edureka!

MODULE-1
DESIGN GOALS, ARCHITECTURE AND INSTALLATION

How it Works?

edureka!

Experienced Instructor

Live Online Class

In-class Questions

Survey Feedback

24x7 Support

Class Recording in LMS

Module Wise Assessment

Project Work

Verifiable Certificate

Android & iOS App

Course Topics

edureka!

- → Module 1
  - » **Design Goals, Architecture and Installation**
- → Module 2
  - » CRUD Operations
- → Module 3
  - » Schema Design and Data Modelling
- → Module 4
  - » Administration
- → Module 5
  - » Scalability and Availability
- → Module 6
  - » Indexing and Aggregation Framework
- → Module 7
  - » Application Engineering and MongoDB Tools
- → Module 8
  - » Project, Additional Concepts and Case Studies

Objectives

edureka!

At the end of this Module, you will be able to understand

→Database Categories

→Mongo DB Overview

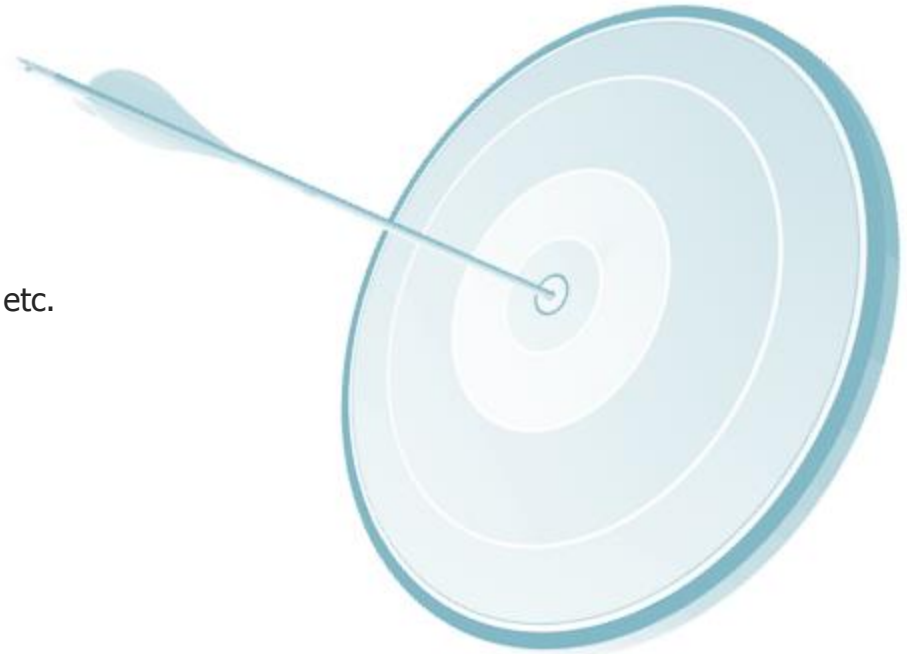
→Design Goals for MongoDB Server and Database

→Mongo DB Tools

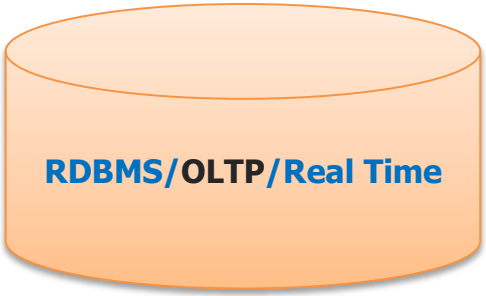
→Introduction to JSON and BSON

→Installation of MongoDB on Windows, Linux, MAC OS etc.

