

MONGODB

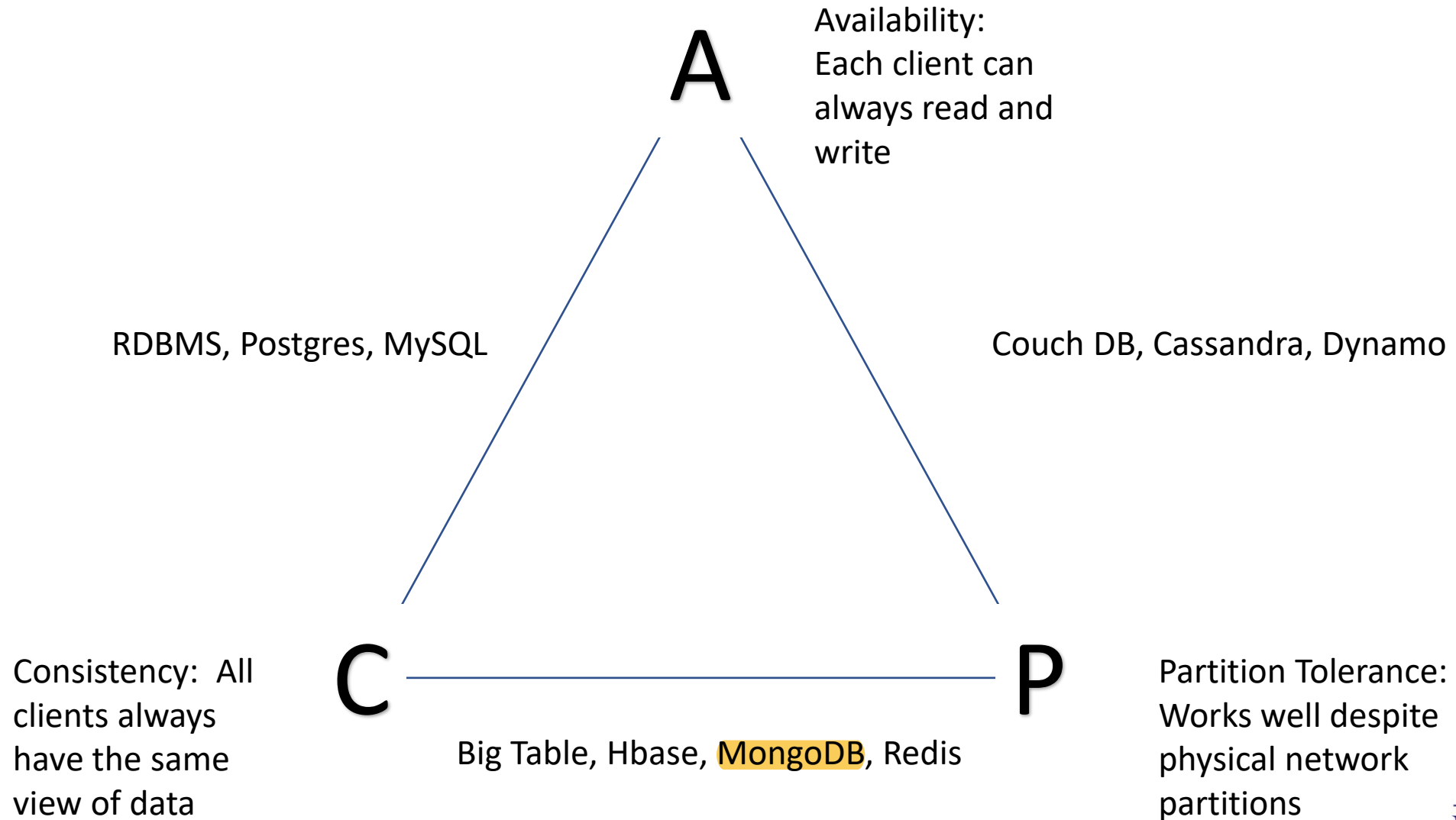
Intro



What is MongoDB

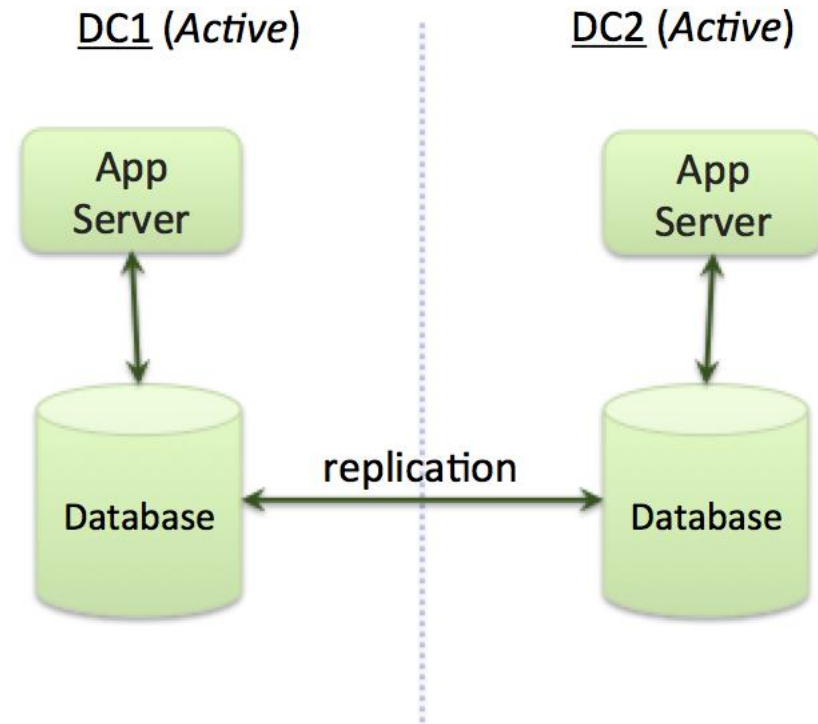
- Developed by 10gen
- Founded in 2007
- A document-oriented, NoSQL database
 - Hash-based, schema-less database
 - No Data Definition Language
 - In practice, this means you can store hashes with any keys and values that you choose
 - Keys are a basic data type but in reality stored as strings
 - Document Identifiers (_id) will be created for each document, field name reserved by system
 - Application tracks the schema and mapping
 - Uses BSON format
 - Based on JSON – B stands for Binary

