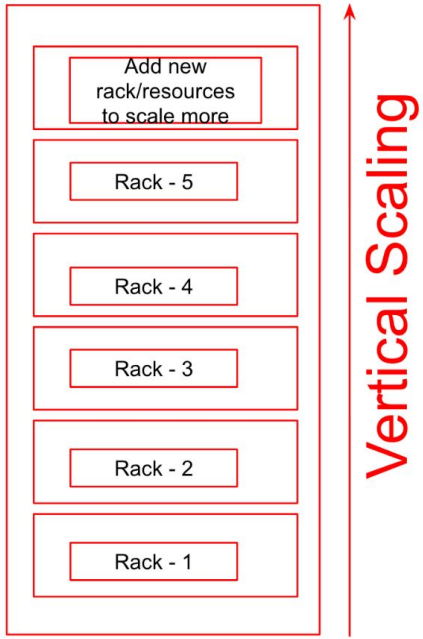


Vertical Scaling

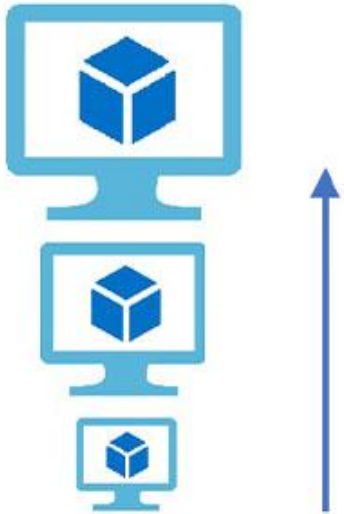
To scale more, Add more RAM, CPU, Memory to the **one existing machine**

Horizontal Scaling

To scale more: Add more machines to existing **group of distributed system**

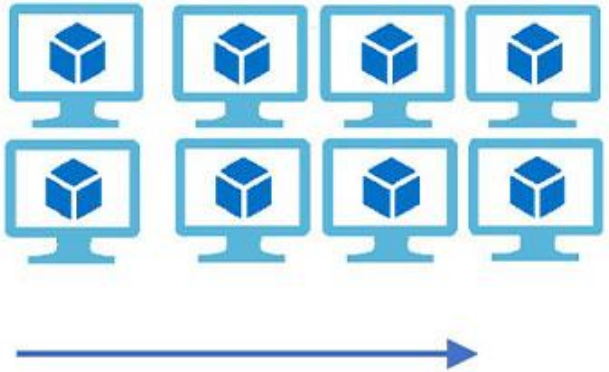
Vertical Scaling

(Increase size of instance (RAM , CPU etc.))



Horizontal Scaling

(Add more instances)