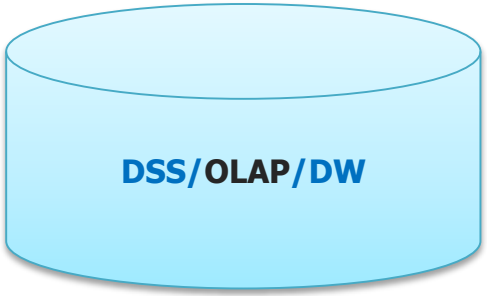
→Environment Setup for MongoDB



Database Categories



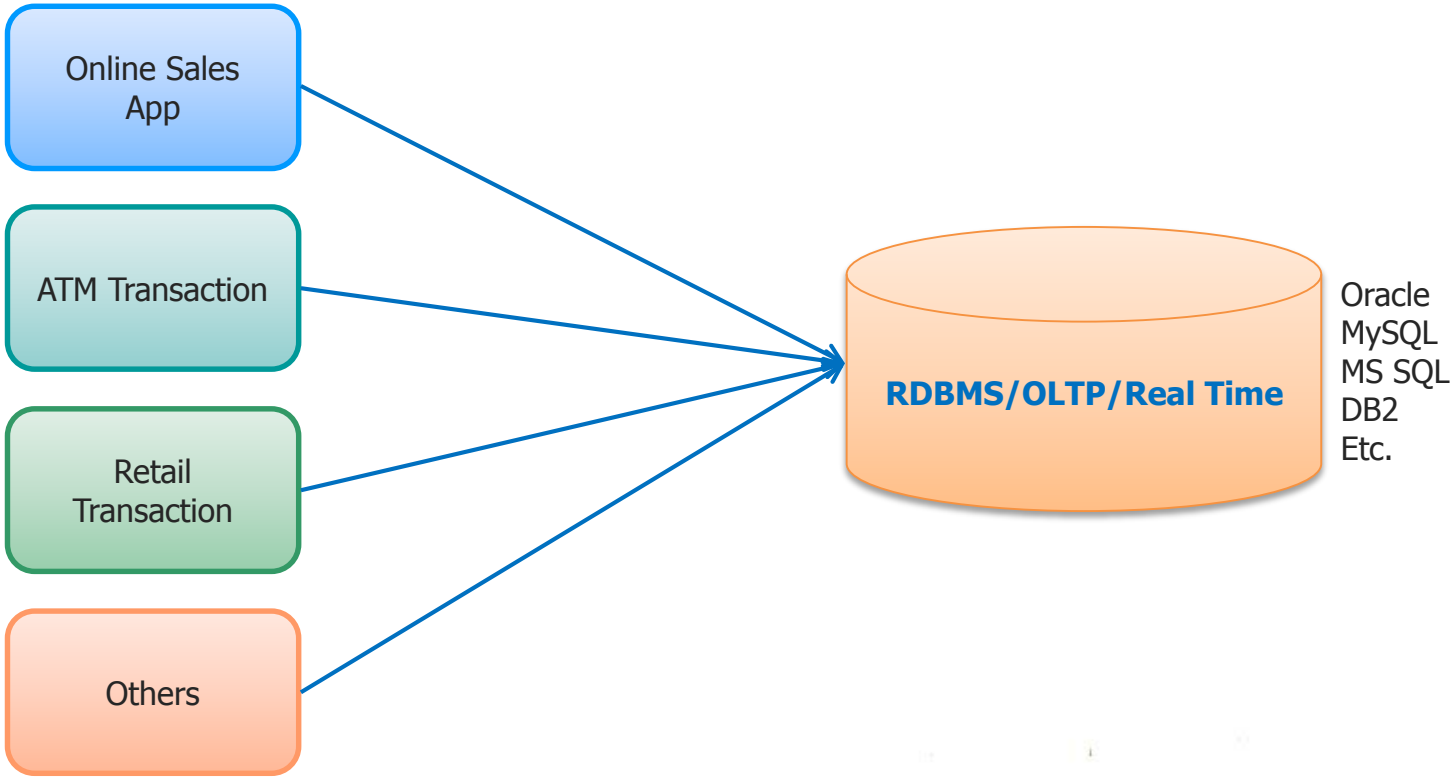
Oracle
MySQL
MS SQL
DB2



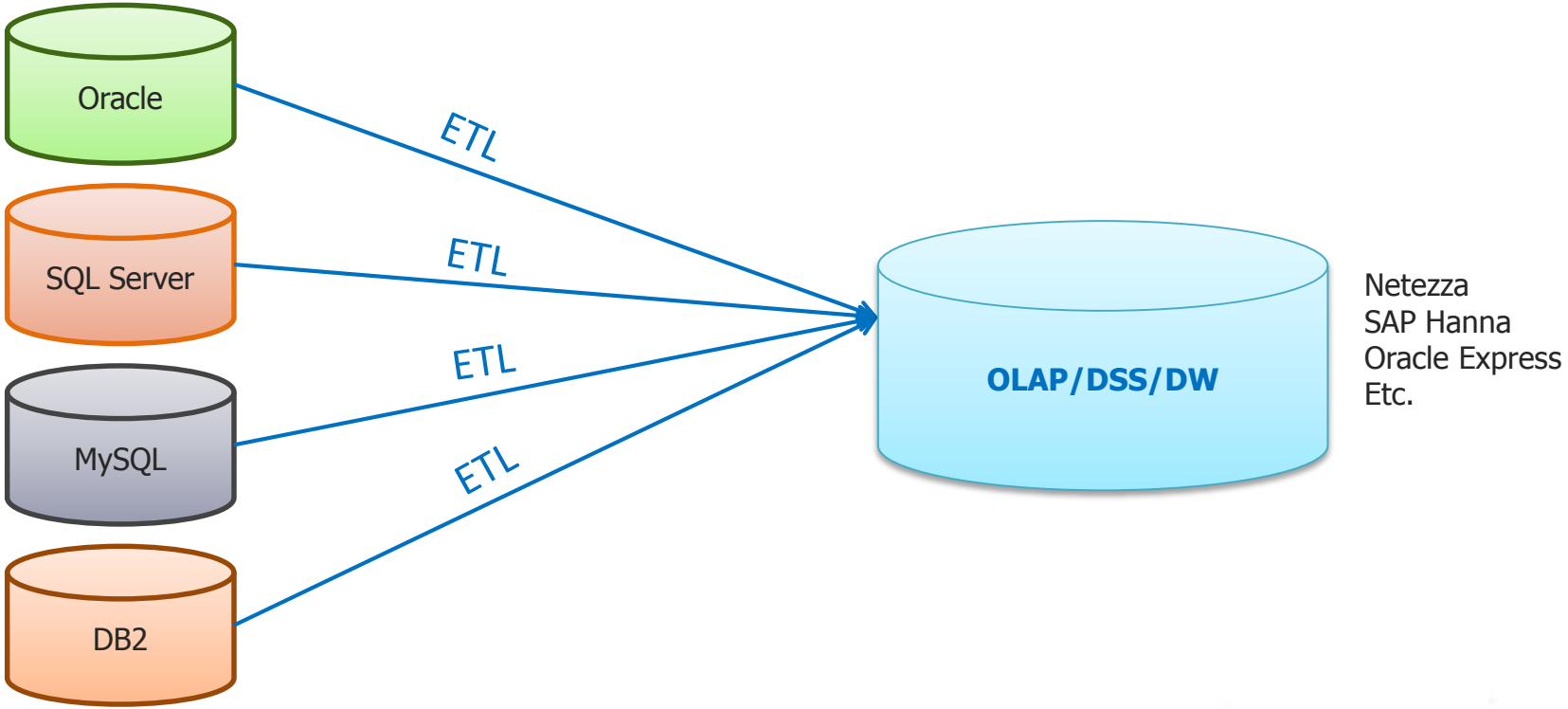
Netezza
SAP Hana
Oracle Express

MongoDB
Hbase
Cassandra
couchDB

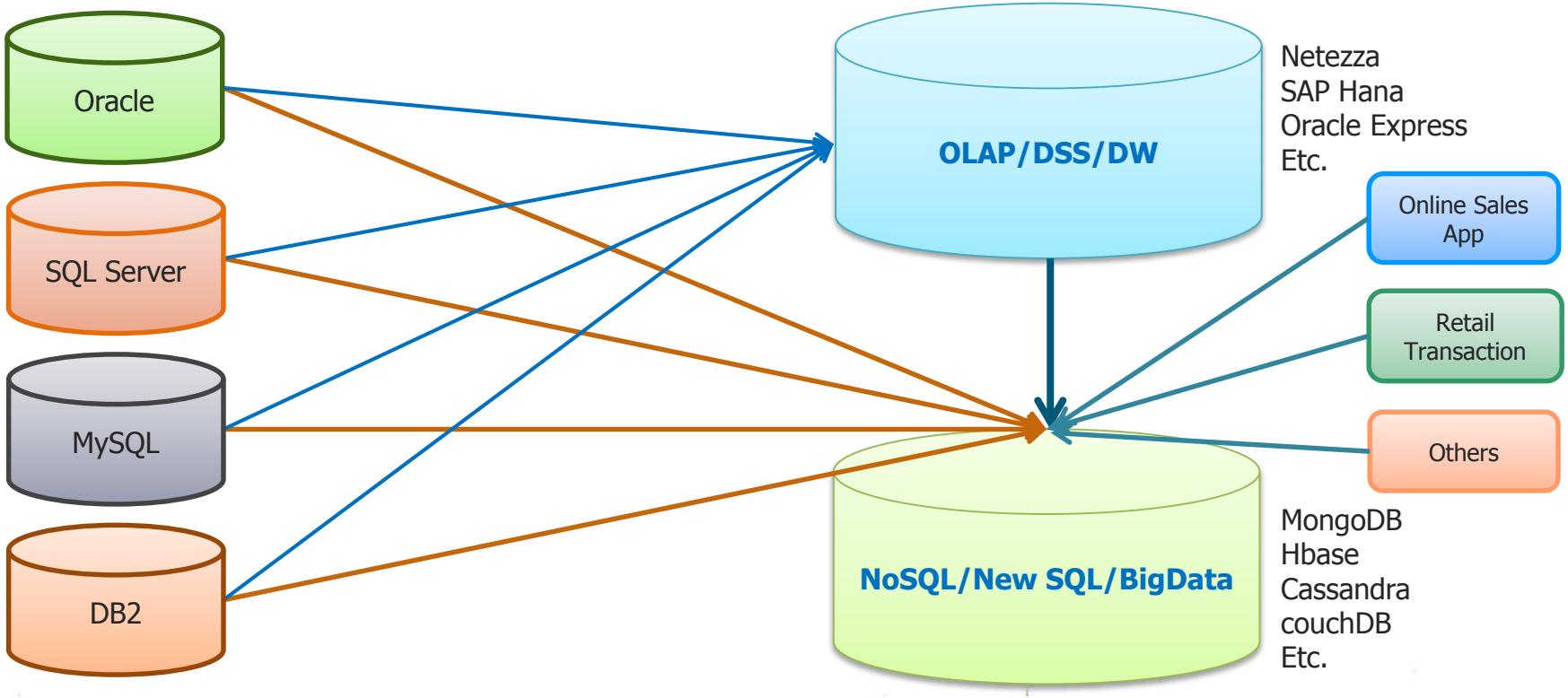
OLTP Database



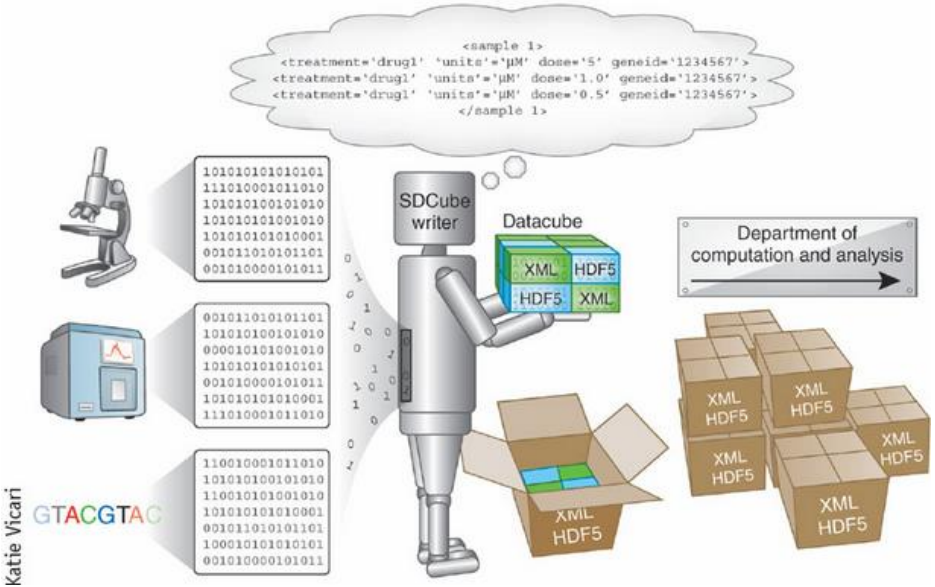
OLTP Database (Contd.)



NoSQL Databses



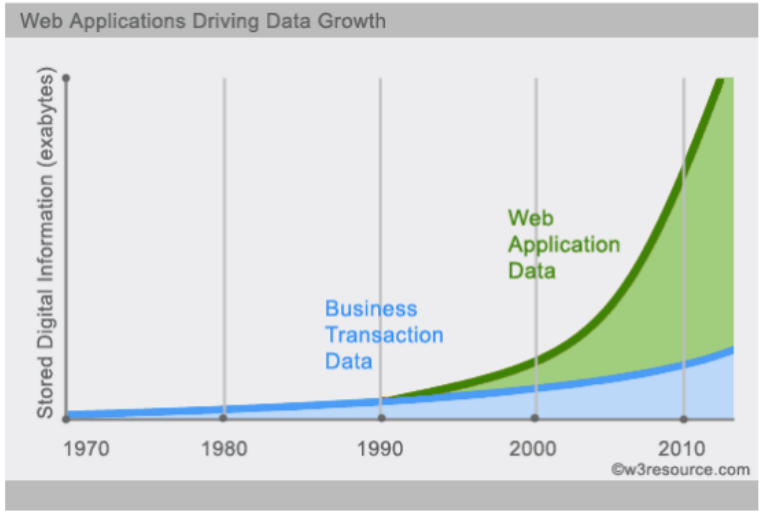
Why NoSQL?

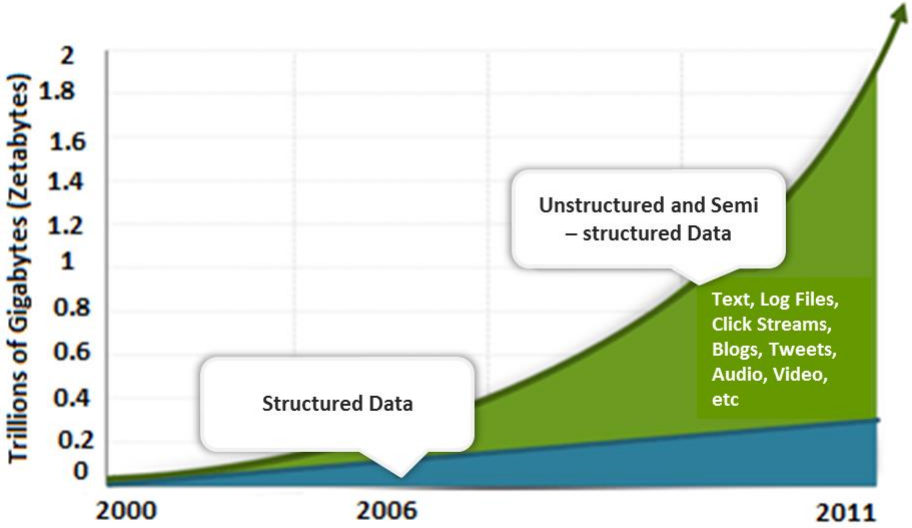


Nature of Data

Application Development (high coding velocity & agility)

Why NoSQL? (Contd.)





