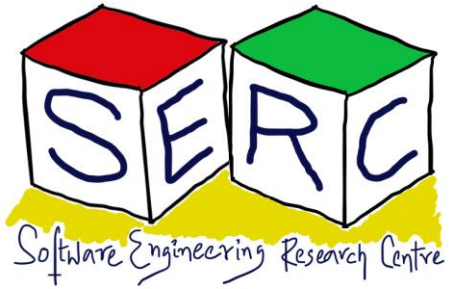
Unstructured Database NoSQL

Spring 2025

CS6.201 - Introduction to Software Systems

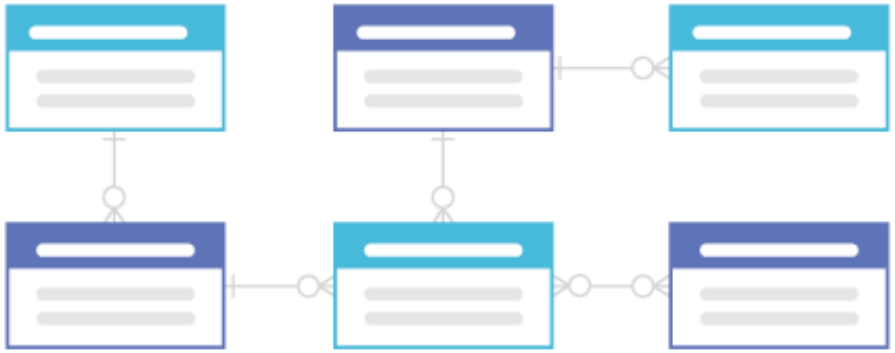


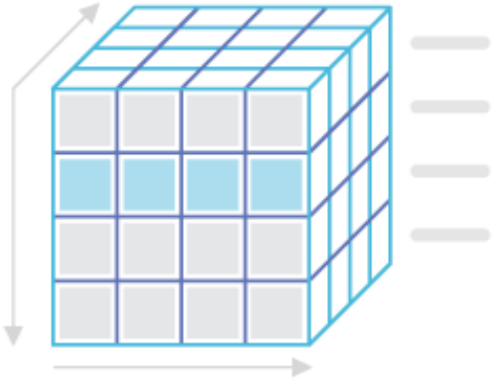
NoSQL (often interpreted as Not only SQL) database. It provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases.

SQL	NoSQL
Relational Database Management System (RDBMS)	Non-relational or distributed database system.
These databases have fixed or static or predefined schema	They have dynamic schema
These databases are best suited for complex queries	These databases are not so good for complex queries
Vertically Scalable	Horizontally scalable
Follows ACID property	Follows BASE property

SQL

Relational Database Management Systems (RDBMS)

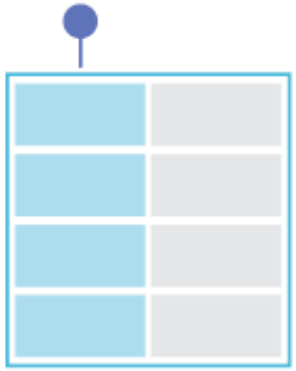




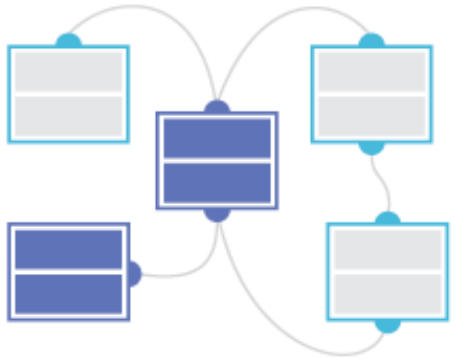
Online Analytical Processing (OLAP) Cube