Where does it fit in CAP Theorem



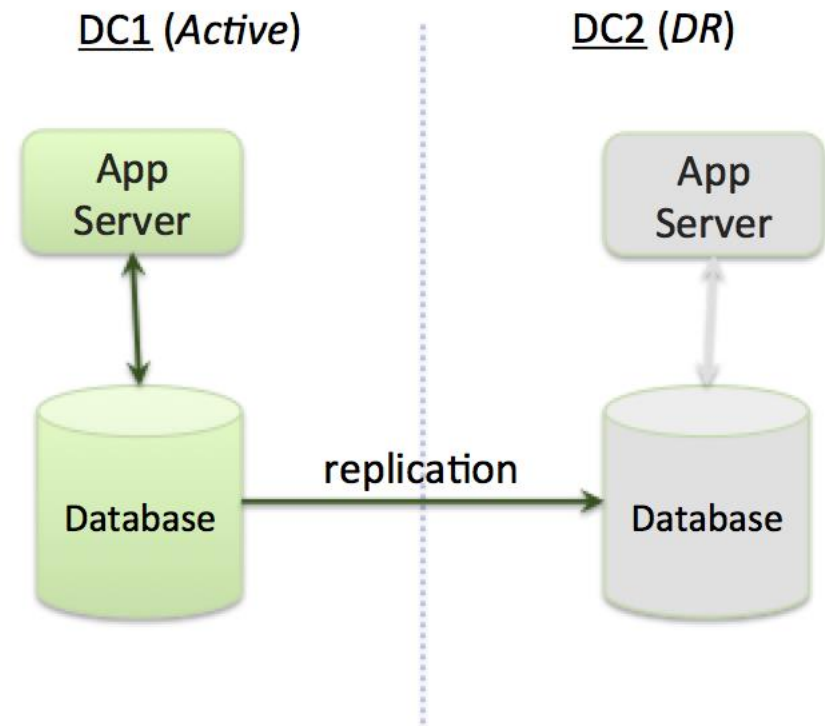
Active Active Architecture

- Serve a globally distributed audience by providing **local processing (low latencies)**
- Maintain **always-on availability**, even in the face of complete regional outages
- Provide the best utilization of platform resources by allowing server resources in multiple data centers to be used in parallel to process application requests.



Active DR Architecture

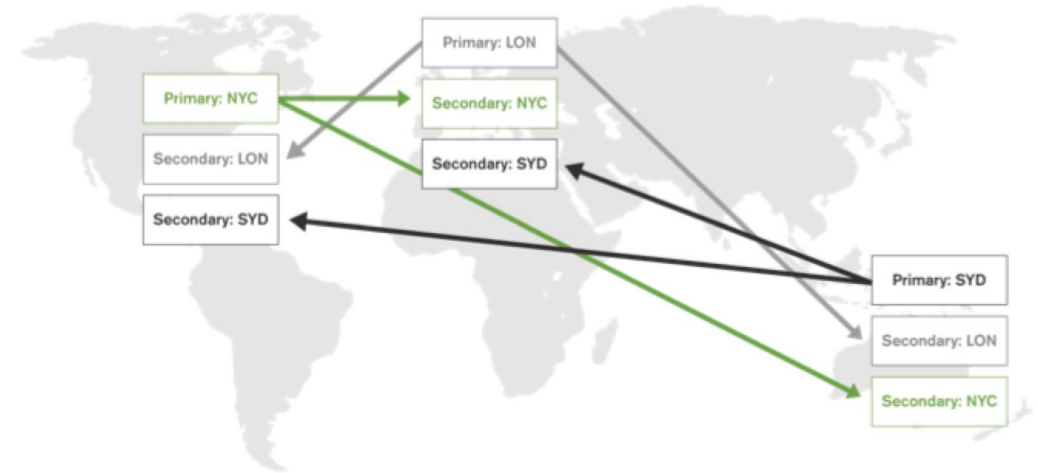
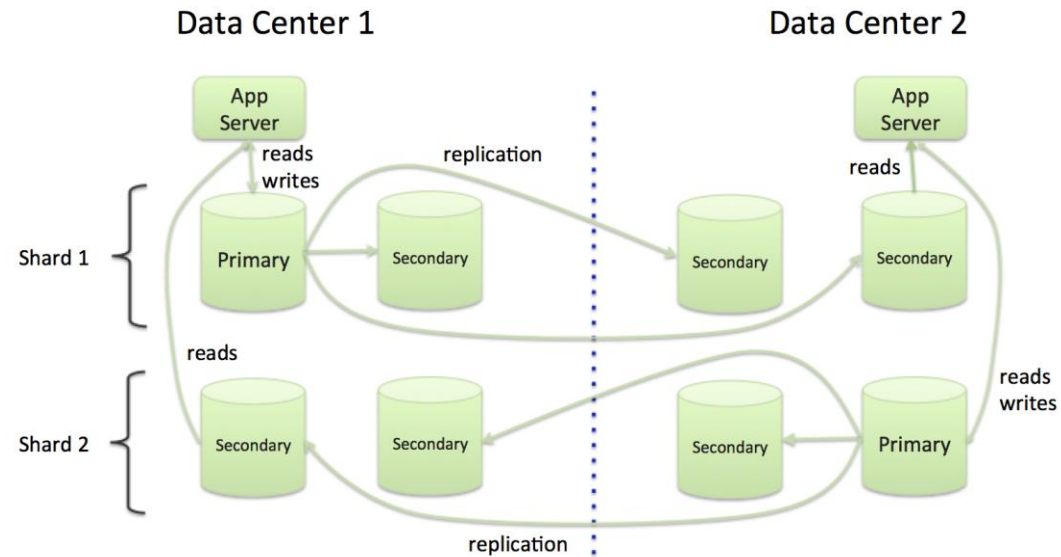
- A simpler architecture to maintain
- Replication copies data from the active node to the DR node
- DR site takes over after failure of the Active node
- More expensive to operate due to idle resources