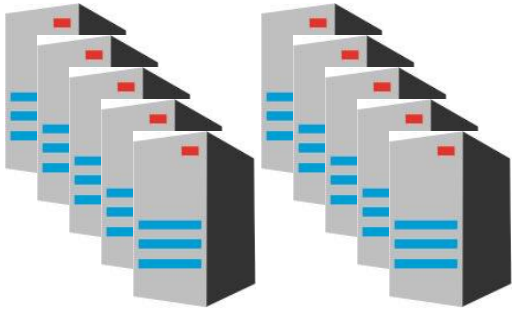
Data Warehousing and Analytics

Source:IDC2011 Digital Universe Study [http://www.emc.com/collateral/demos/microsites/emc-digital-univere-2011index.html]

Distributed Architecture at Its Core

edureka!

Read 1 TB Data



1 Machine

4 I/O Channels
Each Channel – 100 MB/s

10 Machine

4 I/O Channels
Each Channel – 100 MB/s

Distributed Architecture at Its Core (Contd.)

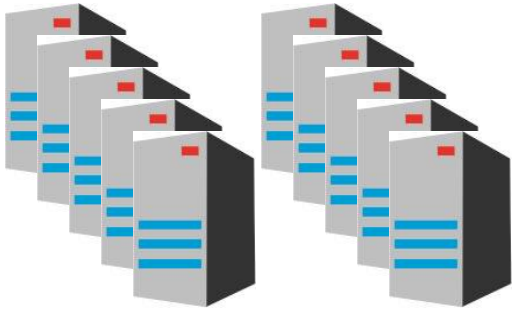
edureka!

Read 1 TB Data

1 Machine

4 I/O Channels
Each Channel – 100 MB/s

43.69 Minutes



10 Machine

4 I/O Channels
Each Channel – 100 MB/s

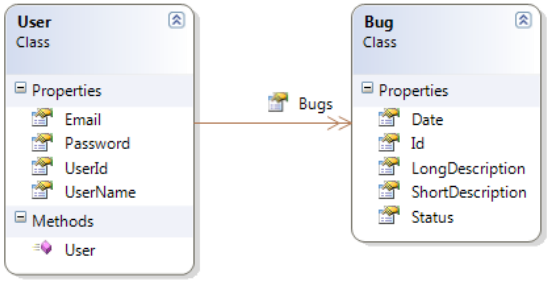
4.36 Minutes

What is NoSQL?

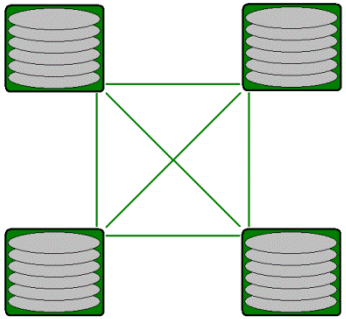
edureka!

Next Generation Databases

Not Only SQL

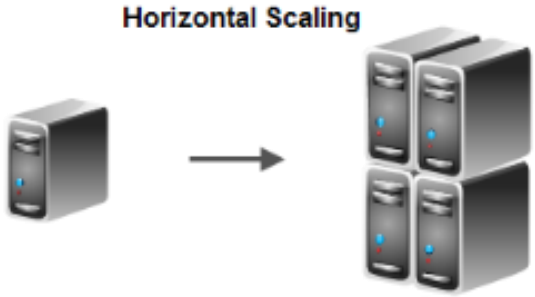


Non – Relational



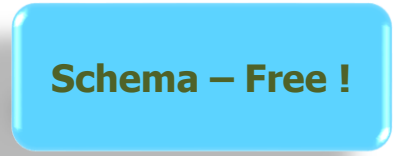
Distributed Architecture

Open Source



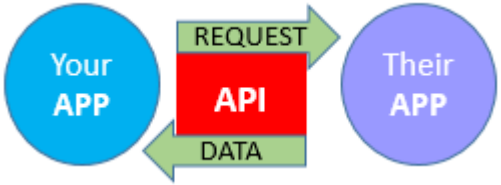
Horizontally Scalable

What is NoSQL? (Contd.)



Schema - Free

Easy – Replication



Simple API

Can Manage Huge Amount of Data

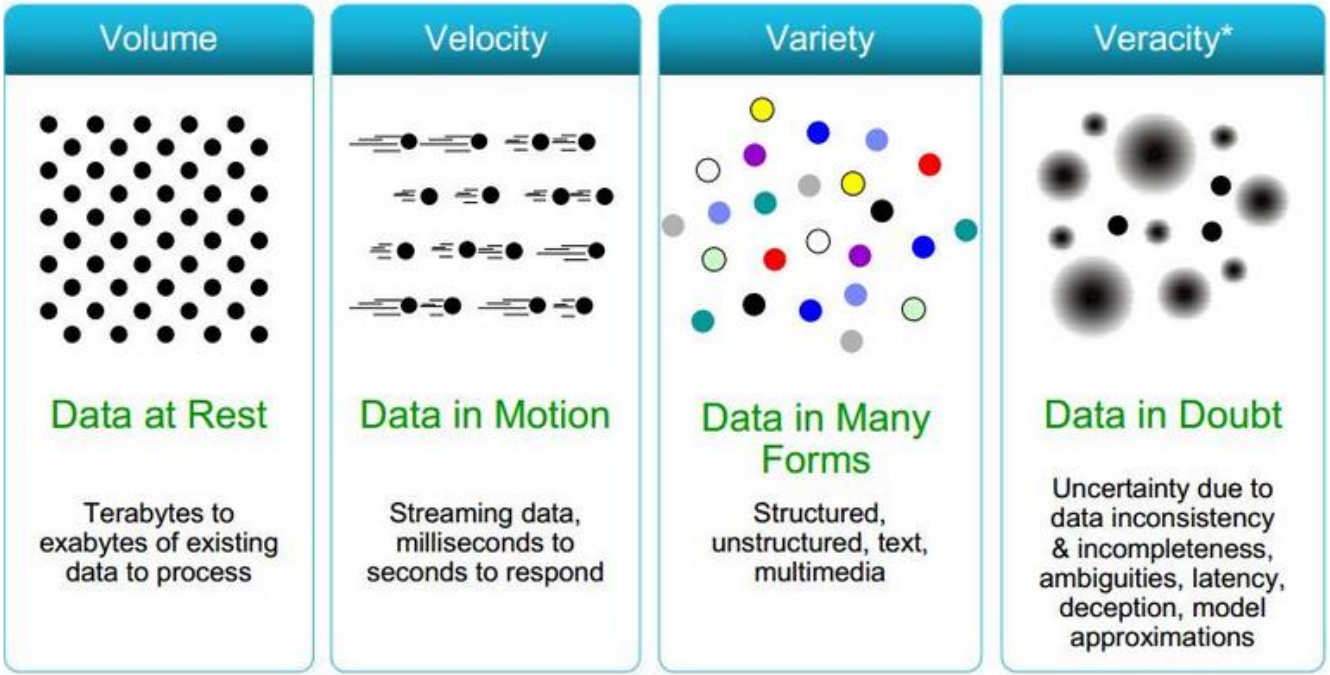
Can be implemented on Commodity Hardware's

~ 150 No SQL Database are there in Market

Benefit of NoSQL

edureka!

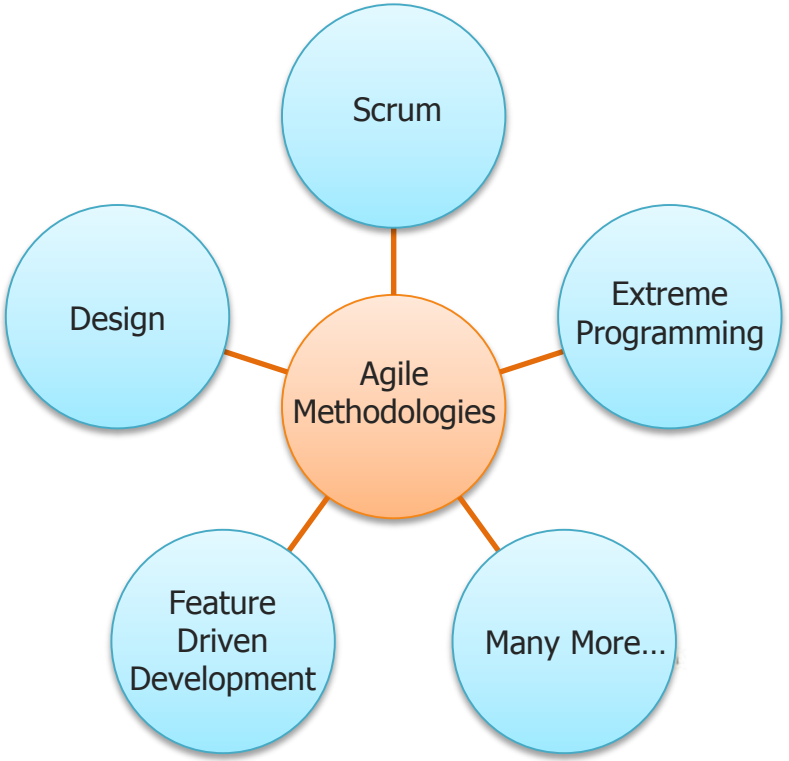
→ Large volumes of structured, semi-structured, and unstructured data



Benefit of NoSQL

edureka!

→Agile development, quick changes, and frequent code pushes

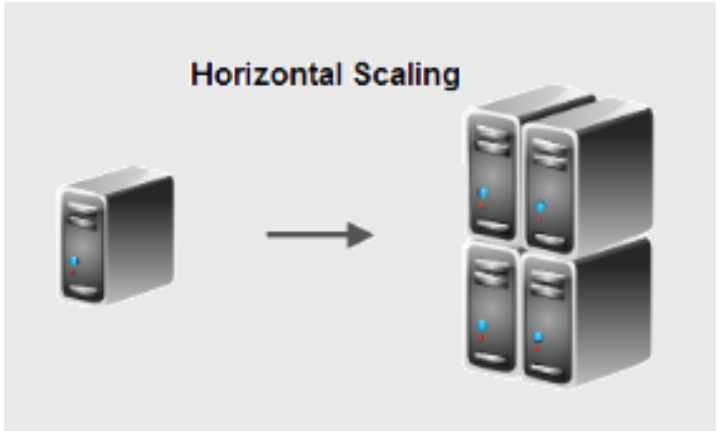


Benefit of NoSQL

edureka!