NoSQL

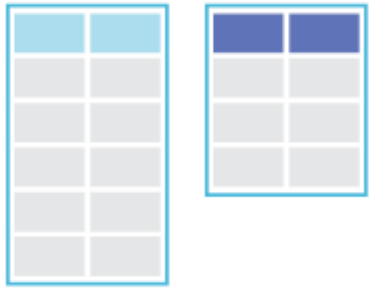
Key-Value



Graph



Document

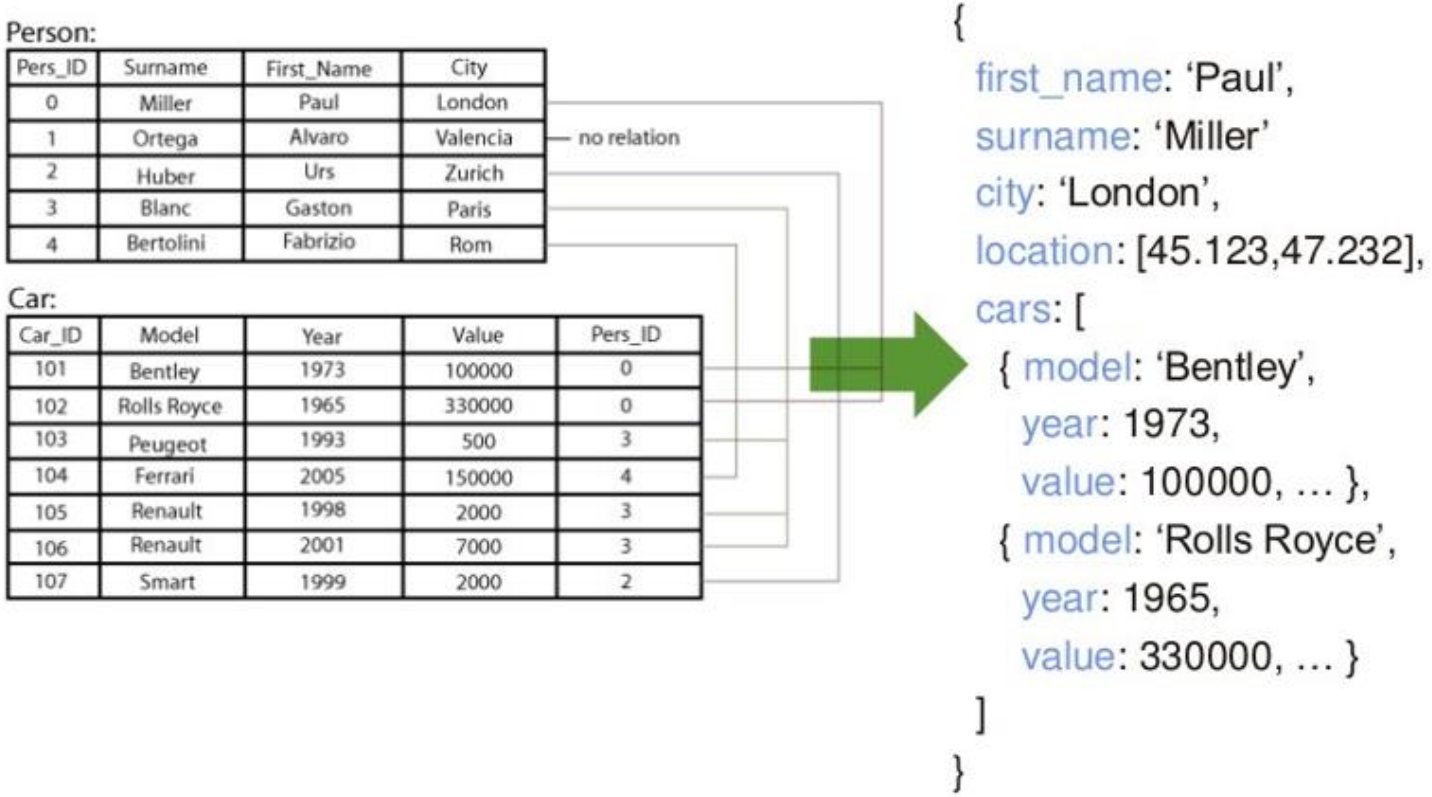


Column store

Atomicity, Consistency, Isolation, Durability

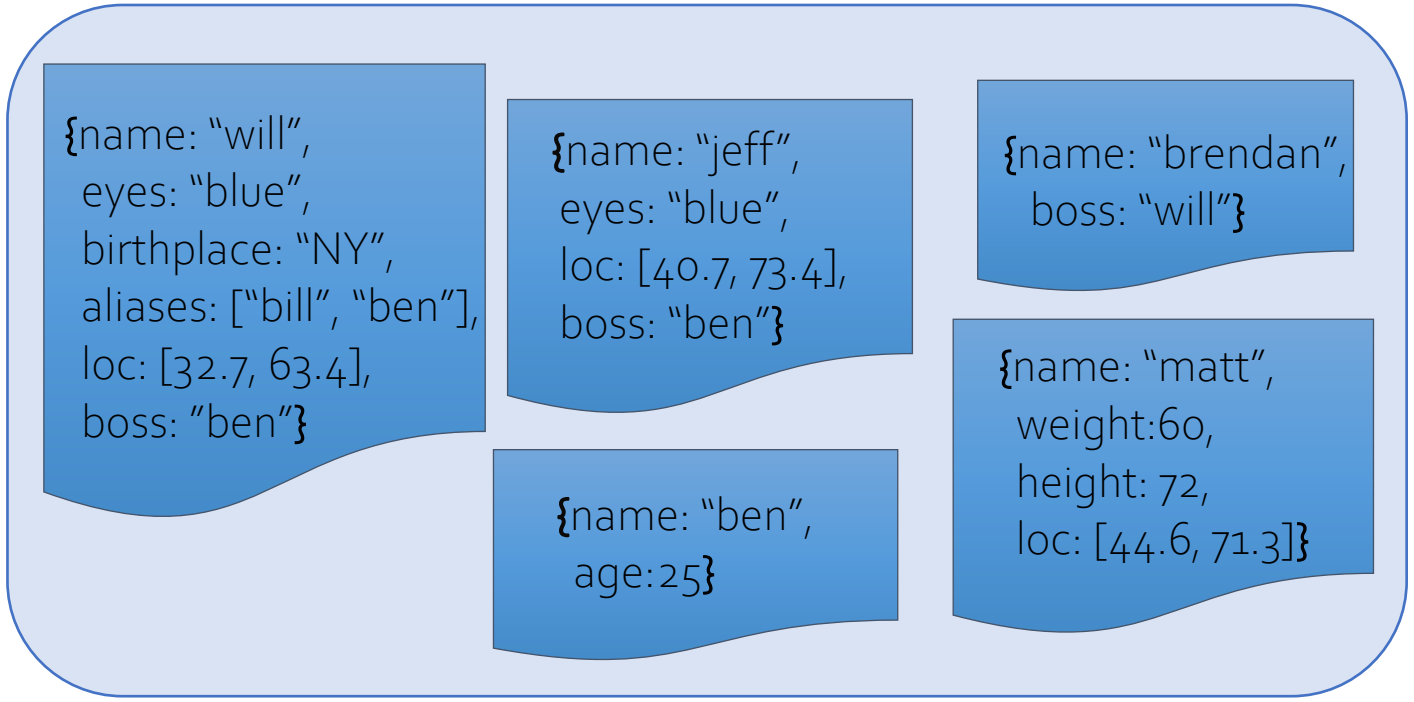
Basically Available, Soft State, Eventually Consistent

MongoDB is an open source, document-oriented database designed with both scalability and developer agility in mind. Instead of storing your data in tables and rows as you would with a relational database, in MongoDB you store JSON-like documents with dynamic schemas(schema-free, schema less).



Relational Database	Document Database
Database	Database
Table	Collection
Row	Document (JSON, BSON)
Column	Field
Index	Index
Join	Embedded Document
Partition	Shard

MongoDB does not need any pre-defined data schema. Every document could have different data!



