simply done because it is more efficient to store than JSON data, so it's faster and more efficient from a space or size perspective and additionally it supports additional types because for example that ObjectId is not valid normal JSON. And there are some other types too like for example there are different types of numbers, with decimals and very big numbers and these are stored in different ways behind the scenes.
* We as developers will always write data in JSON format.
* While creating documents, we can skip double quotation marks for keys provided there is no space in a key.

## \_id field in the document

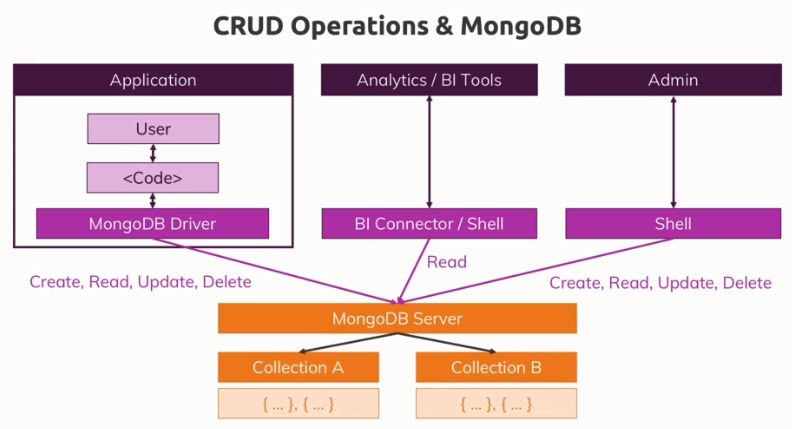
* When we insert documents, we get “\_id” as an auto-generated field.
* We don't have to use the auto-generated ID, we just have to ensure that we have a unique ID and if we can ensure this, we can also assign IDs on our own.
* Each document must have unique key.
* E.g.

> db.flightData.insertOne({ "departureAirport": "MUC", "arrivalAirport": "SFO", "aircraft": "Airbus A380", "distance": 12000, **"\_id": "MUC-SFO"** })

## Embedded Documents

* Nesting of documents inside one document.
* We can have up to 100 levels of nesting in mongodb (but we rarely need 3 or 4 levels of nesting).
* Another hard limit is the overall document’s size has to be below 60 MB.
* We can have arrays of embedded documents.
* Arrays can hold any kind of data.
* You can have documents in an array and you can also have arrays in documents.
* Embedded or nested documents and arrays gives you a lot of flexibility to structure your data.

## CRUD operations on Documents



### Creating Documents

* To create a document inside a collection, we use insertOne() or insertMany() operators.
* MongoDB documents are very flexible, i.e. they do not all have to have the same structure (within the same collection).
* We have seen insertOne() above, let’s use insertMany()
* To create multiple documents inside a collection, use insertMany()

E.g.

> db.tours.insertMany([{ name: "The Sea Explorer", price: 497, rating: 4.8}, { name: "The Snow Adventurer", price: 997, rating: 4.9, difficulty: "easy"}])

{

"acknowledged" : true,

"insertedIds" : [

ObjectId("5fa294d58251e593f8ea0992"),

ObjectId("5fa294d58251e593f8ea0993")

]

}

* + Here we have added 2 documents and the second document has an extra field called ‘difficulty’.
  + For the created documents, we got 2 auto generated IDs.

### Querying/Reading Documents

#### Theory

* find() method does not give us collection of data, but it gives us a so called **cursor** object.
* Find() does not give us an array of all the documents in a collection and that makes a lot of sense because that collection could be very big and if find() would immediately send us back all documents and you think about a collection with let's say 20 million documents, then this would take super long but most importantly, it would send a lot of data over the wire. So instead of that, it gives us back cursor object which is an object with a lot of metadata behind it that allows us to loop through the results and that is what that “it” command does in Mongo Shell, “it” basically use that cursor to fetch the next bunch of data.
* By default, using just find() in Mongo Shell returns us first 20 documents.
* Since find() returns a cursor, there are methods which can be used on it. E.g.
  + toArray() – to get all the data at once, we can use   
    > db.tours.find().toArray()
  + forEach() – to loop through all the documents. (Very efficient)   
    > db.tours.find().forEach((doc) => printjson(doc))
* find().pretty() works but findOne().pretty() fails because **pretty() is a method that simply exists on the cursor**. find() returns a cursor however findOne() does not give us a cursor because it only gives us one document anyways.
* For the other methods like insert, update and delete, cursors do not exist because these methods don't fetch data, they manipulate it.

#### Examples

* To query for all the documents in a certain collection, use find() operator without any arguments.

E.g.

> db.tours.find()

* To query for a certain document with some filtering criteria, just pass the filtering criteria as a BSON object (filter object) to the find() operator.

E.g.

> db.tours.find( { name: "The Forest Hiker" } )

> db.tours.find( { difficulty: "easy" } )

* We can also use some special query operators for more granular filtering. E.g. less than ($lt), greater than ($gt), less than or equal to ($lte), greater than or equal to ($gte).

E.g.

> db.tours.find( { price: {$lte: 500} } )

Here we want to get all documents whose price is less than or equal to 500.

* The special Mongo operators start with $ sign.
* We can use more than one search criteria (AND) to query the documents.

E.g.