→ Object-oriented programming that is easy to use and flexible

→ Horizontal scaling instead of expensive hardware



Categories of NoSQL Database

edureka!

Document Base

→ Document databases pair each key with a complex data structure known as a document.

→ Documents can contain many different key-value pairs, or key-array pairs, or even nested documents.

Key – value Stores

→ Key-value stores are the simplest NoSQL databases.

→ Every single item in the database is stored as an attribute name (or "key"), together with its value.

Graph Store

→ Graph stores are used to store information about networks, such as social connections.

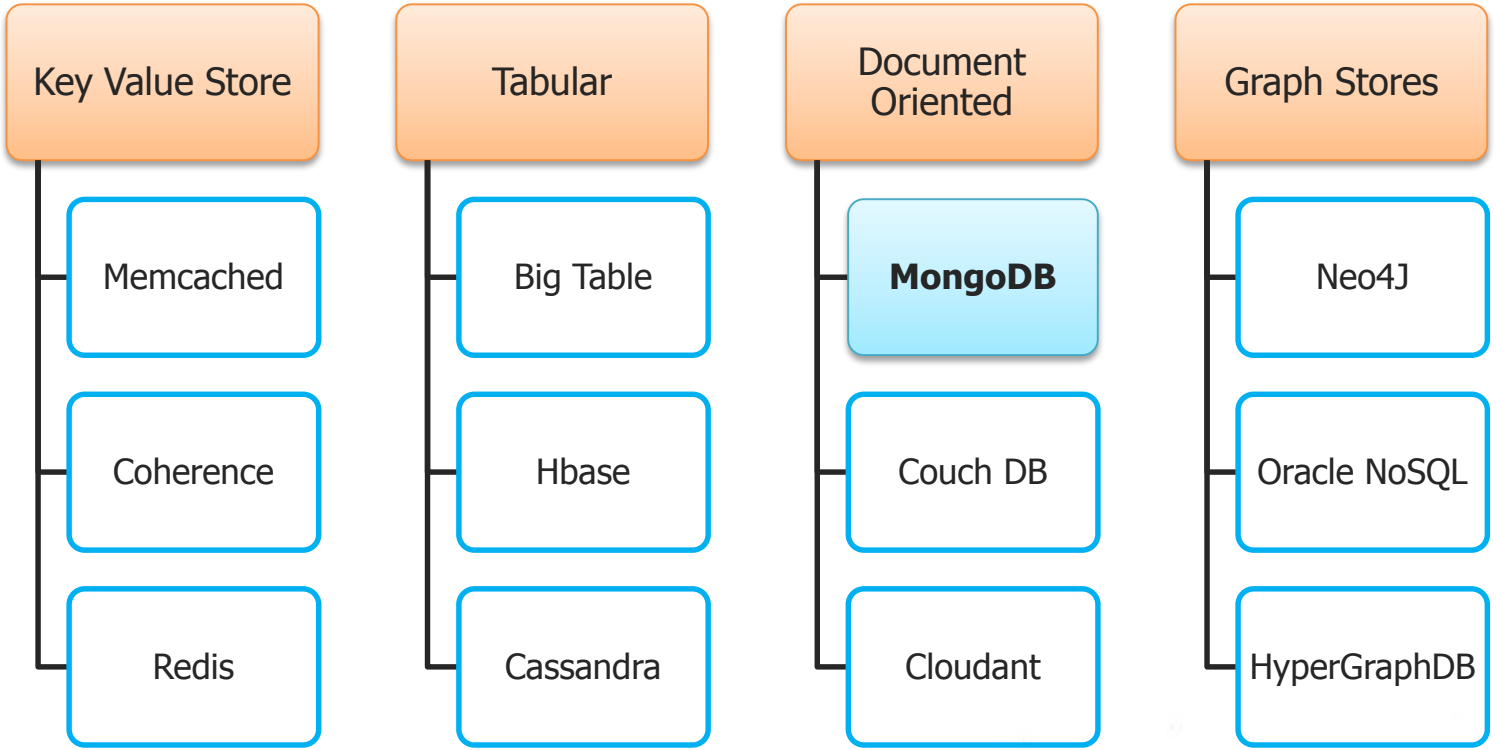
→ Graph stores include Neo4J and HyperGraphDB.

Wide Column Stores

→ Wide-column stores such as Cassandra and HBase are optimized for queries over large datasets, and store columns of data together, instead of rows.

Type of No SQL Databases

edureka!



NoSQL vs. SQL Comparison

edureka!

Entity	SQL Databases	NoSQL Databases
Type	One Type (SQL) with Minor Variation	Many Types (Document, Key-Value, Tabular, Graph)
Development	1970	2000
Examples	Oracle, MSSQL, DB2 etc.	MongoDB, Cassandra, Hbase, Neo4J
Schemas	Fixed	Dynamic
Scaling	Vertical	Horizontal
Dev Model	Mix	Open Source
Consistency	Follow ACID	Follow BASE

SQL Databases (ACID Property)

edureka!

Atomic
→ A transaction is a logical unit of work which must be either completed with all of its data modifications, or none of them is performed.

Consistent
→ At the end of the transaction, all data must be left in a consistent state.

ACID Property

Isolated

Isolated
→ Modifications of data performed by a transaction must be independent of another transaction. Unless this happens, the outcome of a transaction may be erroneous.

Durable
→ When the transaction is completed, effects of the modifications performed by the transaction must be permanent in the system.

Durable

Cap Theorem

edureka!

CAP theorem states that there are 3 basic requirements which exist in a special relation when designing applications for a distributed architecture.

Consistency

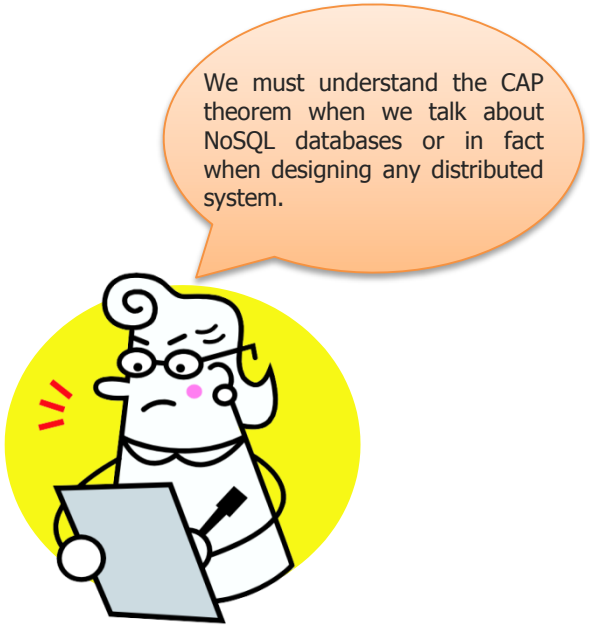
This means that the data in the database remains consistent after the execution of an operation. For example after an update operation all clients see the same data.

Availability

This means that the system is always on (service guarantee availability), no downtime.

Partition Tolerance

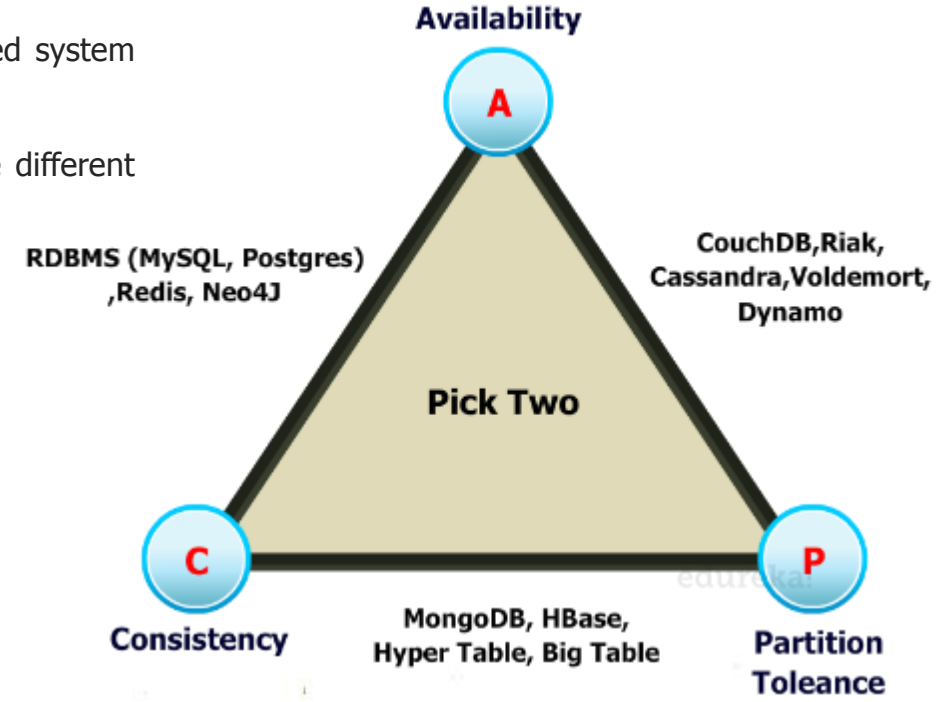
This means that the system continues to function even the communication among the servers is unreliable, i.e. the servers may be partitioned into multiple groups that cannot communicate with one another.



Cap Theorem (Contd.)

edureka!

- → In theoretically it is impossible to fulfill all 3 requirements.
- → CAP provides the basic requirements for a distributed system to follow 2 of the 3 requirements.
- → Therefore all the current NoSQL database follow the different combinations of the C, A, P from the CAP theorem.