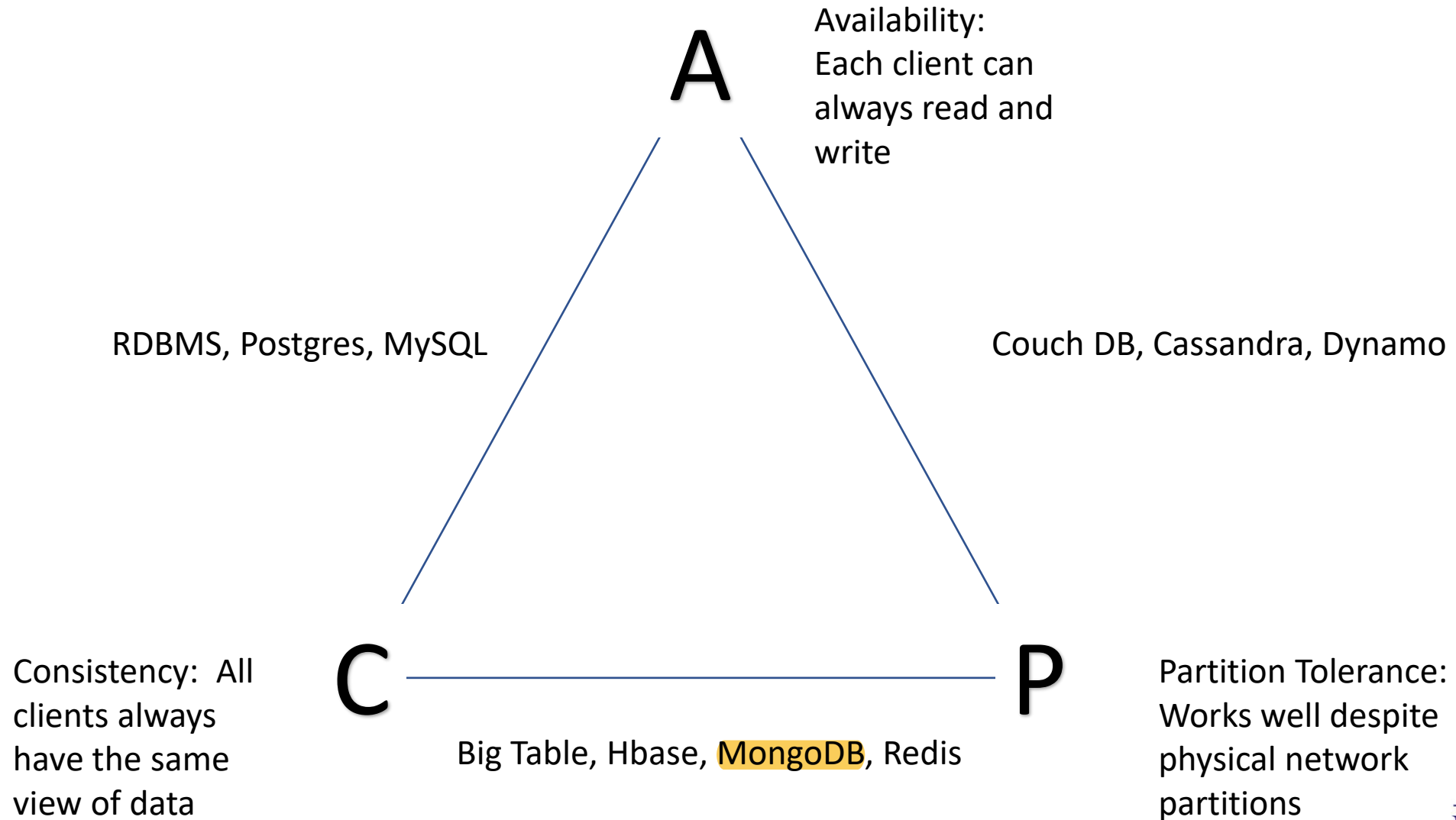
Distributed Architectures

- Distributed transactions using two-phase commit
- Multi-Master, sometimes also called "masterless"
- Partitioned (sharded) database with multiple primaries each responsible for a unique partition of the data

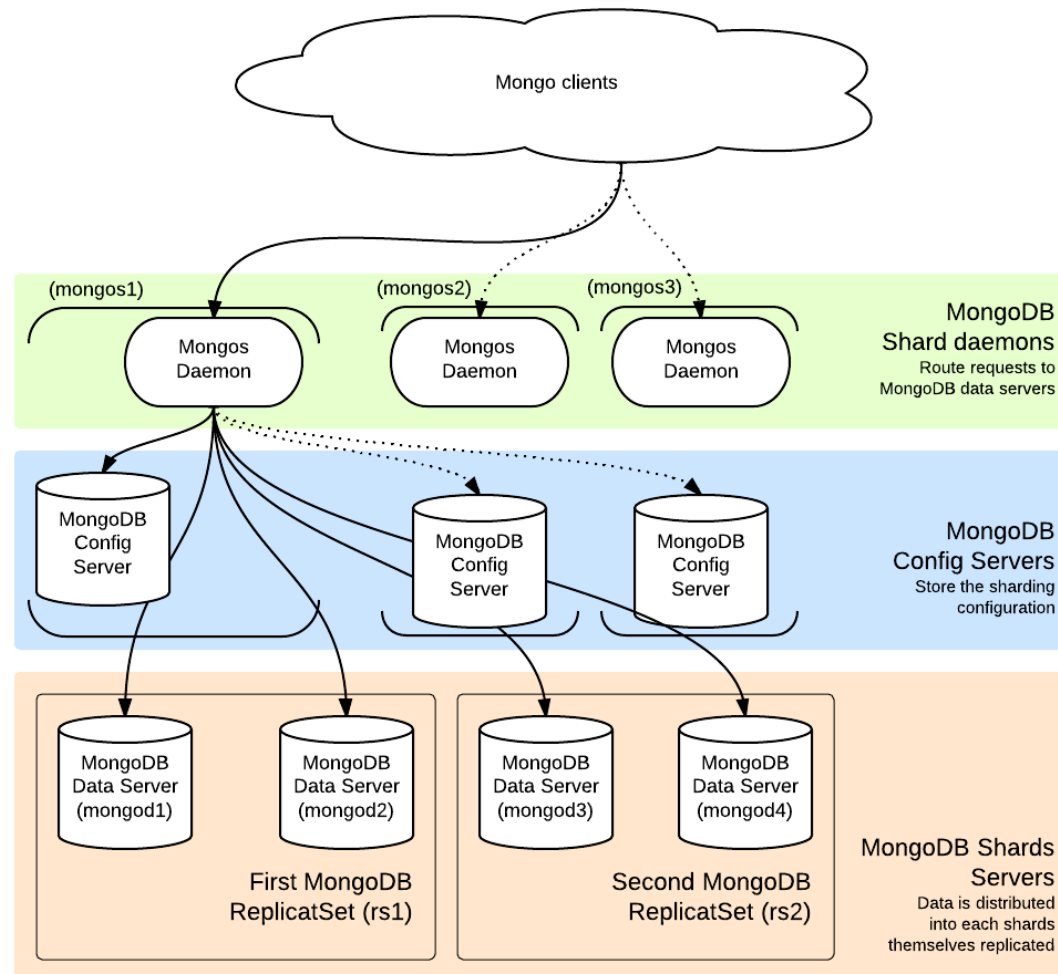
Sharded Architecture



Where does it fit in CAP Theorem



MongoDB Architecture



Replication

- A replica set is a group of mongod instances that maintain the same data set.
- A replica set contains several data bearing nodes and optionally one arbiter node.
 - Of the data bearing nodes, one and only one member is deemed the primary node, while the other nodes are deemed secondary nodes.
- The primary node receives all write operations.
 - A replica set can have only one primary capable of confirming writes with { w: "majority" } write concern; although in some circumstances, another mongod instance may transiently believe itself to also be primary.
 - Primary is Primary until decided otherwise or fails to reply to heartbeat