> db.tours.find( { price: {$lt: 500}, rating: {$gte: 4.8} } )

Here we want to get all documents whose price is less than to 500 AND rating greater than or equal to 4.8.

* We can also use OR criteria to query the documents, i.e. to get documents satisfying either of the conditions. For this, we use $or operator, which takes an array of conditions.

E.g.

> db.tours.find({ $or: [{ price: {$lt: 500 }}, { rating: {$gte: 4.8 }}] })

Here we want to get documents having price less than 500 OR having rating greater than or equal to 4.8

* Besides our filter object, we can also pass in an object for **projection**. Projection means we simply want to select some of the fields in the output.
* For projections, we need to 0 or 1 for the document fields. **0 means exclusion, 1 means inclusion**. We cannot mix both 0 and 1 for different fields in the same query.
* E.g. If we want only name from the above query,

> db.tours.find({ $or: [ { price: { $lt: 500 } }, { rating: { $gte: 4.8 }} ] }, {**name: 1**})  
Here we want to get documents (with names only) having price less than 500 OR having rating greater than or equal to 4.8

* E.g. 2) If you want all fields except ‘name’

> db.tours.find({ $or: [ { price: { $lt: 500 } }, { rating: { $gte: 4.8 }} ] }, {**name: 0**})

* E.g. 2) If you want all fields except ‘name’ and ‘price’

> db.tours.find({ $or: [ { price: { $lt: 500 } }, { rating: { $gte: 4.8 }} ] }, {**name: 0, price: 0**})

* We cannot mix both 0 and 1 for different fields in the same query.

E.g.

> db.tours.find({ price: { $lt: 500 } }, {name: 0, price: 1})

Error: error: {

"ok" : 0,

"errmsg" : "Projection cannot have a mix of inclusion and exclusion.",

"code" : 2,

"codeName" : "BadValue"

}

* By default, “\_id” field is always returned even with projections. We have to explicitly mention in the projection if we don’t want it by setting “\_id”: 0

> db.tours.find({ price: { $lt: 500 } }, {name: 1, **\_id: 0**})

* **Querying embedded documents**. Note: We have to use double quotes.

> db.flightData.find({**"status.description"**: "on-time"}).pretty()

{

"\_id" : ObjectId("5ff856d1c1bd95b8458e8602"),

"departureAirport" : "MUC",

"arrivalAirport" : "SFO",

"aircraft" : "Airbus A380",

"status" : {

"description" : "on-time",

"lastUpdated" : "1 hr ago"

}

}

* **Querying** **arrays**.

> db.passengers.find({**hobbies: "sports"**}).pretty()

{

"\_id" : ObjectId("5ff9723ef82ed11a0231da5b"),

"name" : "Albert Twostone",

"age" : 68,

**"hobbies" : [**

**"sports",**

**"cooking"**

**]**

}

### Updating Documents

* To update one document, we use updateOne() operator. And for updating many, we have updateMany() operator.
* First we need to select the documents we want to update and second we need to pass in the data that should be updated.
* The first argument is basically a filter object, we basically need to query for the documents that we want to update. The filter object can be a complex object.
* E.g.

> db.tours.updateOne({ name: "The Snow Adventurer" }, { $set: { price: 597 } })

{ "acknowledged" : true, "matchedCount" : 1, "modifiedCount" : 1 }

Here the first argument is filter object ({ name: "The Snow Adventurer" }, and the second argument is to set price of 597 using the $set operator.

* In case of updateOne(), if the filter query returns more than one documents, then only first document will be updated. So in order to update multiple documents, use updateMany()
* E.g.

> db.tours.updateMany({ price: {$gt: 500}, rating: {$gte: 4.7} }, { $set: {premium: true} })

{ "acknowledged" : true, "matchedCount" : 1, "modifiedCount" : 1 }

Here we have set a new field ‘premium’ to true for the documents having price greater than 500 and rating greater than or equal to 4.7

* With updateMany() or updateOne(), we only update parts of the document. But we can also completely replace the content of the document using replaceOne() or replaceMany().
* replaceOne() and replaceMany() use the same arguments as updateOne() and updateMany().

### Deleting Documents

* deleteOne() will delete only the first document matching our query. And deleteMany() will delete all the documents matching our query.
* E.g.

> db.tours.deleteMany({ rating: {$lt: 4.8} })

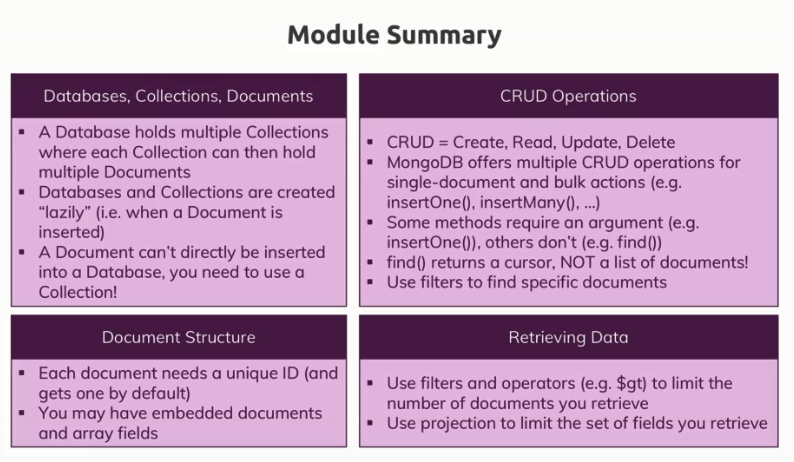
Here we delete all documents having rating less than 4.8

* To delete all the documents in a collection, pass empty filtering object like this

> db.tours.deleteMany({})

Because the empty object is basically a condition that all of the documents always match.