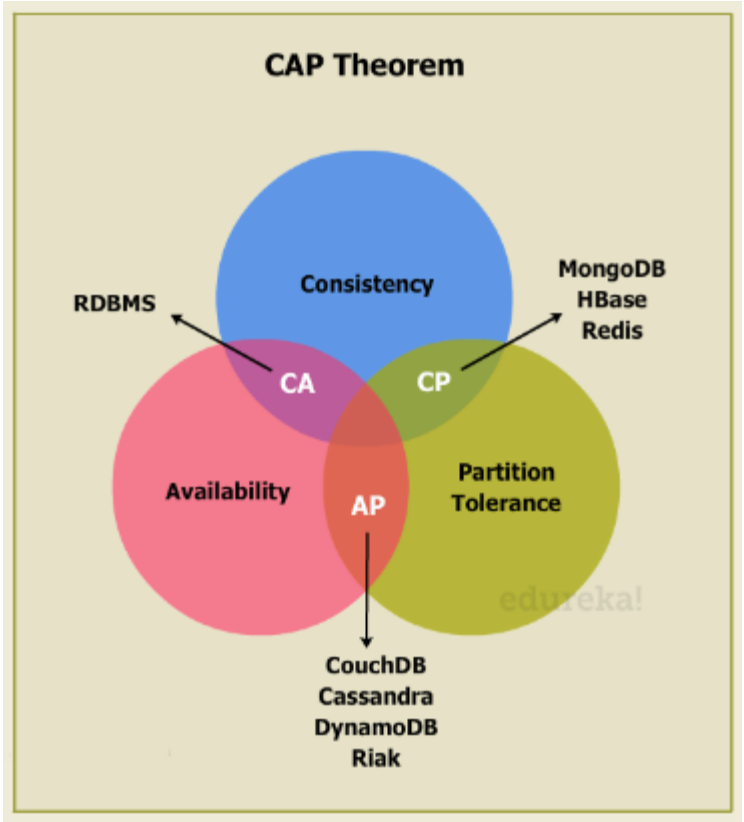
Cap Theorem (Contd.)

Here is the brief description of three combinations CA, CP, AP :

→ CA - Single site cluster, therefore all nodes are always in contact. When a partition occurs, the system blocks.

→ CP - Some data may not be accessible, but the rest is still consistent/accurate.

→ AP - System is still available under partitioning, but some of the data returned may be inaccurate.



BASE

A BASE system gives up on consistency.

→ Basically Available

→Basically Available indicates that the system does guarantee availability, in terms of the CAP theorem.

→ Soft State

→Soft State indicates that the state of the system may change over time, even without input. This is because of the eventual consistency model.

→ Eventual Consistency

→Eventual Consistency indicates that the system will become consistent over time, given that the system doesn't receive input during that time.

Annie’s Introduction

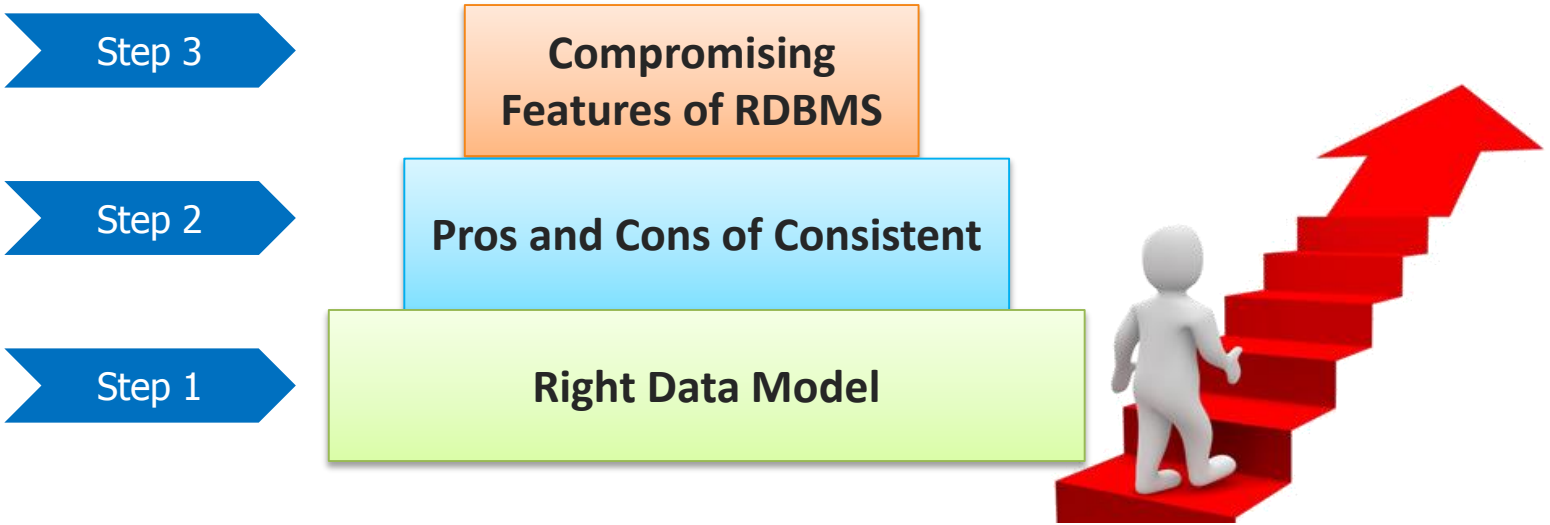
Annie's Question

Map the following to corresponding data bases:
MongoDB
Neo4J
Cassandra
Memcached

Annie's Answer

Ans.
MongoDB →Document Oriented Database
Neo4J → Graph Database
Cassandra → Columnar Database
Memcached →Key-value store

Before Selection and Implementation of NoSQL



Annie's Question

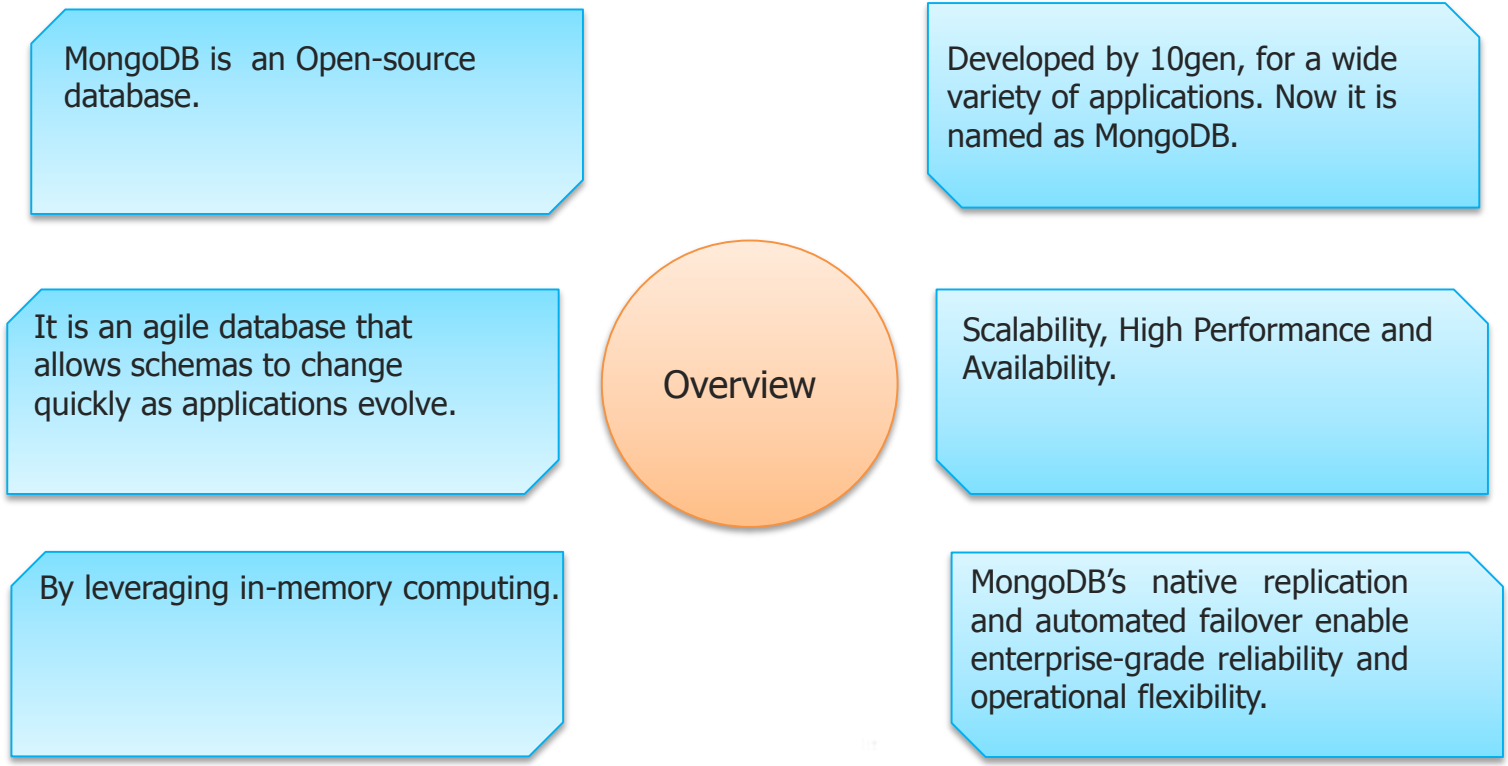
Which concept is followed by NoSQL, choose from below list
1→ ACID
2→ CAP
3→ BASE

Annie’s Answer

Ans.
BASE

MongoDB Overview

MongoDB Overview



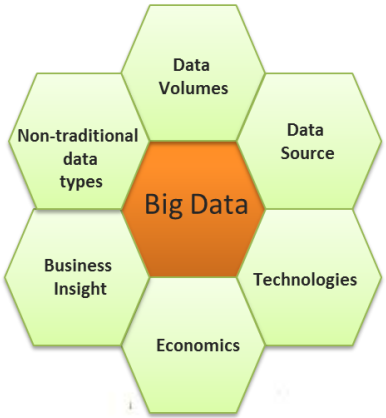
New World

New Apps

New Development Methods

New Data Volumes

New Architectures



New Data Types

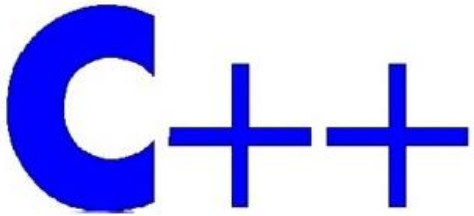
What is MongoDB?

edureka!

