



University of
Salford
MANCHESTER



SCHOOL OF
**SCIENCE, ENGINEERING
& ENVIRONMENT**

Big Data Tools and Techniques

Week 9

NoSQL Databases

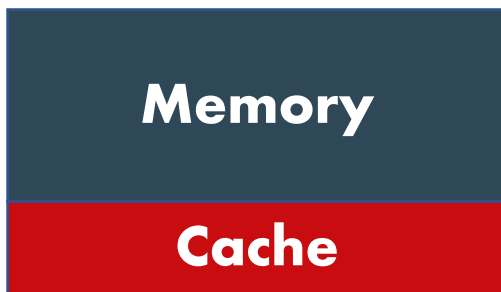
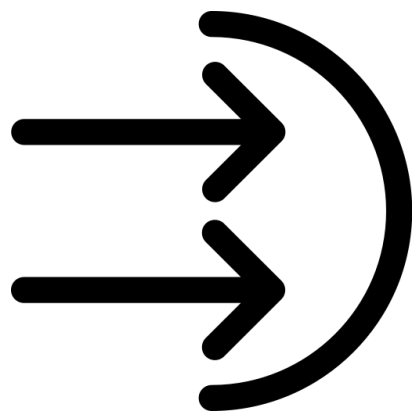
2025

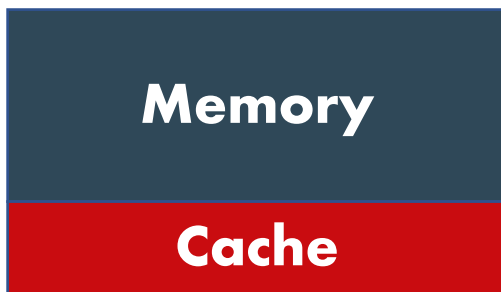
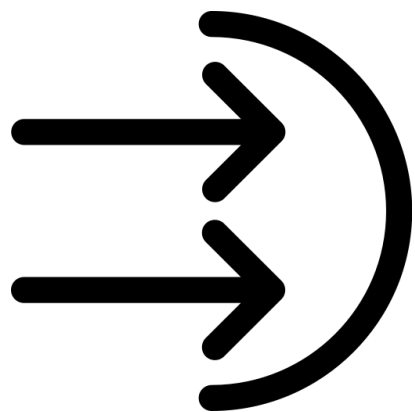
Expectations

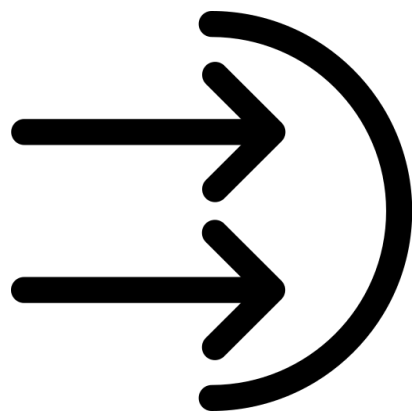
1. Choose a quiet place to attend the class and please concentrate during the lecture.
2. Put your questions in Padlet and I will review them in the due time (Padlet link is in BB, week 9, Lecture folder for Q&A week9).
3. You can find a handout on BB.
4. We will have 5 mins break after the first hour of the lecture (please remind me).
5. Jisc code will be shared during the break time.

Learning Outcomes

1. To recognize the NoSQL
2. To describe differences between SQL and NoSQL
3. To evaluate an appropriate database solution in response to big data requirements
4. To use MongoDB