## Summary

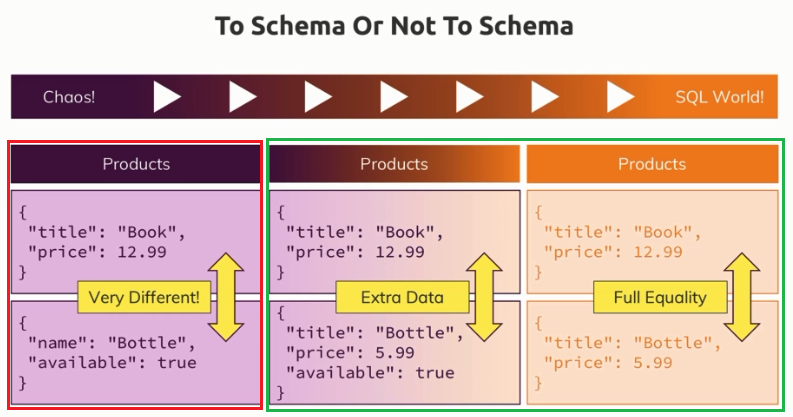


# Schemas and Relations: How to Structure Documents

## Why do we use Schemas

* MongoDB does not enforce any schemas. Documents don’t have to use the same Schema inside of one Collection. So you can mix different documents in a collection but in reality, you will probably have some kind of schema.
* Schema simply means the structure of a document. So how does it look like, which fields does it have, which types of values do these fields have, that is what is a schema.

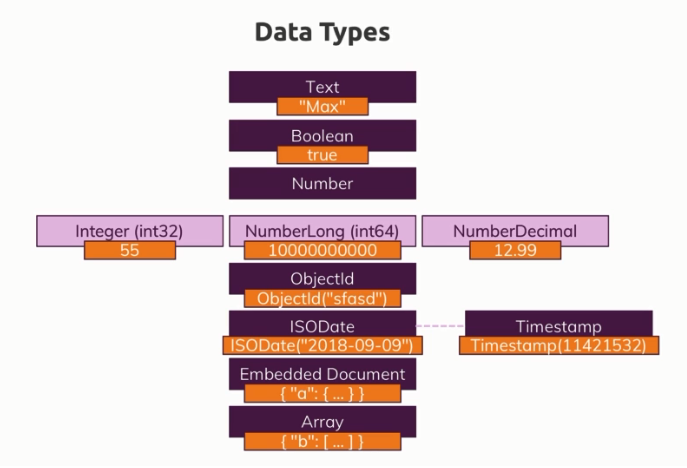
## Structuring Documents



* In reality, I guess either the approach on the right or most often the one in the middle is what you will see because there, you use the best of both worlds. You have some structure because your usecase or application probably needs it but you also take advantage of the flexibility mongodb gives you so that you can store extra information.

## Data Types

### Overview



#### Text

* Text always uses quotation marks around the value
* There is no hard limitation here, the only limitation you have regarding the maximum amount of characters in your text are the 16 MB overall size for the whole document.

#### Boolean

* true or false

#### Number

* Actually, we got a couple of different numbers in mongodb.
* We got **NumberInt** (int32). int32 integers are integers that are 32bits long and therefore if you try to store values that are longer than that, they will overflow that range and you will end up with a different number.
* For longer numbers, mongodb has a an int64 integer (**NumberLong**) which you can also assign to store values.
* In the shell by the way if you just enter a normal value, it is treated as a float value, that is because the normal shell is based on javascript and javascript doesn't differentiate between integers and floating point values, therefore everything will be stored as a 64bit float value in the shell, that is the default value.
* Be aware that mongodb is able to store smaller integers and bigger integers, the solution you choose here of course kind of defines how much space will be allocated and therefore how much space will be eaten up.
* And you also can store decimal numbers, you also got a special type, **NumberDecimal** and that is provided by mongodb **to store high precision floating point values** because normal floating point values also called doubles are basically rounded, so they're not super precise after their decimal place.
* For a lot of use cases, normal floats are enough though, but if you're doing scientific calculations or anything like that, you might need a very high precision. This NumberDecimal offers this, a very high decimal place precision where you got 34 decimal places, so after the comma which are not rounded but guaranteed to be correct and precise.
* It's simply a computing limitation that we got no 100% precision all the time but that rounding occurs.

#### ObjectId

* The ObjectId object is a special object automatically generated by mongodb to give you a unique ID which is not just a unique random string but also a string that contains a temporal component, so that if you create two elements after each other, two documents after each other, you are guaranteed to have the right order due to that ID because the older element will have an ID that comes prior to the other one, so there is this sorting built into the ObjectId because it respects a timestamp.