Open Source

Document Storage

Object Oriented

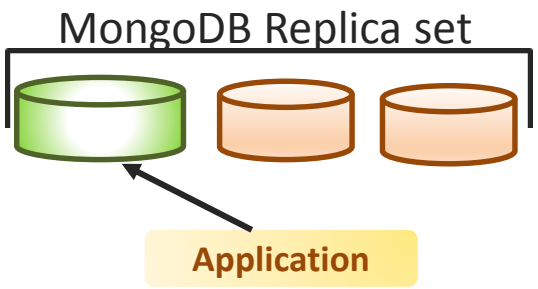


Written in C++

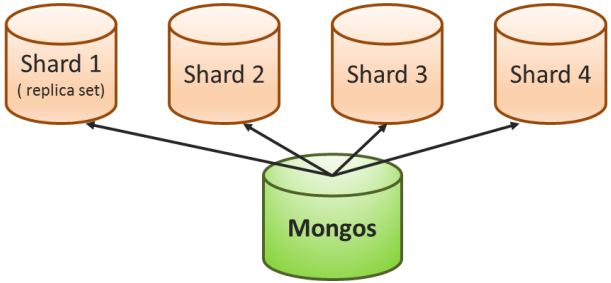
Easy to Use

Full Index Support

What is MongoDB? (Contd.)

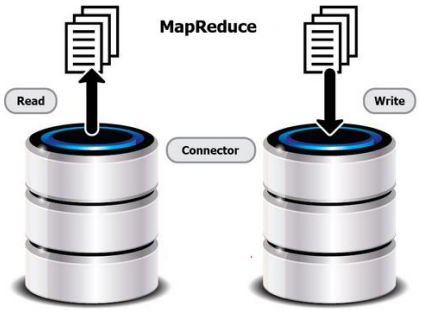


Replication and High Availability



Auto Sharding

Easy Query

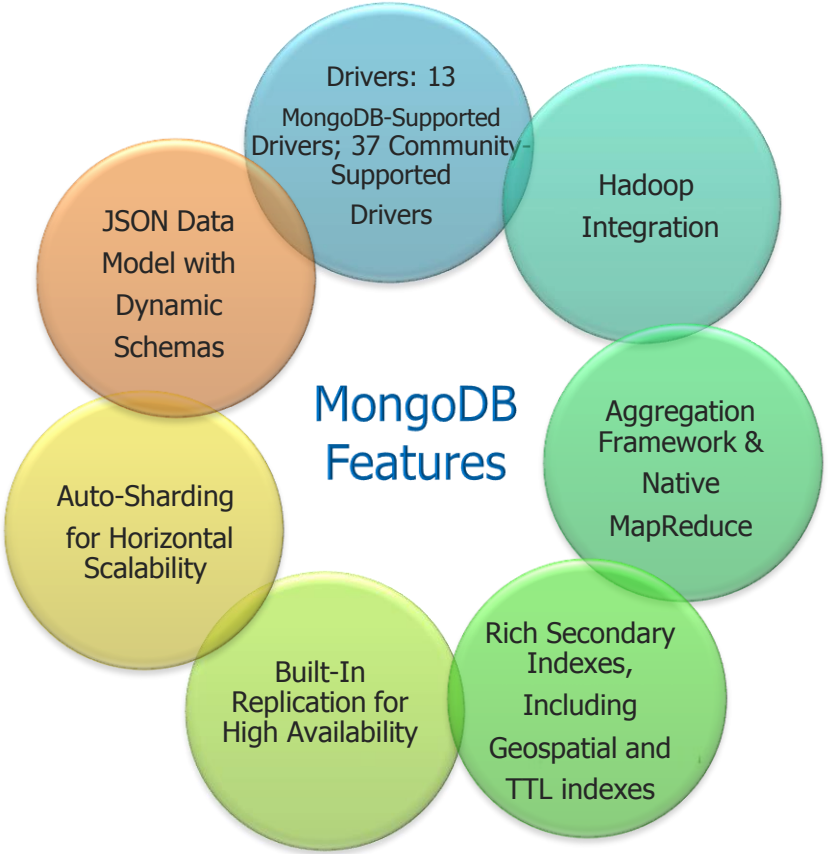


Map Reduce

Gird FS

Support from Expert

MongoDB Features



Annie's Question

Which kind of data can be processed with MongoDB, choose from below option
1→ Online Data
2→ Offline Data
3→ Both

Annie's Answer

Ans. Both

Annie's Question

Can you give example of Big Data?

Annie's Answer

Ans.

1. Facebook ingests 500 terabytes of new data every day.
2. A Boeing 737 will generate 240 terabytes of flight data during a single flight across the US.

Few MongoDB Clients

US Department of Veterans Affairs

US Department of Energy's Berkley Lab and MIT

Few MongoDB Clients (Contd.)

edureka!

→ Metlife uses MongoDB for "The Wall" an innovative customer service application provides a 360-degree, consolidated view of MetLife customers, including policy details and transactions across lines of business.

→ ebay has a number of projects running on MongoDB for search suggestions, metadata storage, cloud management and merchandizing categorization.

→ MongoDB is the repository that powers MTV Networks' next-generation CMS, which is used to manage and distribute content for all of MTV Networks' major websites.

→ MongoDB is used for back-end storage on the SourceForge front pages, project pages, and download pages for all projects.

→ Craigslist uses MongoDB to archive billions of records.

→ ADP uses MongoDB for its high performance, scalability, reliability and its ability to preserve the data manipulation capabilities of traditional relational databases.

Few MongoDB Clients (Contd.)

edureka!

→ CNN Turk uses MongoDB for its infrastructure and content management system, including the tv.cnnturk.com.

→ Foursquare uses MongoDB to store venues and user 'check-ins' into venues, sharding the data over more than 25 machines on Amazon EC2.

→ Justin.tv is the easy, fun, and fast way to share live video online. MongoDB powers Justin.tv's internal analytics tools for virality, user retention, and general usage stats that out-of-the-box solutions can't provide.

→ ibibo ('I build, I bond') is a social network using MongoDB for its dashboard feeds. Each feed is represented as a single document containing an average of 1000 entries; the site currently stores over two million of these documents in MongoDB.

Industry /Domains Where MongoDB is Used

Government

Financial Services

Healthcare

Media and Entertainment

Tele-communications

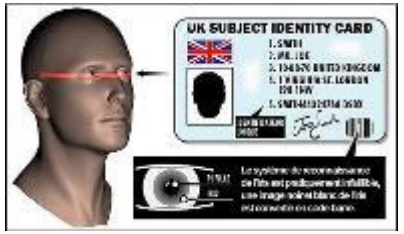
Retail

Financial Services

- → Risk Analytics and Reporting
- → Reference Data Management
- → Market Data Management
- → Portfolio Management
- → Order Capture
- → Time Series Data

Government

- → Surveillance Data Aggregation
- → Crime Data Management and Analytics
- → Citizen Engagement Platform
- → Program Data Management
- → Healthcare Record Management



Health Care

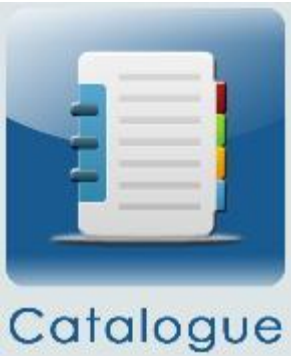
- → 360-Degree Patient View
- → Population Management for At-Risk Demographics
- → Lab Data Management and Analytics
- → Mobile Apps for Doctors and Nurses
- → Electronic Healthcare Records (EHR)

Media and Entertainment

- → Content Management and Delivery
- → User Data Management
- → Digital Asset Management
- → Mobile and Social Apps
- → Content Archiving

Retail

- → Rich Product Catalogs
- → Customer Data Management
- → New Services
- → Digital Coupons
- → Real-Time Price Optimization