Memory

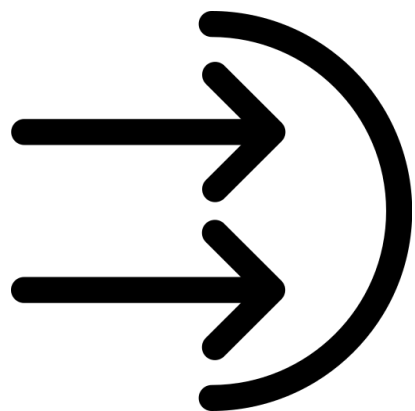
Cache

RDD

SQL

Stream

ML



Memory

Cache

RDD

SQL

Stream

ML

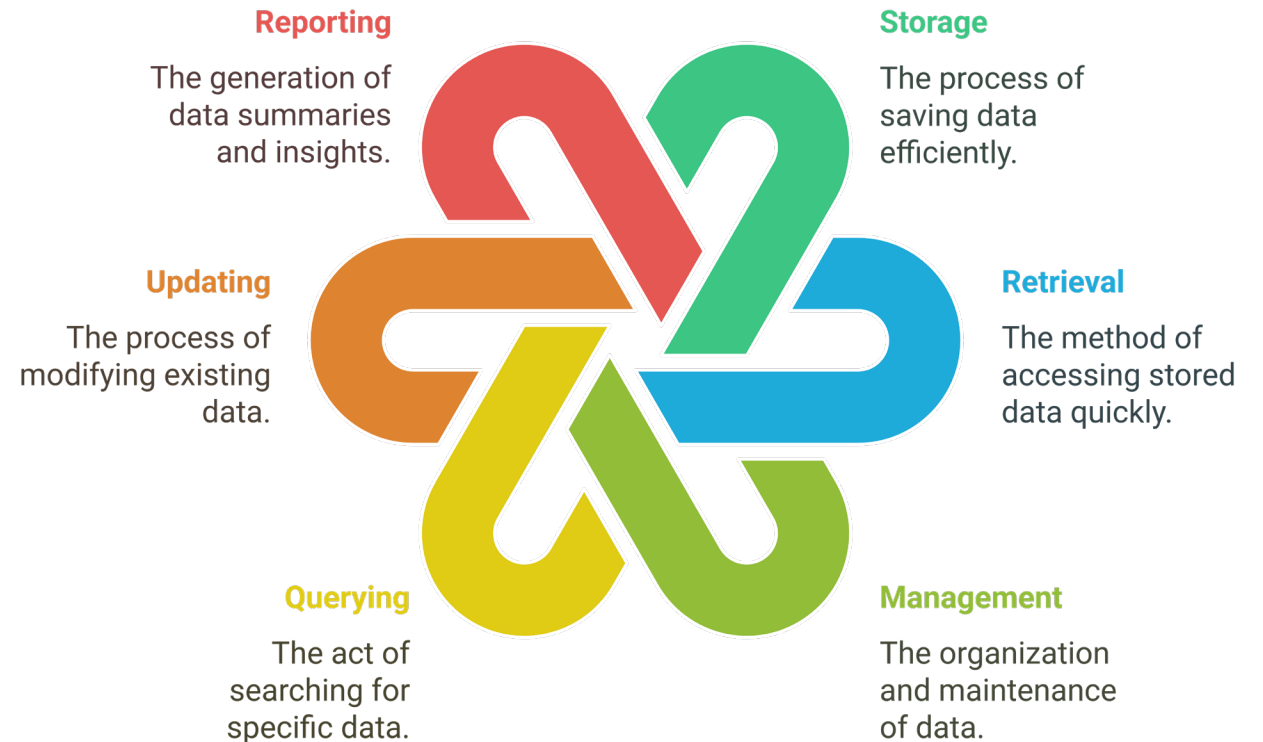
Database

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. It allows for efficient storage, retrieval, and management of data, enabling users to perform various operations such as querying, updating, and reporting.

Databases are essential for:

- Store
- Manage
- Analyse
- Share

Components of a Database



RDBMS

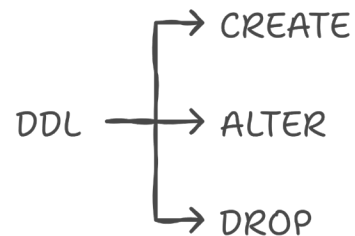
Databases usually consist of two parts:

- Schema
- Instance

SQL

SQL (Structured Query Language) is a standardized programming language specifically designed for managing and manipulating relational databases. It enables users to create, read, update, and delete data within a database, as well as define and modify its structure.