NoSQL Databases

NoSQL


Gaming


Social


IoT


Web


Mobile


Enterprise

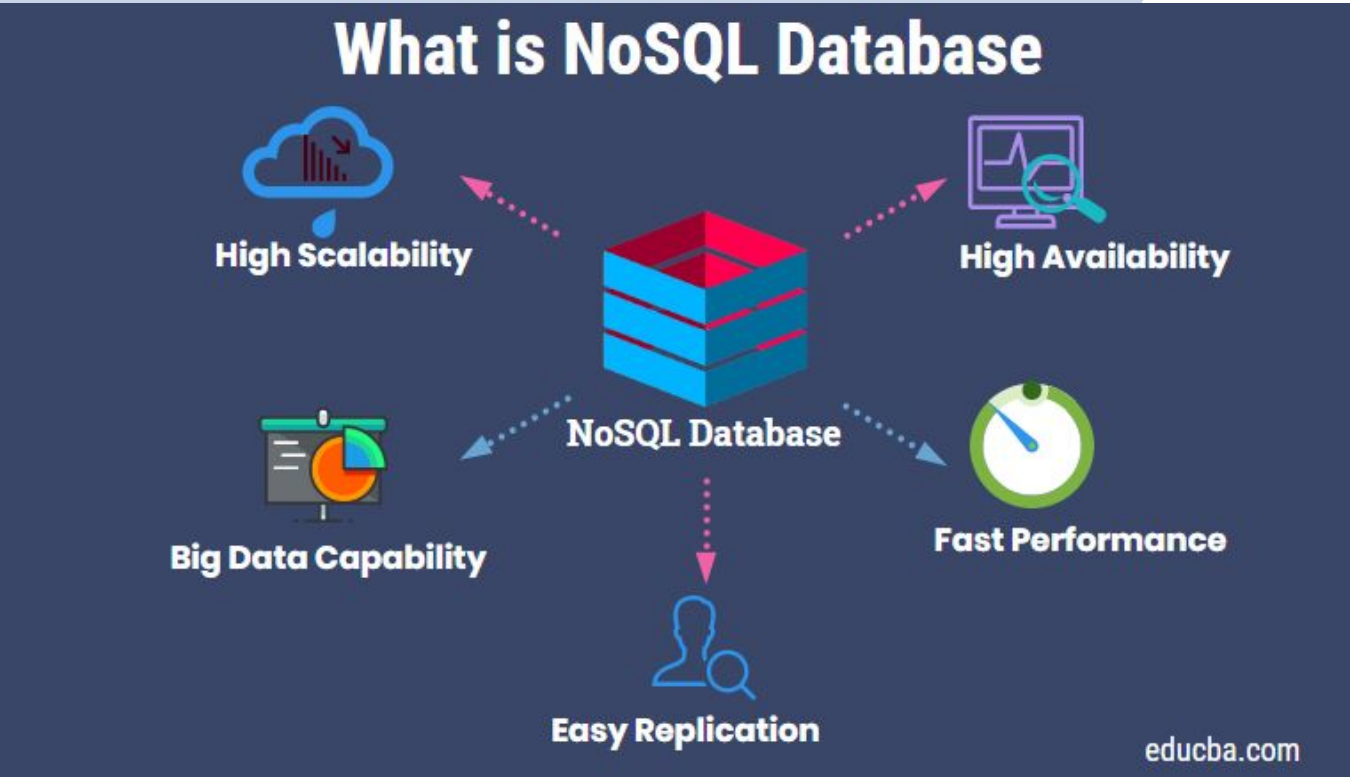
SQL


Web


Mobile


Enterprise


Data mart


What is NoSQL Database
High Scalability
High Availability
NoSQL Database
Big Data Capability
Fast Performance
Easy Replication
educba.com

SQL VS NoSQL

SQL

Relational Data Model

Pros
- Easy to use and setup.
- Universal, compatible with many tools.
- Good at high-performance workloads.
- Good at structure data.

Cons
- Time consuming to understand and design the structure of the database.
- Can be difficult to scale.

No SQL

Document Data Model

Pros
- No investment to design model.
- Rapid development cycles.
- In general faster than SQL.
- Runs well on the cloud.

Cons
- Unsuited for interconnected data.
- Technology still maturing.
- Can have slower response time.

DATA MODELS

Unstructured vs Structured Data

Structured Data

Often numbers or labels, stored in a structured framework of columns and rows relating to pre-set parameters.

- ID CODES IN DATABASES
- NUMERICAL DATA GOOGLE SHEETS
- STAR RATINGS

Semi-unstructured Data

Loosely organized into categories using meta tags

- EMAILS BY INBOX, SENT, DRAFT
- TWEETS ORGANIZED BY HASHTAGS
- FOLDERS ORGANIZED BY TOPIC

Unstructured Data

Text-heavy information that's not organized in a clearly defined framework or model.

- MEDIA POSTS, EMAILS, ONLINE REVIEWS
- VIDEOS, IMAGES
- SPEECH, SOUNDS

UNSTRUCTURED DATA

Human Generated

- Text files: Word processing, spreadsheets, presentations, emails, logs.
- Social Media: Data from Facebook, Twitter, LinkedIn.
- Website: YouTube, Instagram, photo sharing sites.
- Mobile data: Text messages, locations.
- Communications: Chat, IM, phone recordings, collaboration software.
- Media: MP3, digital photos, audio and video files.
- Business applications: MS Office documents

Machine Generated

- Satellite imagery: Weather data, landforms, military movements.
- Scientific data: Oil and gas exploration, space exploration, seismic imagery, atmospheric data.
- Digital surveillance: Surveillance photos and video.
- Sensor data: Traffic, weather, oceanographic sensors.

STRUCTURED DATA

A table in a relational database such as SQL, ORACLE, MySQL is an example of structured data.

CUSTOMER

CUSTOMER_ID	LAST_NAME	FIRST_NAME	STREET	CITY	ZIP_CODE	COUNTRY
10302	Boucher	Leo	54, rue Royale	Nantes	44000	France
11244	Smith	Laurent	8489 Strong St	Las Vegas	83030	USA
11405	Han	James	636 St Kilda Road	Sydney	3004	Australia
11993	Mueller	Tomas	Berliner Weg 15	Tamm	71732	Germany
12111	Carter	Nataly	5 Tomahawk	Los Angeles	90006	USA
14121	Cortez	Nola	Av. Grande, 86	Madrid	28034	Spain
14400	Brown	Frank	165 S 7th St	Chester	33134	USA
14578	Wilson	Sarah	Seestreet #6101	Emory	1734	USA
14622	Jones	John	71 San Diego Ave	Arlington	69004	USA

SEMI-STRUCTURED DATA

Data with some degree of organization.

Examples:

- Hypertext Markup Language (HTML) files
- JavaScript Object Notation (JSON) files
- Extensible Markup Language (XML) files

Needs some pre-processing before it can be analysed by a computer

Need something different for semi-structured data, this is where NoSQL databases are being popular.