#### Date and Timestamp

* ISODate – to store Date. E.g. ISODate("2021-01-11T06:05:49.188Z")
* Timestamp – The timestamp is mostly used internally, you can create it automatically, mostly you let mongodb create that for you and that is guaranteed to be unique too. So even if you create two documents at the same time, they will not have exactly the same timestamp because it will basically take into account the current time and then also add an ordinal value, so that two documents at the same time still don't get the same time stamp but respect the order in which the insert command was issued.
* E.g. Timestamp(1610345149, 1) => here first number is current timestamp in milliseconds and second argument is ordinal value so that two documents at the same time still don't get the same timestamp.
* ObjectId is the kind of based on the timestamp and then just uses some algorithm to spit out a seemingly random string.

#### Embedded documents

* You can embed documents and these documents in turn can embed other documents.

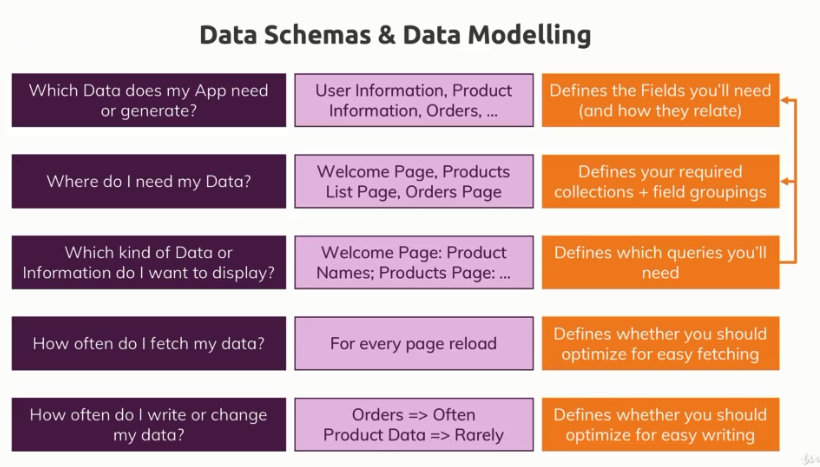
#### Arrays

* Arrays are simply lists of values,
* You can have a list of strings, a list of booleans, a list of numbers or a list of other embedded documents or even a list of lists, a list of arrays.

### Important Notes

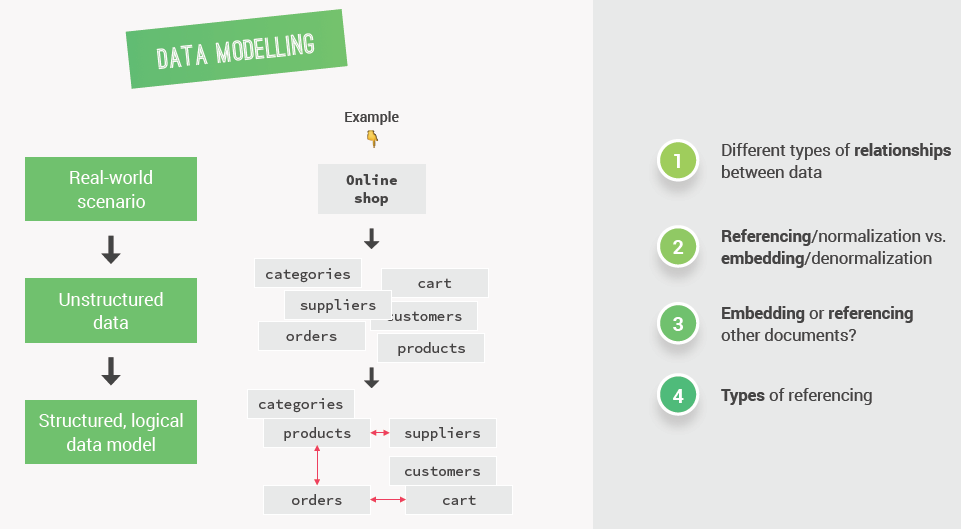
* For the data types, MongoDB supports  
  Refer – <https://docs.mongodb.com/manual/reference/bson-types/>
* MongoDB Limits and Thresholds   
  Refer – <https://docs.mongodb.com/manual/reference/limits/>
* Important data type limits are:
  + Normal integers (int32) can hold a maximum value of +-2,147,483,647
  + Long integers (int64) can hold a maximum value of +-9,223,372,036,854,775,807
  + Text can be as long as you want - the limit is the 16mb restriction for the overall document.
* It's also important to understand the difference between int32 (NumberInt), int64 (NumberLong) and a normal number as you can enter it in the shell. The same goes for a normal double and NumberDecimal.
  + NumberInt creates a int32 value => NumberInt(55)
  + NumberLong creates a int64 value => NumberLong(7489729384792)
  + If you just use a number (e.g. insertOne({a: 1}), this will get added as a normal double into the database. The reason for this is that the shell is based on JS which only knows float/ double values and doesn't differ between integers and floats.
  + NumberDecimal creates a high-precision double value => NumberDecimal("12.99") => This can be helpful for cases where you need (many) exact decimal places for calculations.

## Data Modelling Guidelines



* mongodb really embraces that idea of planning your data structure based on the way you'll retrieve your data, so that you don't have to do complex joins but that you can retrieve your data in the format or almost in the format you need it in your application.