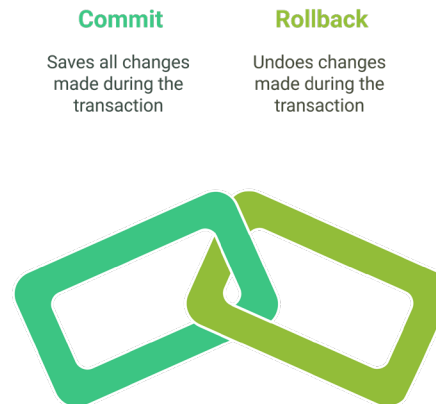
Data Definition Language (DDL) Commands



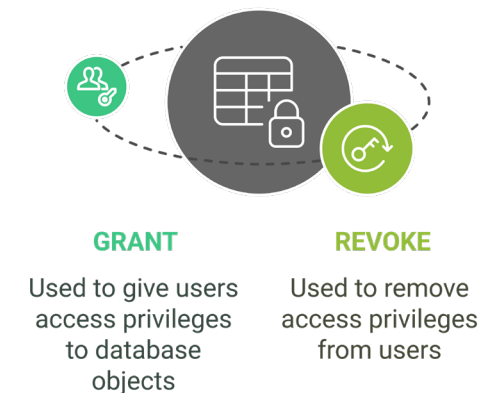
Overview of DML Operations



Understanding Transaction Control Language

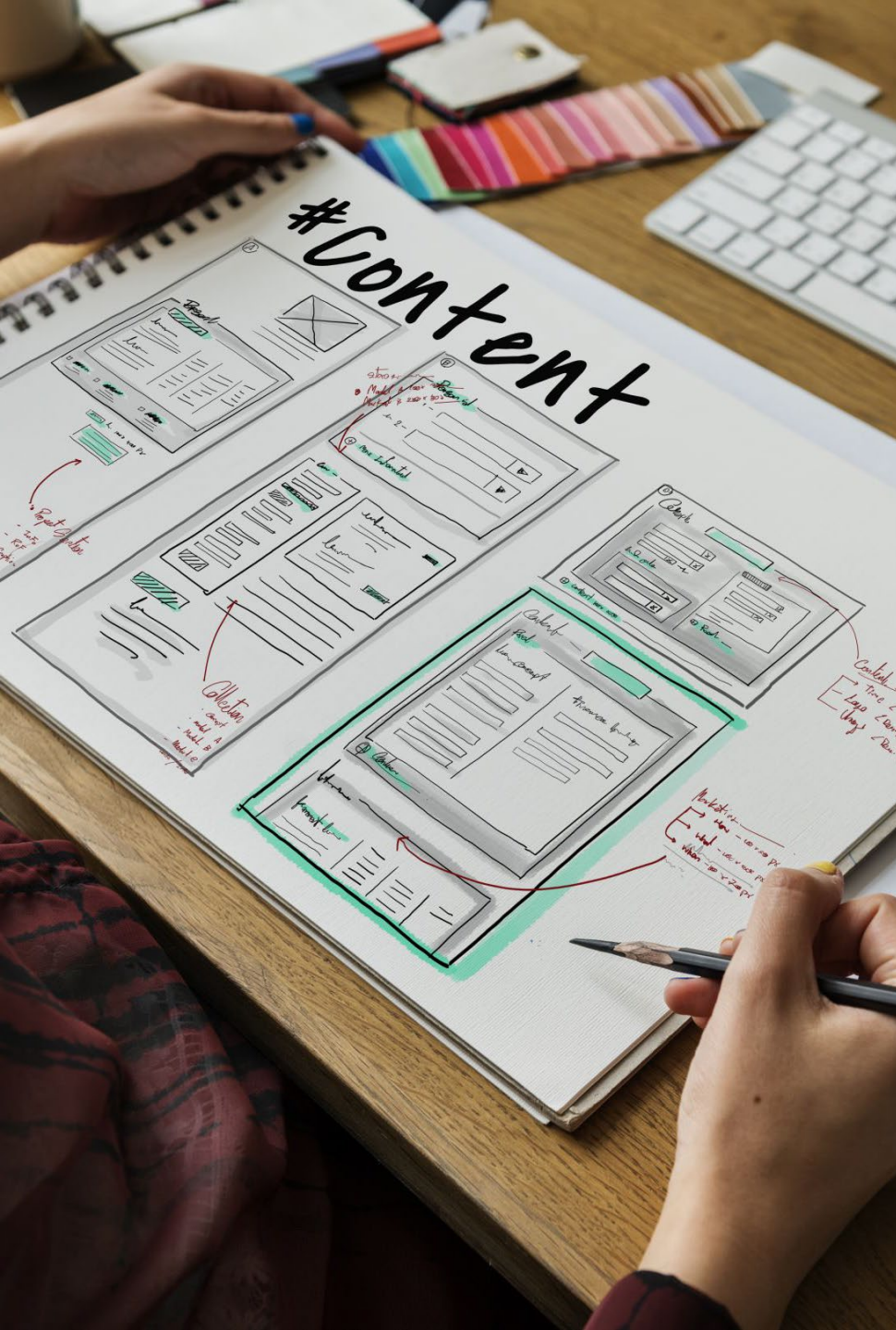


Understanding Data Control Language (DCL)



Question?

Does this approach work for every applications?



Content Management

Text
Image
Video



e-commerce

User profiles

Purchase histories

Product catalogues



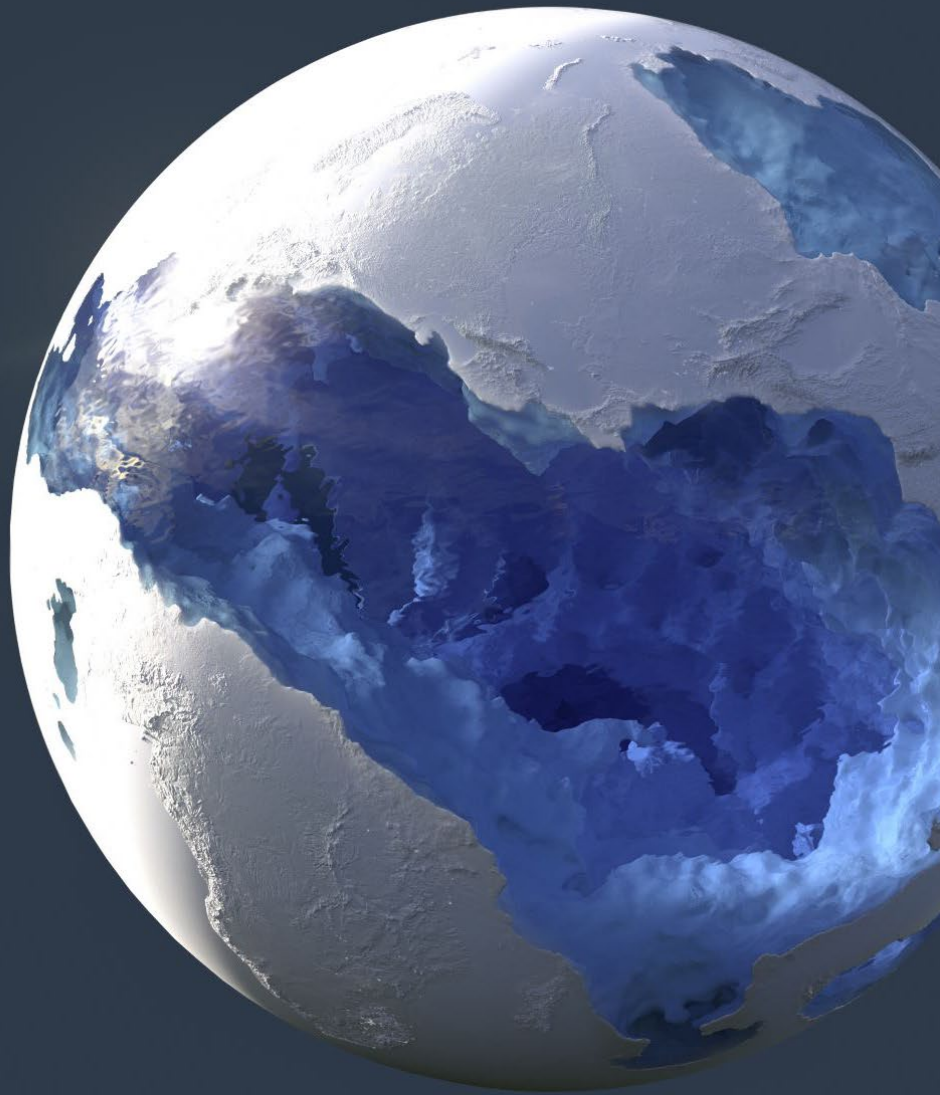
The background is a dark blue gradient with an abstract digital cityscape. It features several 3D wireframe cubes and rectangular blocks of varying sizes, some of which are illuminated from within, casting a soft blue glow. Scattered throughout the scene are numerous small, bright dots in red, green, and blue, resembling data points or stars. Some of these dots are connected by thin, faint lines, suggesting a network or data flow. The overall aesthetic is futuristic and technological.