Sharding

- *Vertical Scaling* involves increasing the capacity of a single server, such as using a more powerful CPU, adding more RAM, or increasing the amount of storage space. Limitations in available technology may restrict a single machine from being sufficiently powerful for a given workload. Additionally, Cloud-based providers have hard ceilings based on available hardware configurations. As a result, there is a practical maximum for vertical scaling.
- *Horizontal Scaling* involves dividing the system dataset and load over multiple servers, adding additional servers to increase capacity as required. While the overall speed or capacity of a single machine may not be high, each machine handles a subset of the overall workload, potentially providing better efficiency than a single high-speed high-capacity server. Expanding the capacity of the deployment only requires adding additional servers as needed, which can be a lower overall cost than high-end hardware for a single machine. The trade off is increased complexity in infrastructure and maintenance for the deployment.

Structure

- **The database** : In simple words it can be called as the physical container for data.
- **The Collection** : A group of database documents can be called as a collection.
- **The Document** : A set of key-value pairs can be designated as a document.

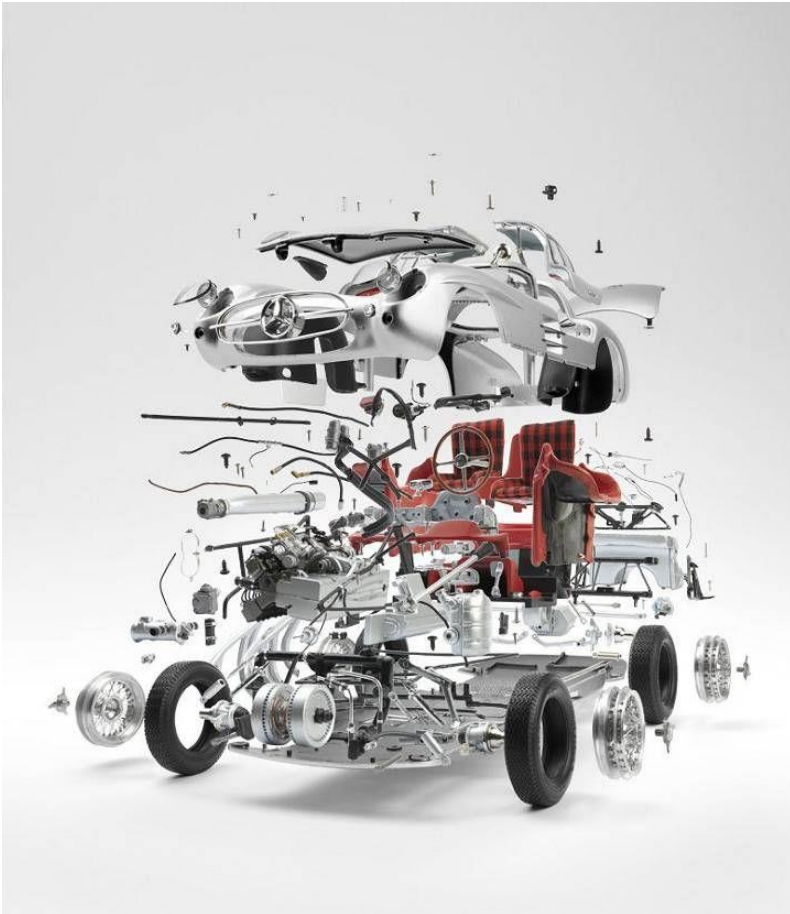
Data Model

- Stores data in form of BSON (binary JavaScript Object Notation) documents

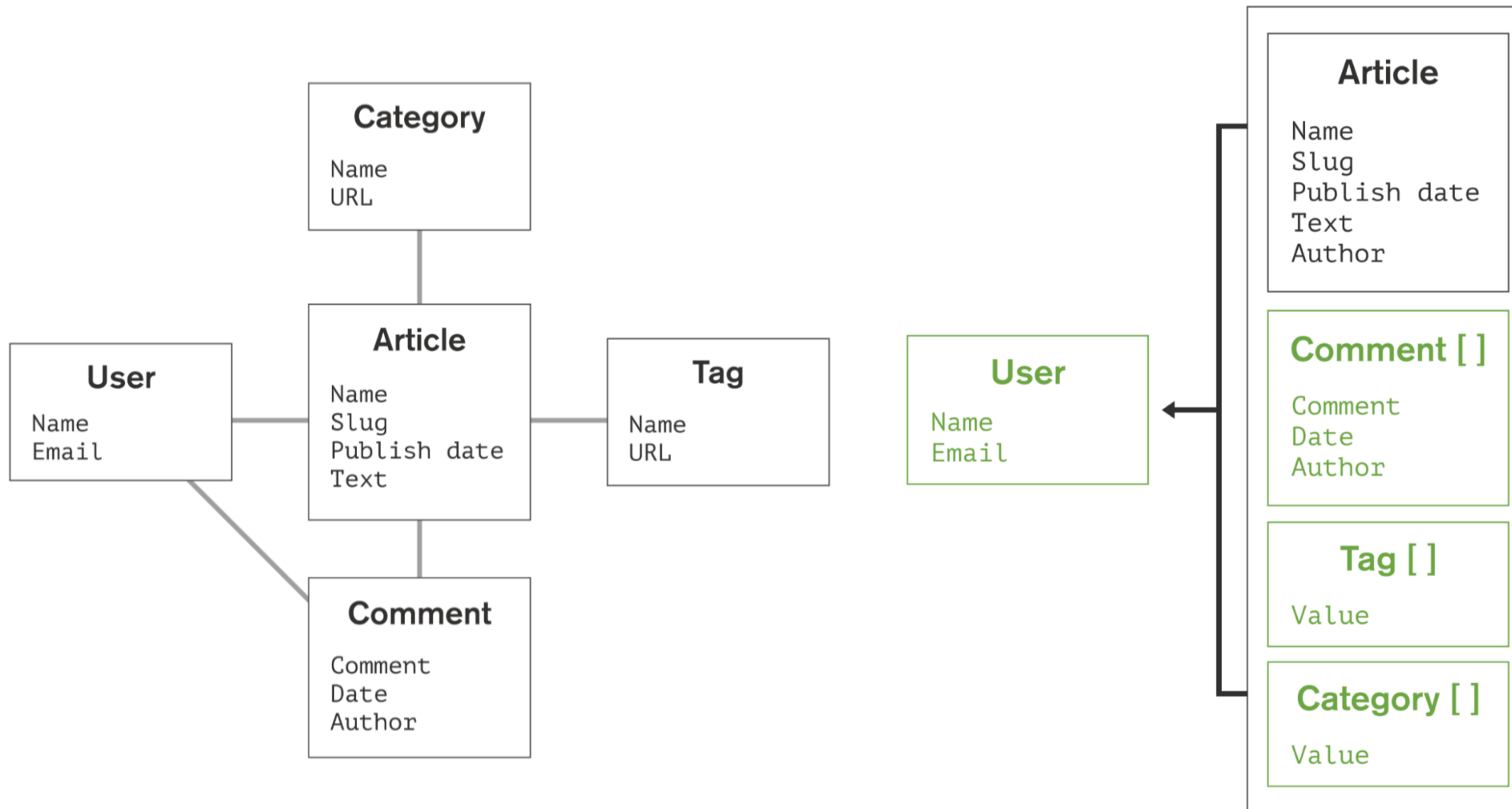
```
{    name: "Ray Ready",  
    salary: 30000,  
    designation: "Adjunct Professor",  
    courses: [ "database", "database 2", "data warehousing" ]  
}
```