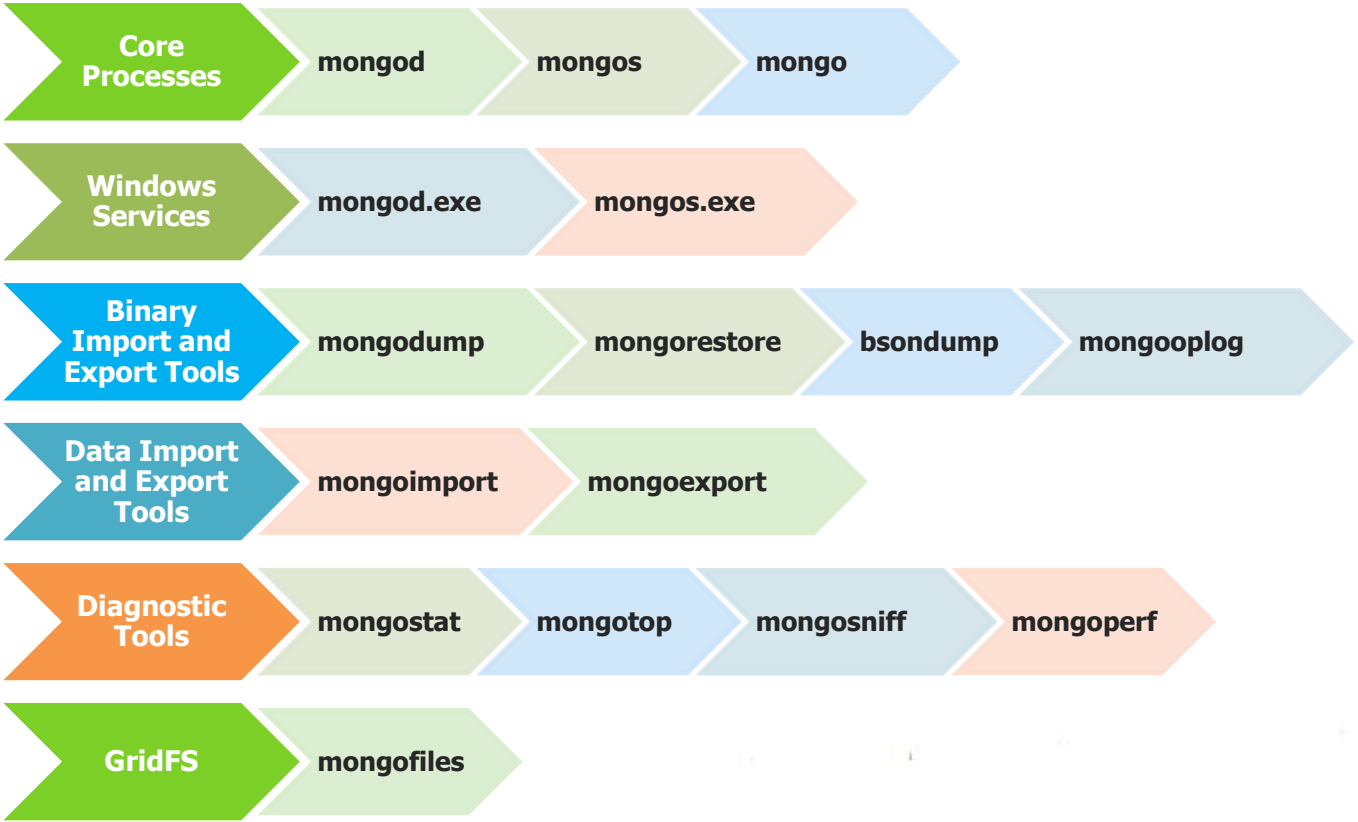


Telecommunication

- → Consumer Cloud
- → Product Catalog
- → Customer Service Improvement
- → Machine-to-Machine (M2M) Platform
- → Real-Time Network Analysis and Optimization

MongoDB Tools

MongoDB Package Components (Tools)



Annie's Question

Which binary do we use for Data Import & Export?

Annie's Answer

Ans.
mongoimport & mongoexport

Annie’s Question

Which binary do we use for backup & recovery?

Annie's Answer

Ans.
mongodump & mongorestore

Annie's Question

Which binary do we use to process large files?

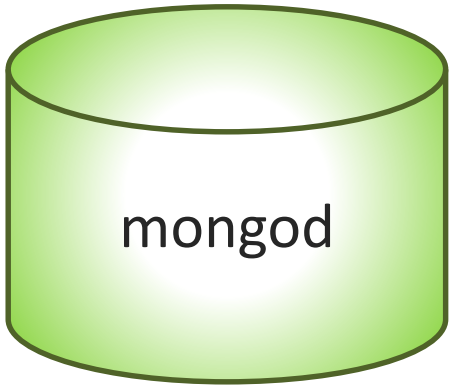
Annie’s Answer

Ans.
mongofiles

MongoDB Database

edureka!

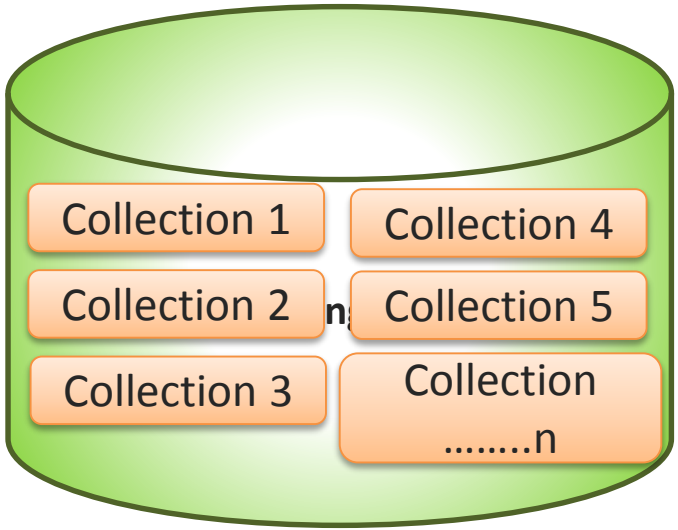
- → Mongod is the primary daemon process for the MongoDB system.
- → Database is a physical container for collections.
- → Each database gets its own set of files on the file system.
- → A single MongoDB server typically has multiple databases.
- → It handles data requests, manages data format, and performs background management operations.



MongoDB Collection

edureka!

- → Collection is a group of MongoDB documents.
- → It is the equivalent of an RDBMS table.
- → A collection exists within a single database.
- → Collections do not enforce a schema.
- → Documents within a collection can have different fields.
- → Typically, all documents in a collection are of similar or related purpose.

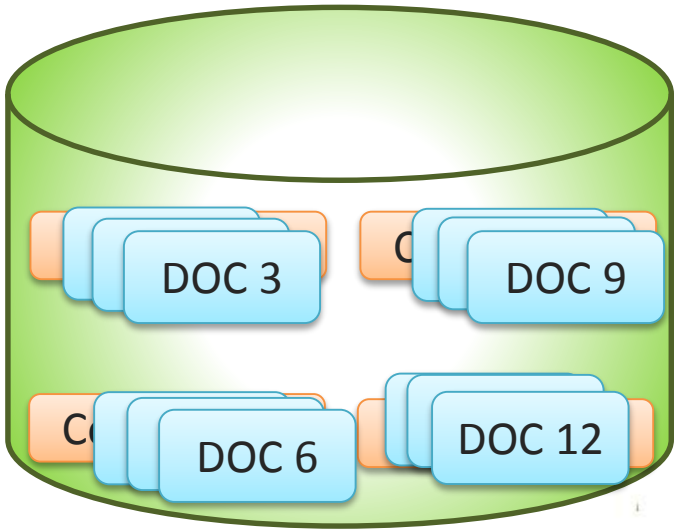


MongoDB Document

edureka!

→ A document is a set of key-value pairs.

→ Documents have dynamic schema.



MongoDB Sample Document

```
{
   _id: ObjectId(7df78ad8902c)
   title: 'edureka',
   description: 'Leading Training Provider Across Glob',
   by: 'edureka',
   url: 'http://www.edureka.in',
   tags: ['mongodb', 'database', 'NoSQL'],
   likes: 100,
   comments: [
      {
         user:'user1',
         message: 'My first comment',
         dateCreated: new Date(2011,1,20,2,15),
         like: 0
      },
      {
         user:'user2',
         message: 'My second comments',
         dateCreated: new Date(2011,1,25,7,45),
         like: 5
      }
   ]
}
```

RDBMS Terminology with MongoDB

edureka!

RDBMS	MongoDB
Database	Database
Table	Collection
Tuple/Row	Document
Column/Attribute/Variable	Field
Table Join	Embedded Documents
Database Server and Client	
Primary Key	Primary Key (Default key _id provided by mongodb itself)
Mysqld/Oracle	mongod
mysql/sqlplus	mongo

Annie's Question

Can MongoDB support join across multiple collections?

Annie's Answer

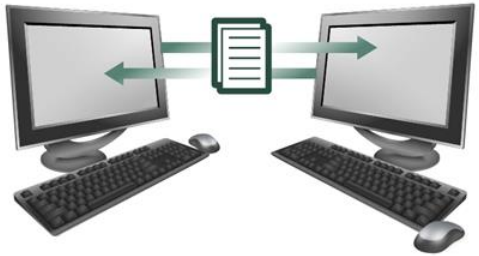
Ans.
No. It doesn't. To overcome this we do pre-join while designing the data model by embedding child document inside parent.

Introduction to JSON and BSON

edureka!