Document-Oriented storege

Full Index Support

Replication & High Availability

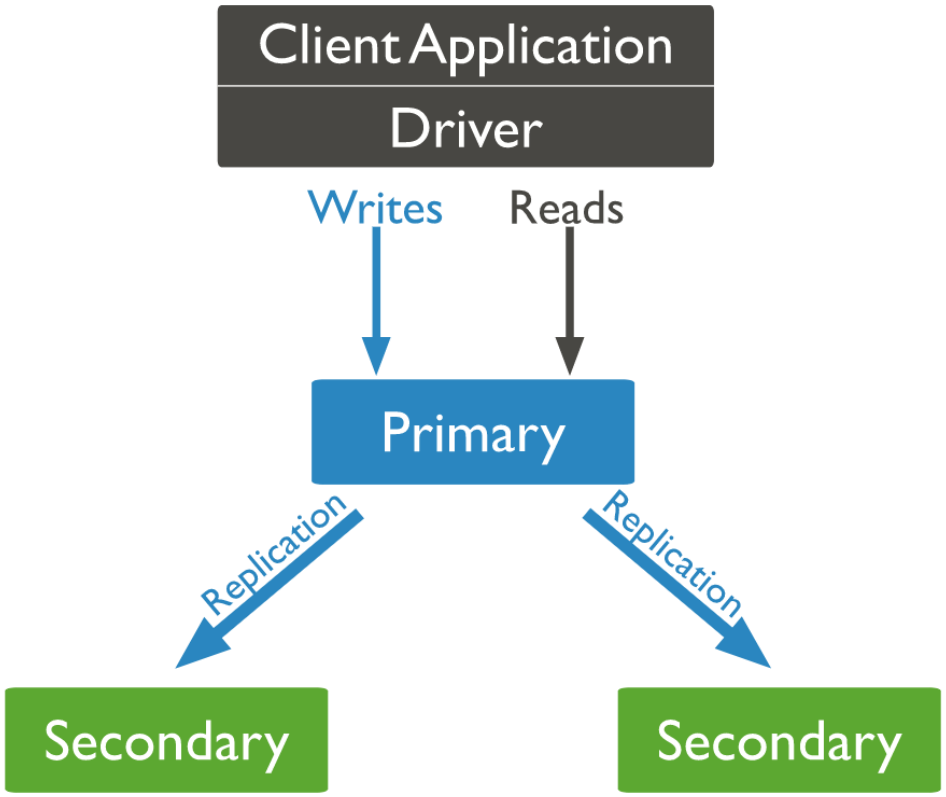
Auto-Sharding

Aggregation