Internet of Things

Sensors

Actuators

Controllers



Geospatial Data



SQL vs NoSQL



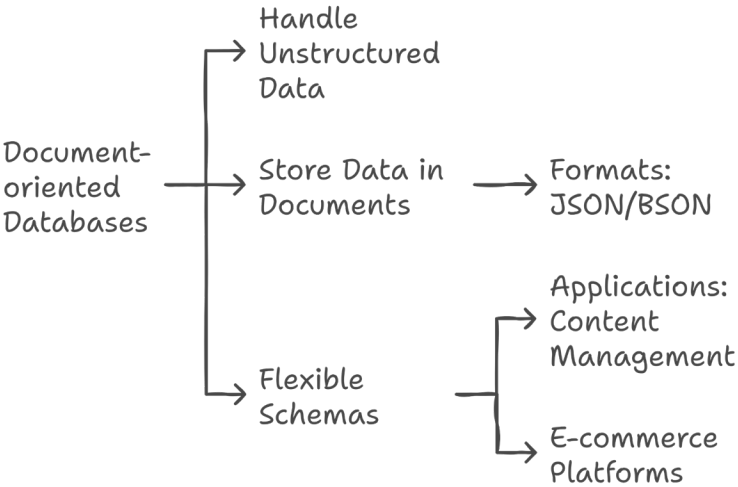
SQL vs NoSQL ...