NoSQL

NoSQL

Non-tabular database.

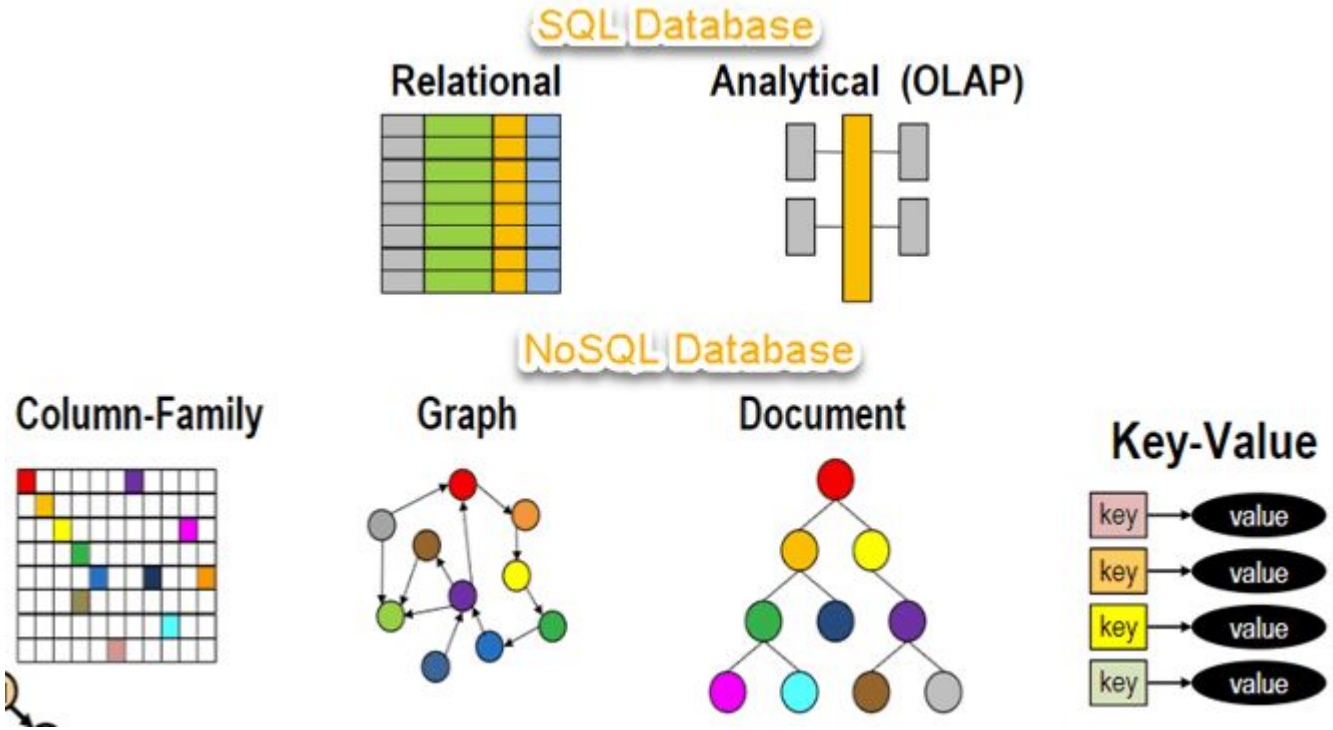
Used to manage variety of data types in big data.

Example: MongoDB, Cassandra and HIVE

Features:

- Flexible schemas
- Horizontal scaling
- Fast queries due to the data model
- Ease of use for developers