## MongoDB Data Modelling

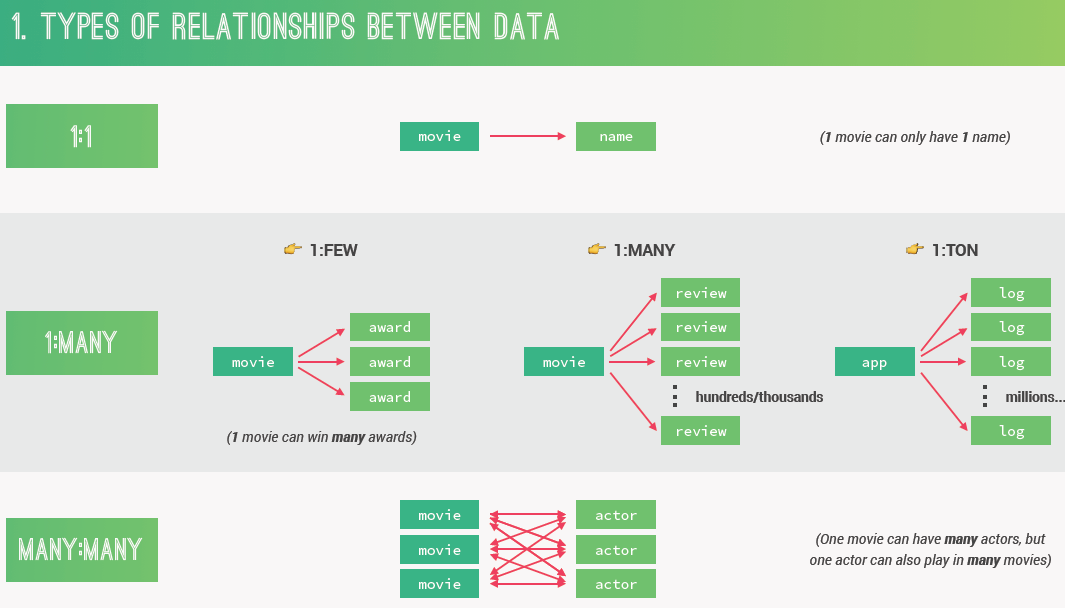


* One of the most important steps in building data intensive apps is to actually model all this data in MongoDB.
* **Data Modeling** is the process of taking unstructured data generated by a real world scenario and then structure it into a logical data model in a database.
* It’s the most demanding part of building an entire application. Because it really is not always straight-forward. And sometimes there are simply no right answers. So not just one unique correct way of structuring the data.
* There are 4 steps in the process of doing data modelling

1. Identify different types of relationships between data.
2. Understand the difference between referencing or normalization and embedding or denormalization.
3. Deciding whether we should embed documents or reference to other documents based on a couple of different factors.
4. Types of referencing.

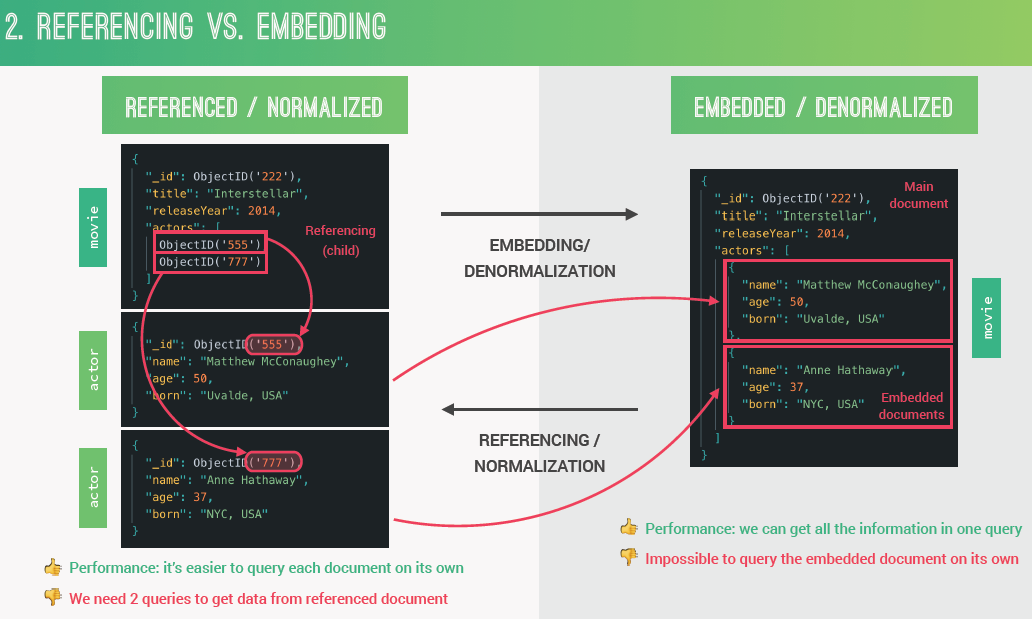
* The way we design data so the way we model our data can make or break our entire application.
* Let’s go through the steps one by one