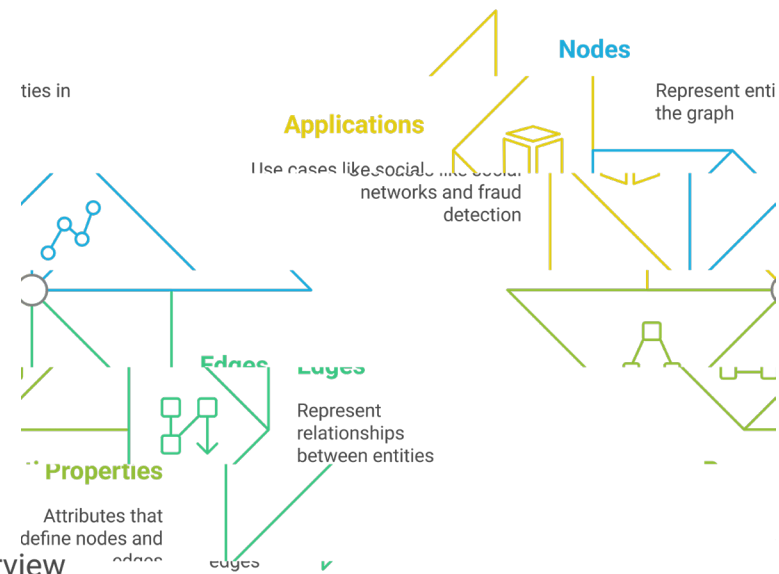
NoSQL



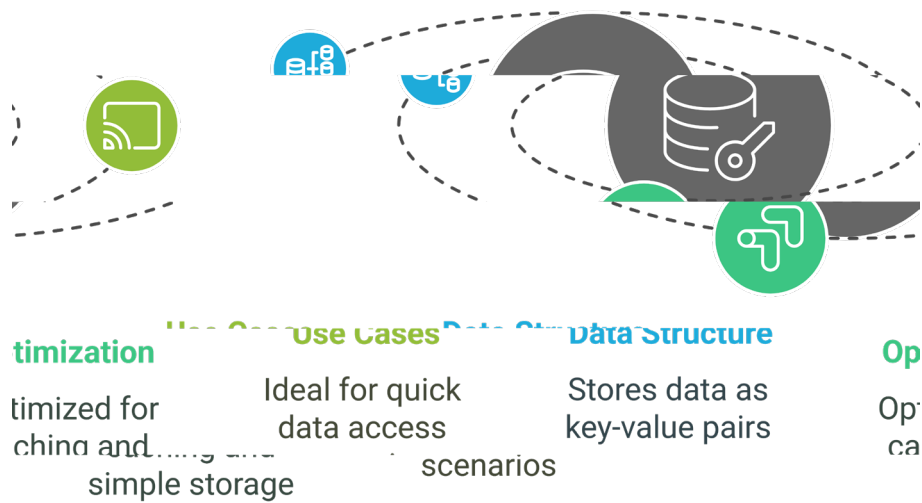
Document-oriented Databases Flowchart



Understanding Graph Databases

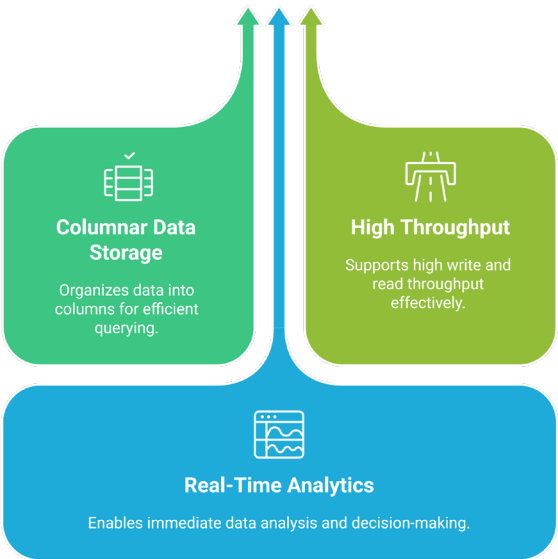


Key-Value Databases Overview



Types of NoSQL

Column-Family Database Efficiency





Popular NoSQL Databases

Why they have been popular?