MongoDB Atlas

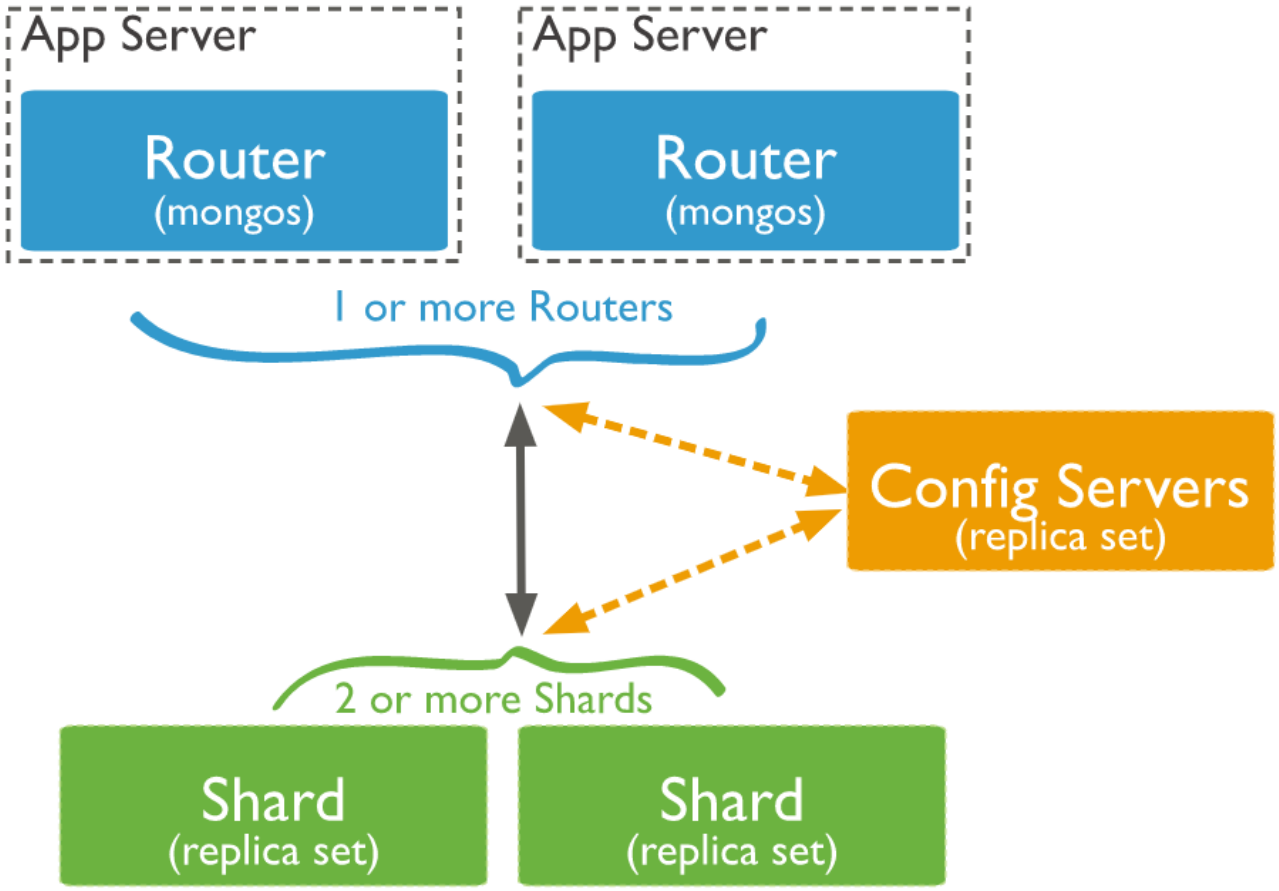
Various APIs - JavaScript, Python, Ruby, Perl, Java, Java, Scala, C#, C++, Haskell, Erlang