TYPES OF NoSQL DATABASES



TYPES OF NoSQL DATABASES

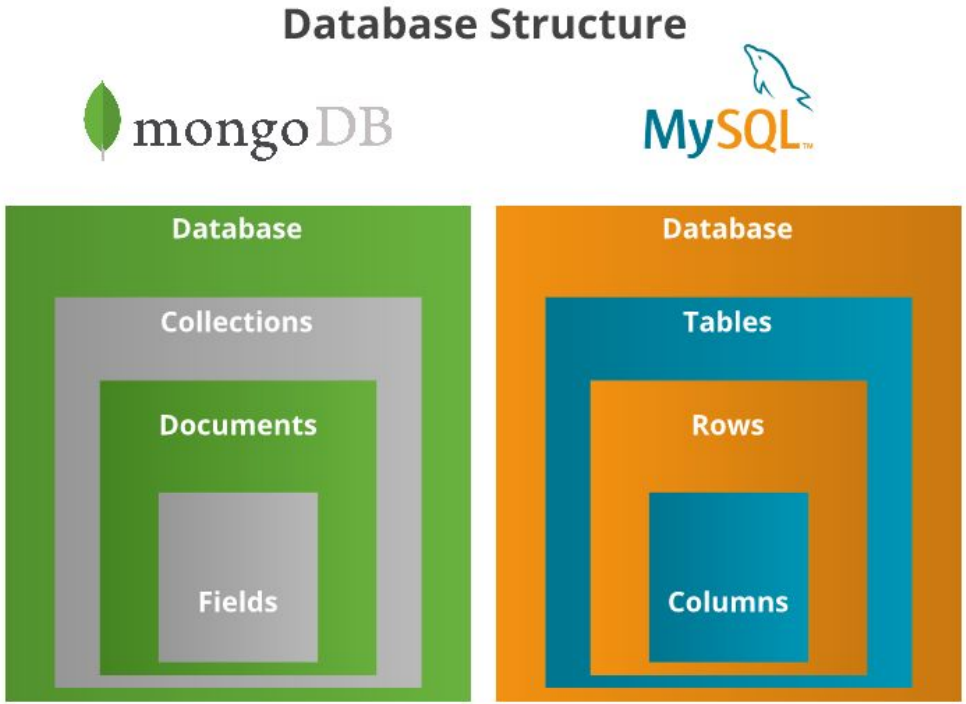
Document Database	Graph Databases
Couchbase, MarkLogic, mongoDB	Neo4j, InfiniteGraph The Distributed Graph Database
Wide Column Stores	**Key-Value Databases**
redis, amazon DynamoDB, AEROSPIKE, riak	accumulo, HYPERTABLE INC, Cassandra, APACHE HBASE, Amazon SimpleDB

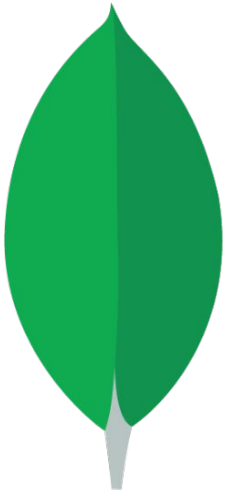
@cloudtxt http://www.aryannava.com

TYPES OF NoSQL DATABASES

1. **Document databases** store data in documents similar to JSON (JavaScript Object Notation) objects. Each document contains pairs of fields and values. The values can typically be a variety of types including things like strings, numbers, booleans, arrays, or objects.
2. **Key-value databases** are a simpler type of database where each item contains keys and values.
3. **Wide-column stores** store data in tables, rows, and dynamic columns.
4. **Graph databases** store data in nodes and edges. Nodes typically store information about people, places, and things, while edges store information about the relationships between the nodes.

DOCUMENT VS RELATIONAL DATABASE





mongoDB®

MongoDB

- MongoDB stores data records as documents (specifically BSON documents) which are gathered together in collections.
- A database stores one or more collections of documents.
- MongoDB is a distributed database at its core.
- Follows concept of Collections and Documents.
  - Database
    - Collection
      - Document

MongoDB- Collection

- A collection is a grouping of MongoDB documents.
- A collection is the equivalent of a table in a relational database system.
- If a collection does not exist, MongoDB will

automatically create the collection when you

first store data for that collection.

MongoDB- Document

- Data is stored as BSON documents in a collection.
  - BSON is a binary representation of JSON documents.
  - JSON is built on two structures:
    - A collection of name/value pairs
    - An ordered list of values, such as an array

MongoDB creates a unique object id **“_id”** for objects in the database.

MongoDB: Object IDs

ObjectIds are small, likely unique, fast to generate, and ordered. ObjectId values are 12 bytes in length, consisting of:

- a 4-byte timestamp value, representing the ObjectID's creation, measured in seconds since the Unix epoch
- a 5-byte random value generated once per process. This random value is unique to the machine and process.
- a 3-byte incrementing counter, initialized to a random value

MongoDB- Object IDs

In MongoDB, each document stored in a collection requires a unique _id field that acts as a primary key.

If an inserted document omits the _id field, the MongoDB driver automatically generates an ObjectId for the _id field.

MongoDB- Queries

1. **Create Operations**: db.collection.insertOne(), db.collection.insertMany()
2. **Update Operations**: db.collection.updateOne(), db.collection.updateMany(), db.collection.replaceOne()
3. **Delete Operations**: db.collection.deleteOne(), db.collection.deleteMany()

MongoDB- Queries

Collections can be queried using the **find()** method.

The basic syntax is:

db.collectionName.find()

This will return the data in an unstructured way.

To improve this, use the **pretty()** method:

db.collectionName.find().pretty()

find() is similar to the WHERE part of a SQL statement

Basic Guide of starting MongoDB will be covered in Workshop along with Basic CRUD operations.

THANKS!

Any questions?

You can find me at

LinkedIn@poudelyamu

yamu.poudel@heraldcollege.edu.np