### Types of Relationships between Data



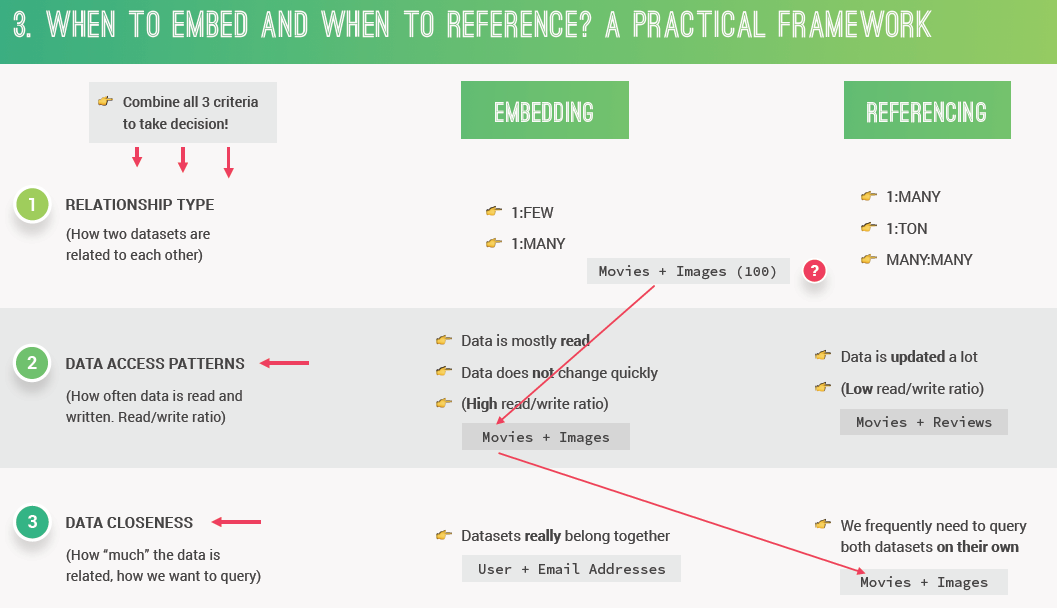
* The one to many relationships are the most important ones to know.
* The difference between ‘many’ and a ‘ton’ is of course a bit fuzzy but just think that if something can grow almost to infinity then it’s definitely a one to a ton relationship.
* In Relational databases, there is just one to many without quantifying how much that many actually is. However with NoSQL databases like MongoDB, this difference (one-to-few, one-to-many, one-to-ton) is extremely important because it’s one of the factors that we're going to use to decide if we should denormalize or normalize data.

### Referencing and Embedding two datasets



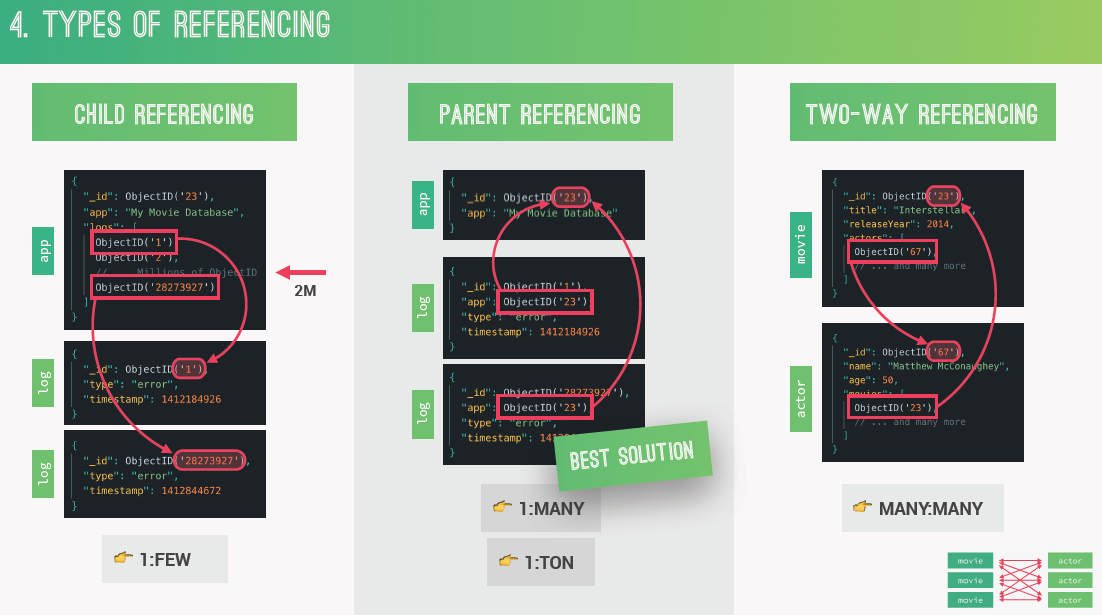
* Each time we have two related datasets we can either represent that related data in a **referenced or normalized** form or in an **embedded or denormalized** form.
* In the **referenced** form, we keep two related datasets and all the documents separated. So all the data is nicely separated which is exactly what normalized means.
* Child referencing is when the parent (e.g. Movie) references its children (e.g. Actors).
* In relational databases, all data is always represented in normalized form like this. But in a NoSQL database like MongoDB we can denormalize data into a denormalized form simply by **embedding** the related document right into the main document.
* If we choose to **denormalize** or to **embed** our data, we will have one main document containing all the main data as well as the related data.
* The result of embedding is that our application will need to fewer queries to the database. Because we can get all the data about movies and actors all at the same time which will of course increase our performance.
* The downside of embedding is of course that we can't really query the embedded data on its own. And so if that's a requirement for the application, you would have to choose a normalized design.
* Note: We could begin our thought process with denormlized data and then come to the conclusion that if its best to actually normalize the data.