- Group of related documents with a shared common index is a collection

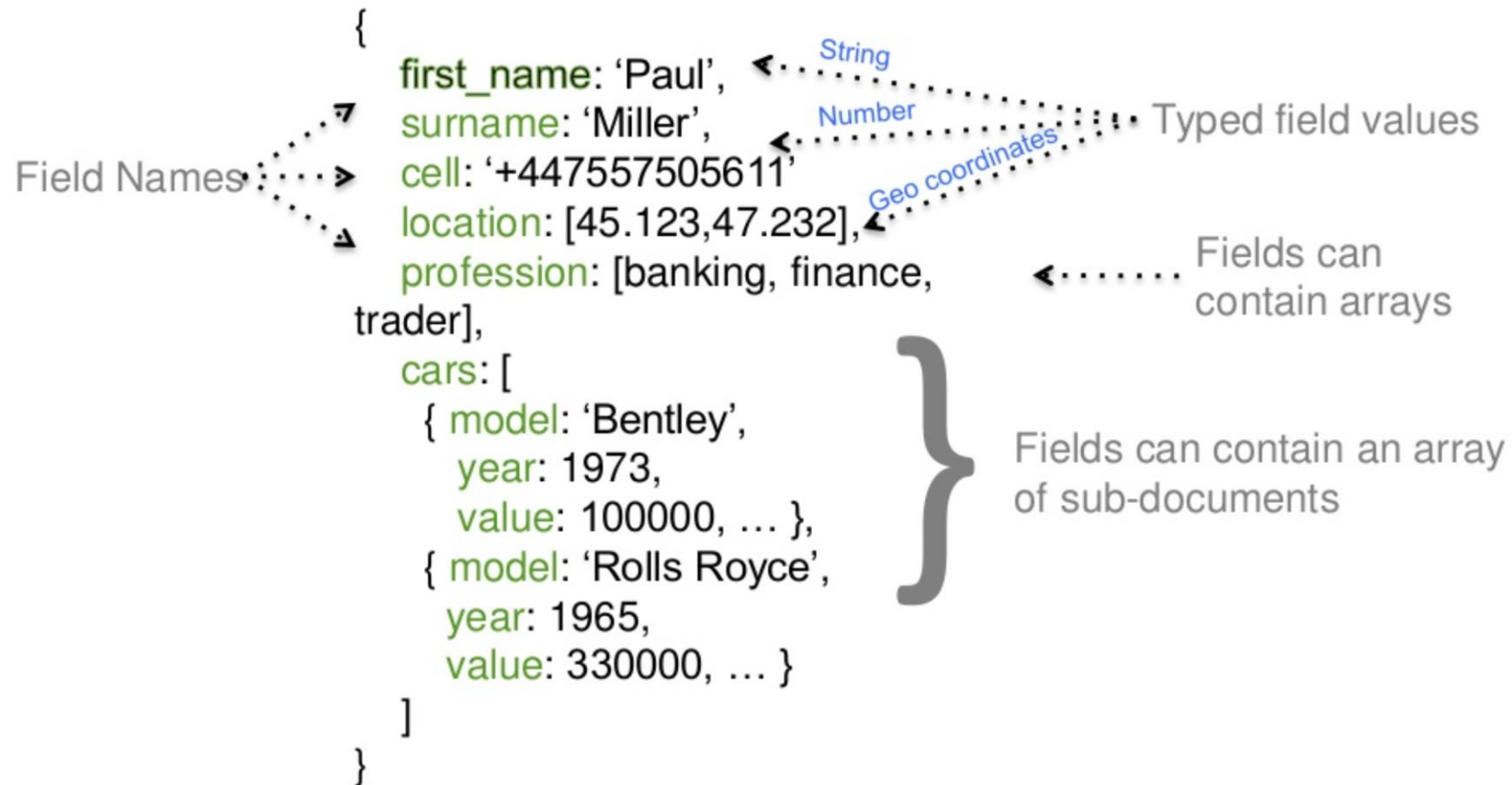
RDBMS vs Document/Object Model



Example Relational to Mongo



Natural Model



Data Types

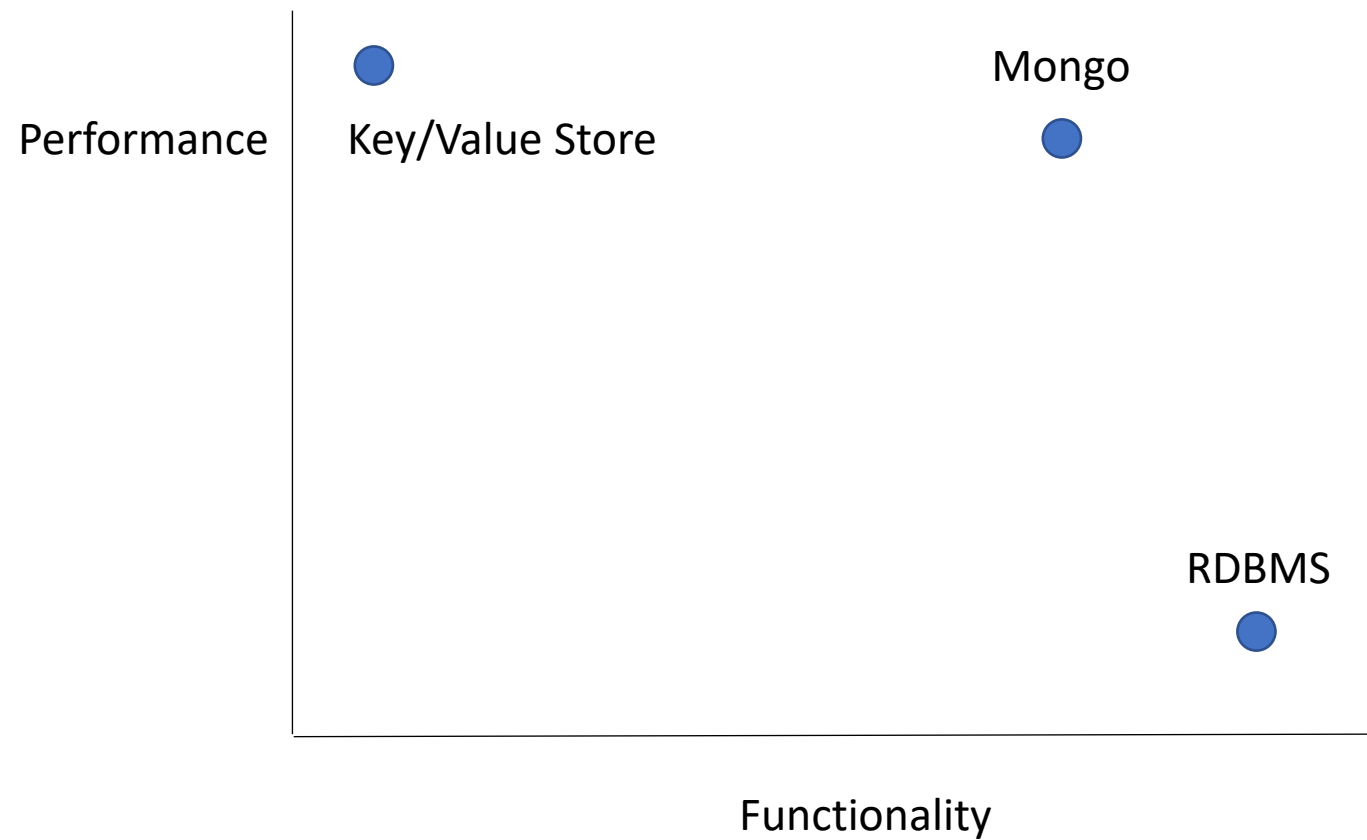
Type	Description	Number
Double	Represents a float value.	1
String	BSON strings are UTF-8. In general, drivers for each programming language convert from the language's string format to UTF-8 when serializing and deserializing BSON. This makes it possible to store most international characters in BSON strings with ease. [1] In addition, MongoDB \$regex queries support UTF-8 in the regex string.	2
Object	Represents an embedded documents.	3
Array	Sets or lists of values can be represented as arrays:	4
Binary data	Binary data is a string of arbitrary bytes, it cannot be manipulated from the shell.	5
Object id	ObjectIds (MongoDB document identifier, equivalent to a Primary key) are: small, likely unique, fast to generate, and ordered. These values consists of 12-bytes, where the first four bytes are a timestamp that reflect the ObjectId's creation.	7
Boolean	A logical true or false. Use to evaluate whether a condition is true or false	8
Date	BSON Date is a 64-bit integer that represents the number of milliseconds since the Unix epoch (Jan 1, 1970). This results in a representable date range of about 290 million years into the past and future.	9
Null	It represents both a null value and a nonexistent field.	10
Regular Expression	RegExp maps directly to a Javascript RegExp	11
JavaScript		13
Symbol	Not supported by the shell. If the shell gets a symbol from the database, it will convert it into a string.	14
32-bit integer	Numbers without decimal points will be saved as 32-bit integers.	16
Timestamp	BSON has a special timestamp type for internal MongoDB use and is not associated with the regular Date type. Timestamp values are a 64 bit value where : the first 32 bits are a time_t value (seconds since the Unix epoch). the second 32 bits are an incrementing ordinal for operations within a given second.	17
64-bit integer	Numbers without a decimal point will be saved and returned as 64-bit integers.	18