Scalability



Performance



Flexibility



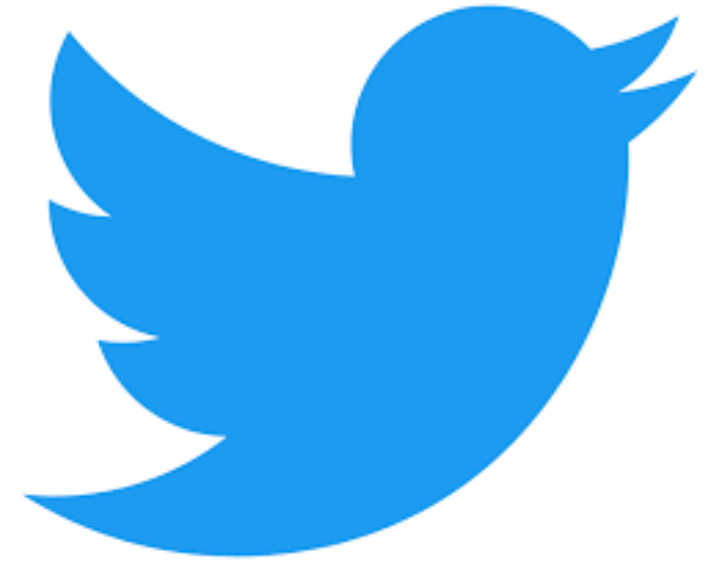
Cost-effectiveness



A few examples



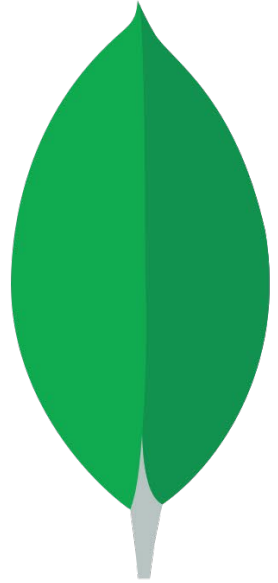
Uber



Challenges

- Lack of standardization
- Limited query capabilities
- Data consistency issues
- Higher learning curve
- Limited ecosystem





mongoDB®

MongoDB

MongoDB is a cross-platform, **document-oriented** NoSQL database. It is designed to store and manage **unstructured** or **semi-structured** data.

Key Features

- High Performance
 - Support for embedded data models reduces I/O activity on database system.
 - Indexes support faster queries and can include keys from embedded documents and arrays.
- Query API
 - Data Aggregation
 - Text Search and Geospatial Queries.

Key Features ...

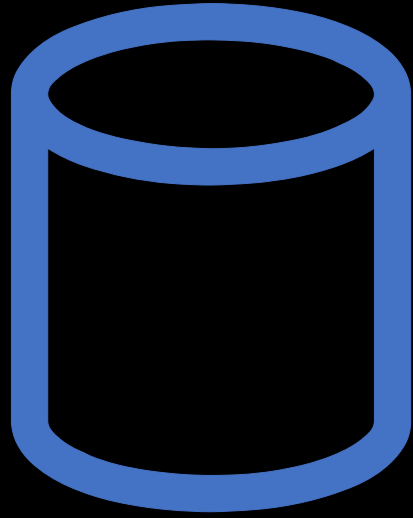
- High Availability through replica sets
- Horizontal Scalability through sharding
- Support for Multiple Storage Engines
 - WiredTiger Storage Engine
 - In-Memory Storage Engine
 - Pluggable storage engine API

```
{  
  name: "sue",  
  age: 26,  
  status: "A",  
  groups: [ "news", "sports" ]  
}
```