### When to Embed and When to Reference



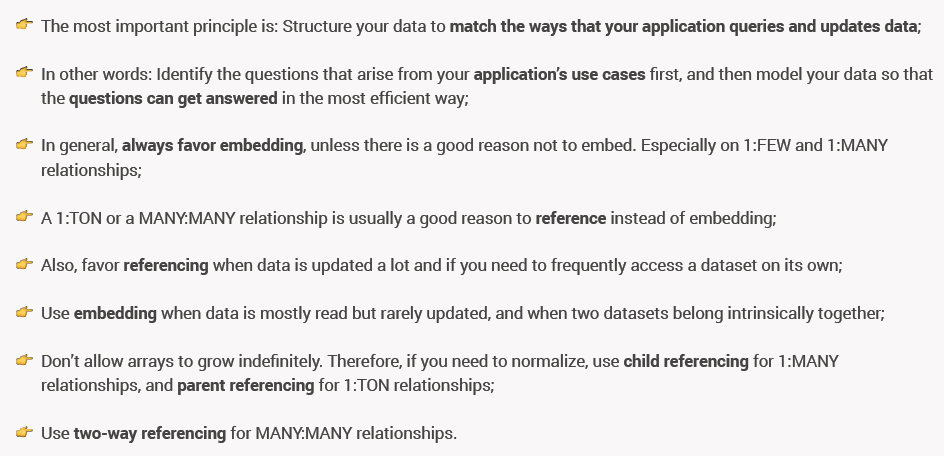
* When we have two related datasets; we have to decide if we're going to embed the datasets or if we're going to keep them separated and reference them from one dataset to the other.
* How do we actually decide if we should normalize or denormalize the data? Three criteria –
  + First look at the type of relationships that exists between datasets.
  + Second try to determine the data access pattern of the dataset that we want to either embed or reference. This just means to analyze how often data is read and written in that dataset.
  + Last, look at data closeness. Data closeness means how much the data is really related and how we want to query the data from the database.
* Now to actually take the decision; we need to combine all of these three criteria and not just use one of them in isolation.
* Usually when we have one to few relationship we will always embed the related dataset into the main dataset.
* Now in a one to many relationship; things are a bit fuzzy so it’s okay to either embed or reference. In that case we will have to decide according to the other two criteria.
* On a one to a ton or a many to many relationship we usually always reference the data. That's because if we actually did embed in this case we could quickly create way too large document. Even potentially surpassing the maximum of 16 MB.
* Data access patterns is just a fancy description for evaluating whether a certain dataset is mostly written to or mostly read from.
* So if the dataset that we're deciding about is **mostly read** and the data is not updated a lot then we should probably **embed** that dataset. The reason for this is that by embedding we only need one trip to the database per query.
* On the other hand, if our data is updated a lot then we should consider referencing or normalizing the data. That's because it’s more work for the database engine to update and embed a document than a simpler standalone document. And since our main goal is performance, we just normalize the dataset.
* Data closeness is just like a measure for how much the data is related.
* If the two datasets really intrinsically belong together then they should probably be embedded into one another.
* Now if we frequently need to query both of datasets on their own then that's a very good reason to normalize the data into two separate datasets even if they are closely related.
* There are not really completely right or completely wrong ways of modeling our data. There are no hard rules; these are just like guidelines that you can follow to find the probably most correct way of structuring your data.

### Types of Referencing



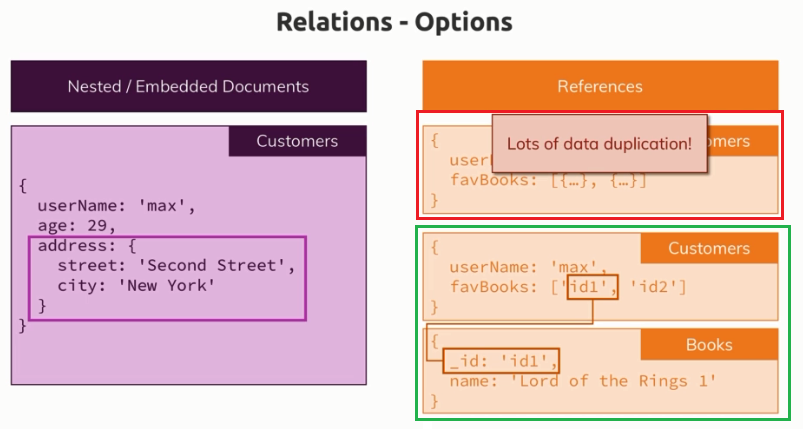
* If we have chosen to normalize/reference our datasets, then after that we still have to choose between three different types of referencing – Child referencing, parent referencing and two-way referencing.
* In child referencing, we basically keep references to the related child documents in a parent document, usually stored in an array.
* However, the problem with child referencing is that this array of IDs can become very large if there are lots of children. And this is an **anti-pattern** in MongoDB, so something that we should avoid at all costs. Also, child referencing makes it so that parents and children are very tightly coupled which is not always ideal. But that's exactly why we have parent referencing.
* In parent referencing, it actually works the other way around. So here in each child document we keep a reference to the parent element. The child always knows its parent but the parent actually knows nothing about the children.
* The **conclusion** of all this is that in general child referencing is best used for one to a few relationships, where we know before hand that the array of child documents won't grow that much. On the other hand, parent referencing is best used for one to many and one to a ton relationships.
* Note: The most important principals of MongoDB data modeling is that array should never be allowed to grow indefinitely. In order to never break that 16 MB limit.
* Two-way referencing is for many-to-many relationships.
  + Considering many to many relationship of Movie and Actor (one movie can have many actors, one actor can act in many movies), tow-way referencing works like this – in each movie we will keep references to all the actors that star in that movie. So a bit like in child referencing. However and at the same time in each actor we also keep references to all the movies that the actor played in. So movies and actors are connected in both directions. In therefore the name two-way referencing.
  + This makes it really easy to search for both movies and actors completely independently. While also making it easy to find the actors associated to each movie and the movies associated to each actor.
  + In case of **few-to-few** relationships, we can use embedding for intrinsic dataset.