Extensive Query Model

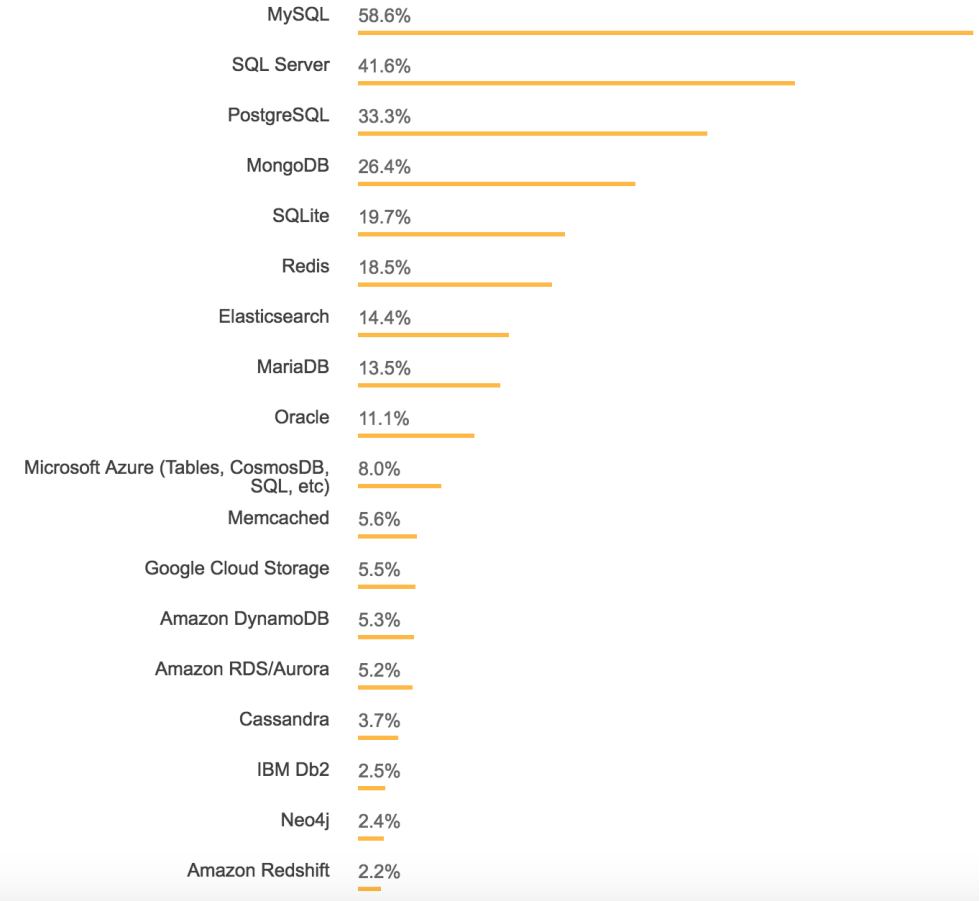
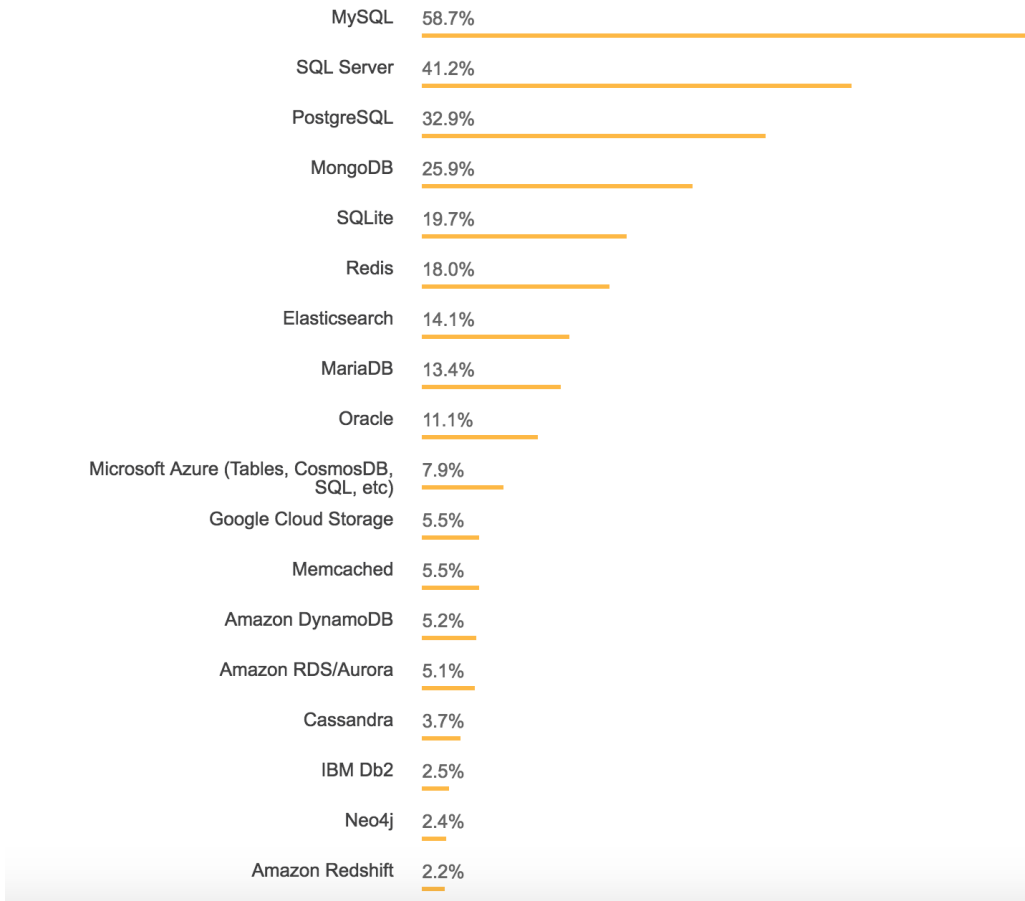
Rich Queries	<ul style="list-style-type: none">Find Paul's carFind everybody in London with a car built between 1970 and 1980
Geo	<ul style="list-style-type: none">Find all car owners within 10m of Manhattan
Text Search	<ul style="list-style-type: none">Find all cars whose VIN starts with ZA1
Aggregation	<ul style="list-style-type: none">Calculate the average value of Paul's car collection
Map Reduce	<ul style="list-style-type: none">What is the ownership pattern of colors by geography over time? (Is silver trending up in the US?)

```
{  
  first_name: 'Paul',  
  surname: 'Miller',  
  cell: '+447557505611'  
  location: [45.123,47.232],  
  profession: [banking, finance,  
trader],  
  cars: [  
    { model: 'Bentley',  
      year: 1973,  
      value: 100000, ... },  
    { model: 'Rolls Royce',  
      year: 1965,  
      value: 330000, ... }  
  ]  
}
```