Community



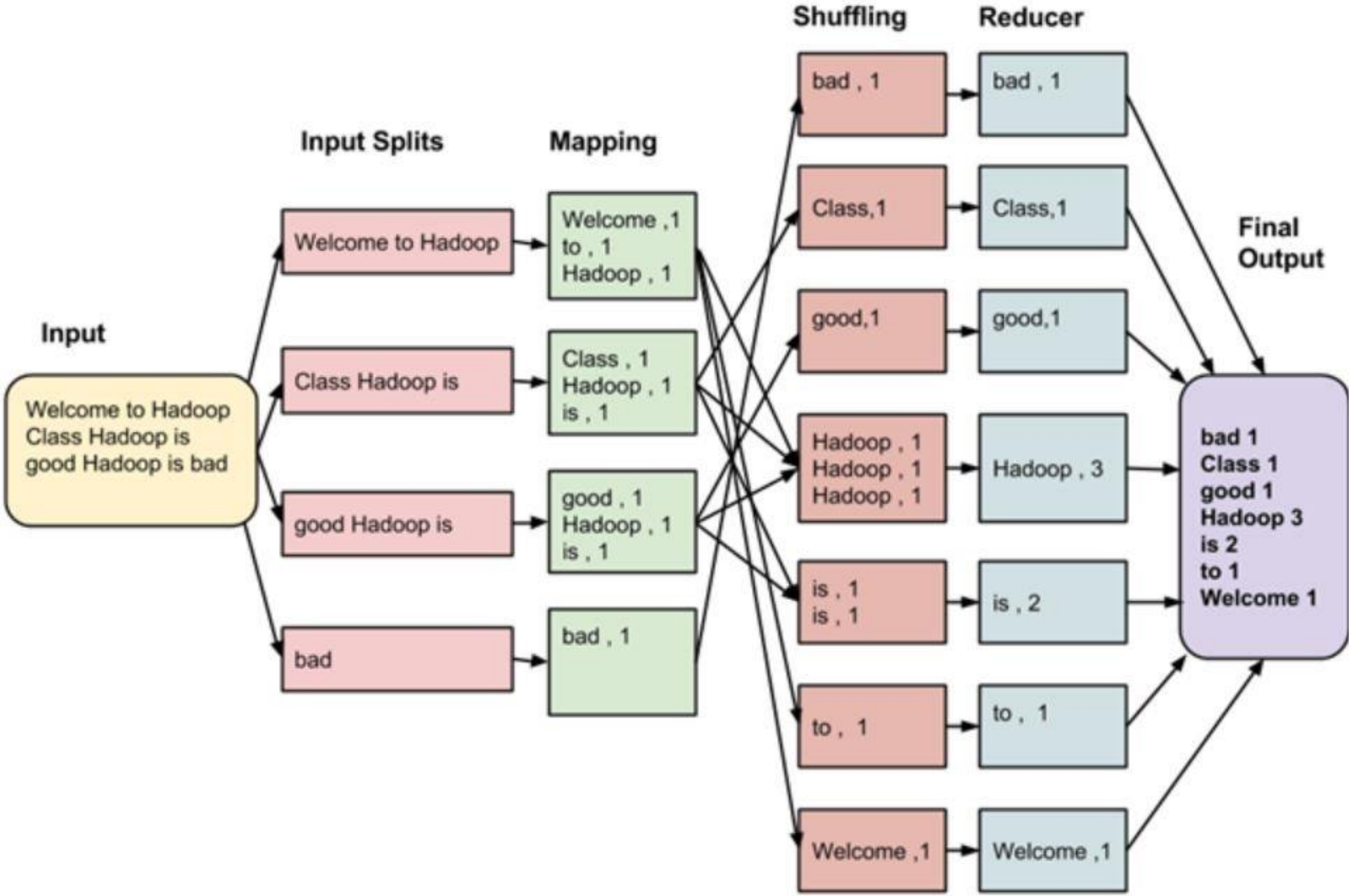
Replication

It provides redundancy and increases data availability. With multiple copies of data on different database servers, replication provides a level of fault tolerance against the loss of a single database server.



Sharding (Replica-Sets)

It is a method for distributing data across multiple machines. MongoDB uses sharding to support deployments with very large data sets and **high throughput operations.**


Input
Welcome to Hadoop
Class Hadoop is
good Hadoop is bad
Input Splits
Welcome to Hadoop
Class Hadoop is
good Hadoop is
bad
Mapping
Welcome ,1
to , 1
Hadoop , 1
Class , 1
Hadoop , 1
is , 1
good , 1
Hadoop , 1
is , 1
bad , 1
Shuffling
bad , 1
Class,1
good,1
Hadoop , 1
Hadoop , 1
Hadoop , 1
is , 1
is , 1
to , 1
Welcome ,1
Reducer
bad , 1
Class,1
good,1
Hadoop , 3
is , 2
to , 1
Welcome ,1
Final Output
bad 1
Class 1
good 1
Hadoop 3
is 2
to 1
Welcome 1

Collection

```
db.orders.mapReduce(
          map      →   function() { emit( this.cust_id, this.amount ); },
          reduce   →   function(key, values) { return Array.sum( values ) },
                       {
          query    →     query: { status: "A" },
          output   →     out: "order_totals"
                       }
                   )
```