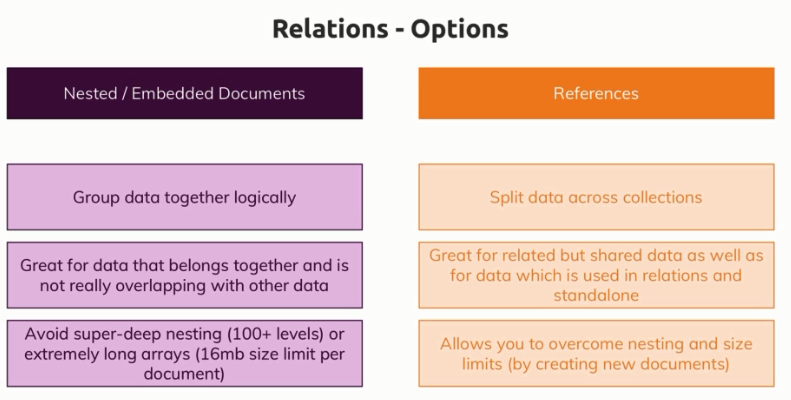
### Summary



## Understanding Relations

* How do you store related data?
* Do you use embedded documents? Or do you work with references?



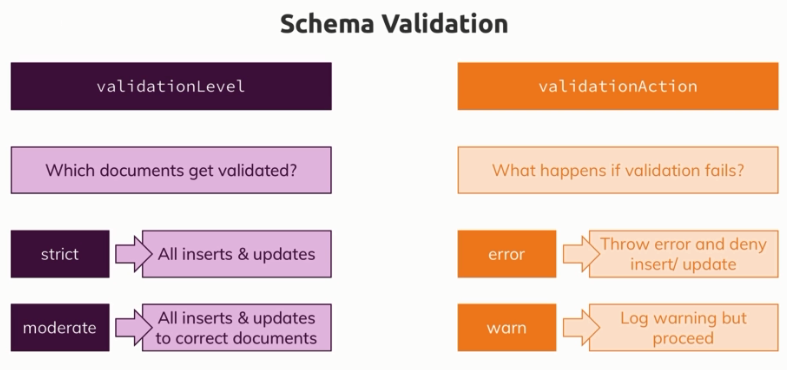


## Using “lookup()” for merging reference relations

* $lookup operator is essentially a helpful tool that allows you to fetch two related documents, merge together in one document in one step instead of having to do two steps.
* This mitigates some of the disadvantages of splitting your documents across collections because now you can at least merge them in one go.
* $lookup operator / function is very useful for merging data in one step and it allows you to get the best of both worlds, having it split up and still fetching it in one go. Now this still is not an excuse for always using references because obviously this costs more performance than having an embedded document. So if you can and if your application needs it, go for an embedded document. If you have to use references or if you want to use references, well then this lookup step in the aggregate method can at least help you get the data you need.

## Understanding Schema Validation

* Sometimes you need a strict schema because you know your application is going to fetch posts and it is going to access the title on each post and it does expect that each title is a string and for cases like this, schema validation can help you.
* If we add a validation schema, then the schema will validate or the mongodb will validate the incoming data based on the schema we defined and either it accepts it and then allows the write to the database or it rejects the incoming data, hence your database is not touched and is not changed and the user gets an error.



* JSON Schema (using $jsonSchema operator) is the recommended means of performing schema validation.
* Validation occurs during updates and inserts. When you add validation to a collection, existing documents do not undergo validation checks until modification.
* The validationLevel option determines which operations MongoDB applies the validation rules:
  + strict (the default), MongoDB applies validation rules to all inserts and updates.
  + moderate, MongoDB applies validation rules to inserts and to updates to existing documents that already fulfill the validation criteria. With the moderate level, updates to existing documents that do not fulfill the validation criteria are not checked for validity.
  + To disable validation entirely, you can set validationLevel to off.
* More one Schema Validation (Recommended):   
  <https://docs.mongodb.com/manual/core/schema-validation/>

### Configuring a Collection

* In MongoDB, a collection is automatically created when we insert at least one document.
* If you want to configure your collection in a special way, well then you can use the db.createCollection() command.
* createCollection() command takes two parameters. 1st arg is – name of the collection and 2nd arg is a document where you can configure that collection.
* You can define that validation schema when you create the collection using createCollection() command or also thereafter with the help of runCommand().

# Tips and Tricks

* As database administrator, you can use db.runCommand in the shell to run administrative command.