← field: value
← field: value
← field: value
← field: value

Document Database

Databases and Collections

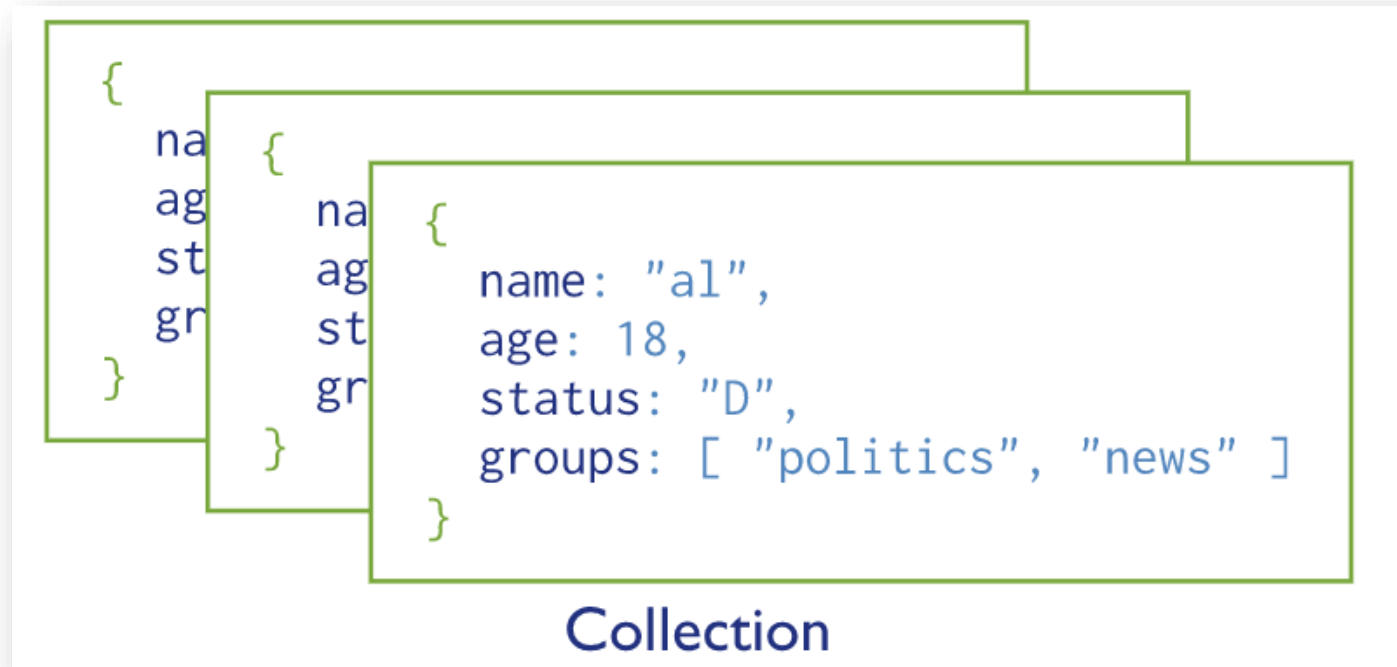


Databases

In MongoDB, databases hold one or more collections of documents.

Collections

MongoDB stores documents in collections. Collections are analogous to tables in relational databases.



Documents

- MongoDB stores data records as BSON documents. BSON is a binary representation of JSON documents, though it contains more data types than JSON.

```
{  
  name: "sue",  
  age: 26,  
  status: "A",  
  groups: [ "news", "sports" ]  
}
```



← field: value
← field: value
← field: value
← field: value

CRUD

- Create
- Read
- Update
- Delete

```
db.users.insertOne(  ← collection
{
  name: "sue",        ← field: value
  age: 26,             ← field: value
  status: "pending"   ← field: value } document
})
```

Create Operations

`db.collection.insertOne()`

`db.collection.insertMany()`


```
db.users.find(  
  { age: { $gt: 18 } },  
  { name: 1, address: 1 }  
) .limit(5)
```

← collection
← query criteria
← projection
← cursor modifier

`db.collection.find()`

Read Operations

```
db.users.deleteMany(  
  { status: "reject" }  
)
```



Delete Operations

- `db.collection.deleteOne()`
- `db.collection.deleteMany()`

```
db.users.updateMany(  
  { age: { $lt: 18 } },  
  { $set: { status: "reject" } }  
)
```



← collection
← update filter
← update action

Update Operations

db.collection.updateOne()

db.collection.updateMany()

db.collection.replaceOne()

Summary

- Defined NoSQL
- Compared NoSQL with SQL
- Reviewed
 - Use cases
 - Types of NoSQL
 - Technologies
 - Key features
 - Challenges
 - Guidelines to choose a right solution
- Introduced MongoDB



Next

In the workshop we will work on reading data from MongoDB

Keep reading and practicing as much as you can

Further reading

- NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, by Pramod J. Sadalage and Martin Fowler.
- MongoDB University. MongoDB offers a range of online courses and certifications that cover various aspects of MongoDB development, administration, and deployment. These courses are self-paced and are designed to help you learn at your own pace (<https://learn.mongodb.com/>).
- <https://www.mongodb.com/docs/manual/tutorial/query-documents/>