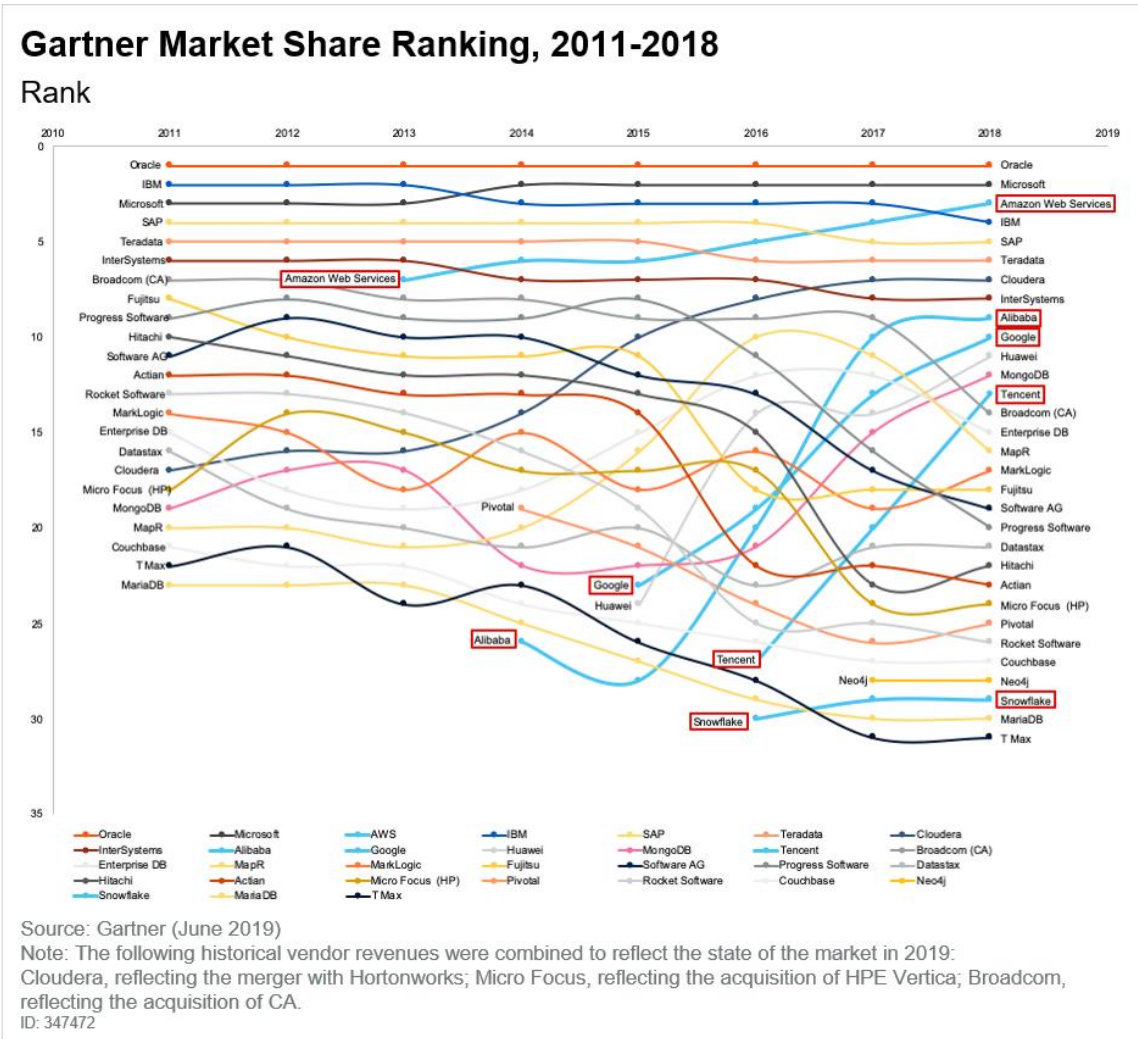
Mongo Strengths



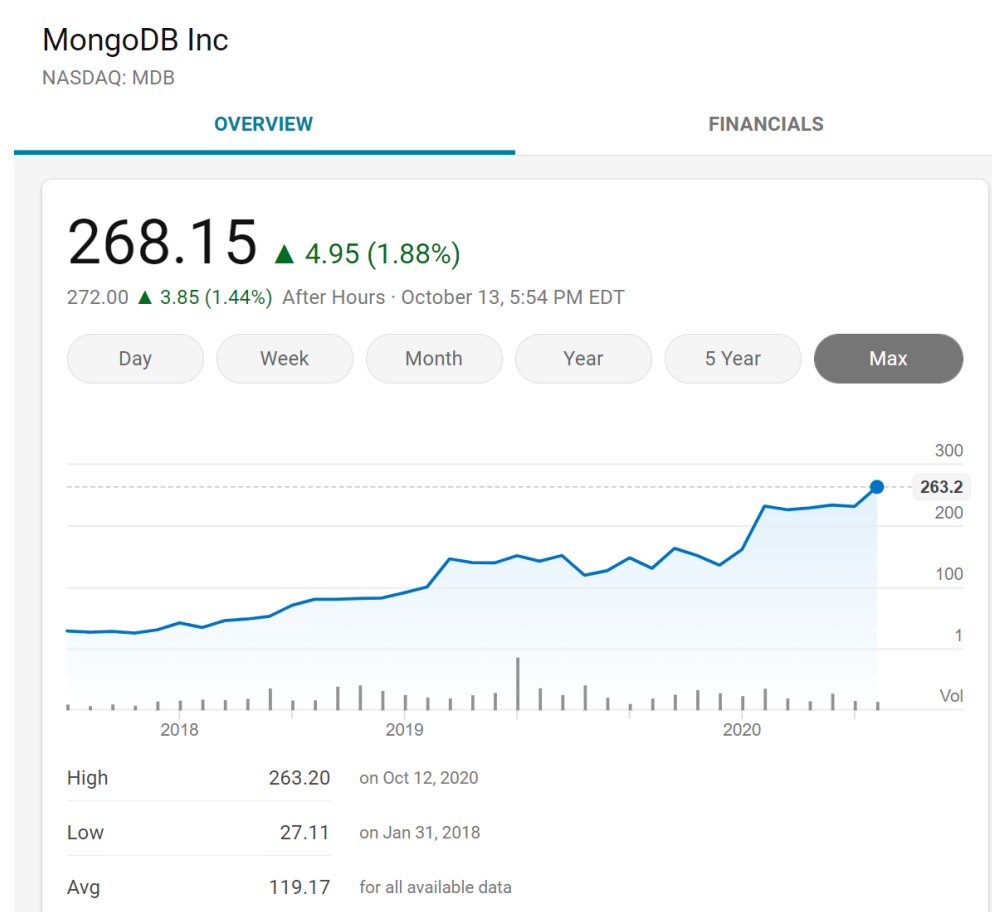
Database Market Share



Market Share Ranking



Mongo Stock Price



Customers

Telefonica

“

MongoDB helped us to deliver that 360 view of the customer in just 90 days.”

— John Bungert, Metlife

 Tweet



AstraZeneca 

CHICO'S

NOKIA

 **Medtronic**


Adobe

bouygues 



ebay™

“

Without MongoDB, it would've taken a much bigger team to get the app live so quickly.”

— Prashanth Kokati, Expedia

 Tweet

Google


CISCO™

SAP®

facebook

Genentech


verizon

NBCUniversal

vivint.

 HM Revenue & Customs

FLORIDA
POLYTECHNIC
UNIVERSITY

The Weather Channel

- 2 million requests per day
- 40 million users with alerts
- Mobile backend as a service for Facebook
- 270,000 apps
- Unknown Workload
- API requests growing 500%

Craigslist

- 1.5 million new ads posted every day
- Archive billions of records
- Changing schema
- Scalability

Expedia Scratch Pad

- Needs to handle a lot of data and process a huge amount of information to automate, filter, and personalize to deliver the exact content
- Easy to customize all of the information pretty quickly and without any hassle
- Speed and ease of scalability as you can store the values in a single view
- Deliver results in real-time