JavaScript Object Notation

JSON Abbreviation

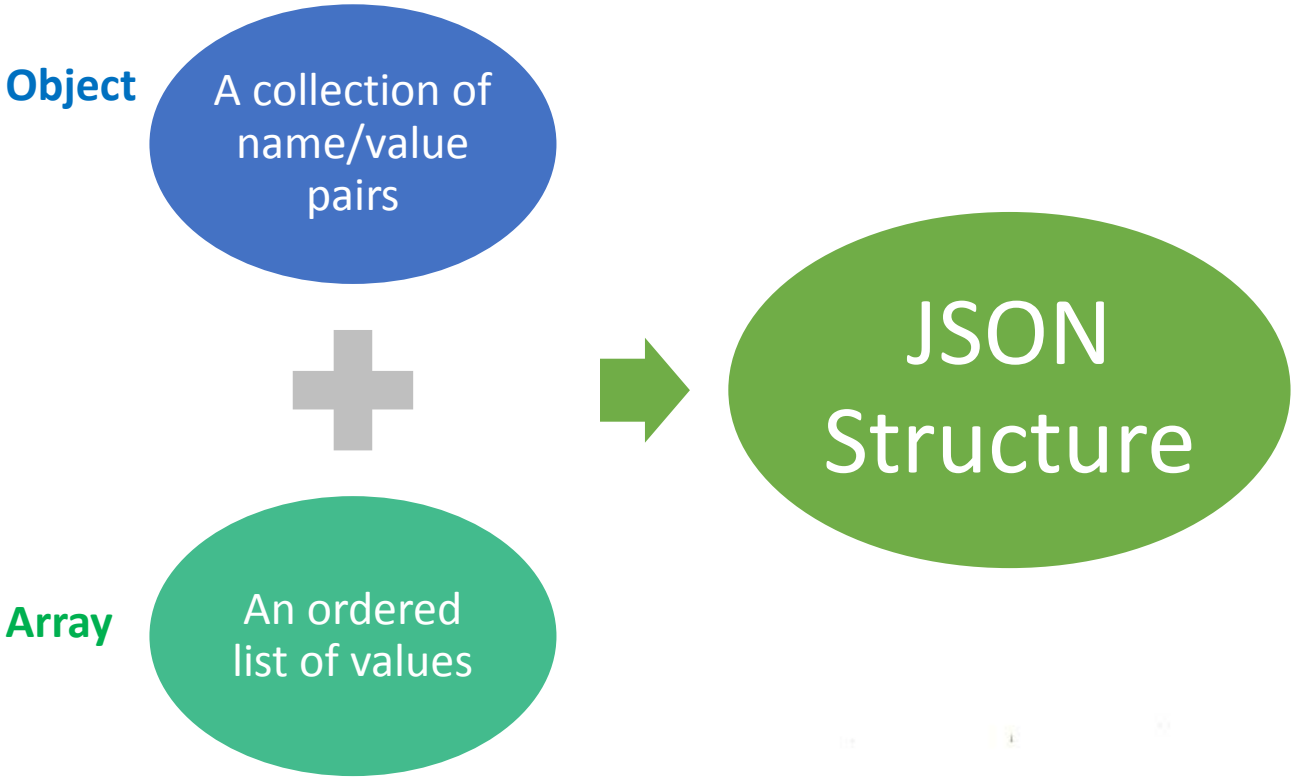


Lightweight data-interchange format

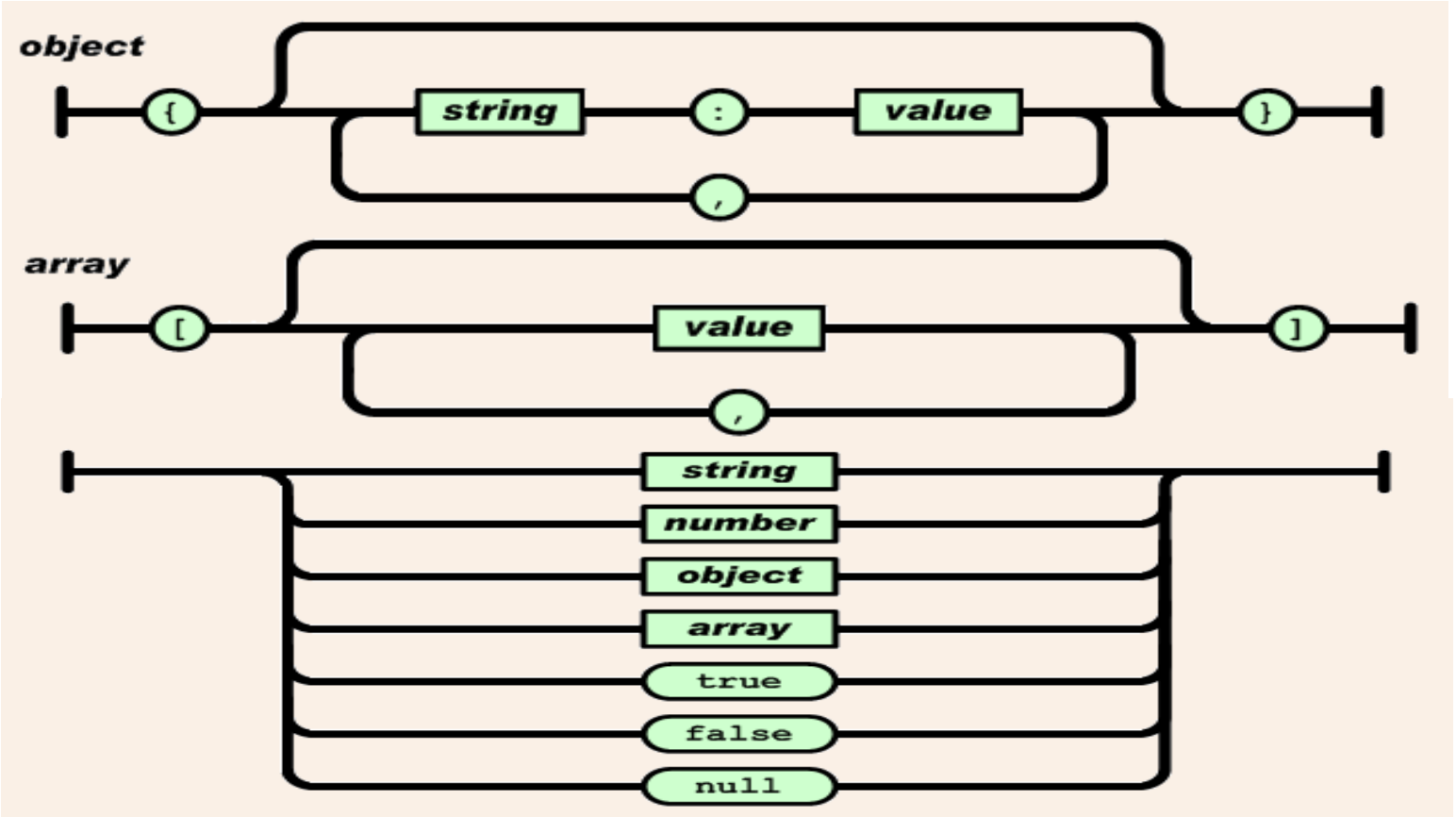
Text format that is completely language independent

Easy for humans to read and write

JSON (Contd.)


Object
A collection of name/value pairs
Array
An ordered list of values
JSON Structure

JSON (Contd.)


object
{
string
:
value
}
,
array
[
value
]
,
string
number
object
array
true
false
null

edureka!

Binary JavaScript Object Notation

BJSON Abbreviation

Supports the embedding of documents and arrays within other documents and arrays

Contains extensions that allow representation of data types that are not part of the JSON spec

Easy for machines to parse and generate

Text format that is completely language independent

JSON vs. BSON

edureka!

```
{"hello": "world"}  →  "\x16\x00\x00\x00\x02hello\x00
                        \x06\x00\x00\x00world\x00\x00"

{"BSON": ["awesome", 5.05, 1986]}  →  "1\x00\x00\x00\x04BSON\x00&\x00
                                       \x00\x00\x020\x00\x08\x00\x00
                                       \x00awesome\x00\x011\x00333333
                                       \x14@\x102\x00\xc2\x07\x00\x00
                                       \x00\x00"
```

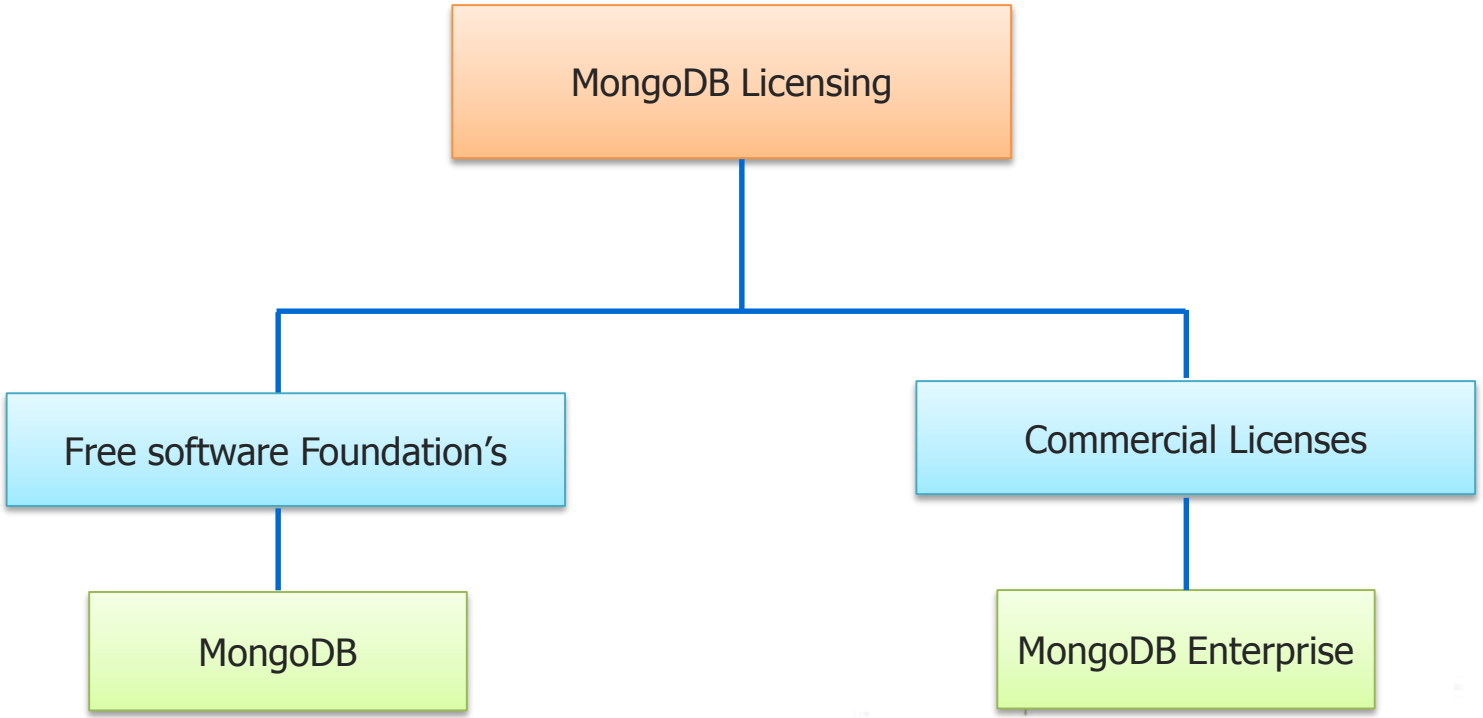
Annie's Question

What is difference between JSON & BSON ?

Annie's Answer

Ans.
JSON→ Java Script Object Notation
BSON→ Binary JSON (Represents JSON Data in Binary Format)

MongoDB Licensing



MongoDB Enterprise

edureka!

- → MongoDB Enterprise is the commercial edition of MongoDB that provides enterprise-grade capabilities.
- → MongoDB Enterprise includes advanced security features, management tools, software integrations and certifications.
- → These value-added capabilities are not included in the open-source edition of MongoDB.

MongoDB Enterprise Includes

MongoDB Enterprise Includes

edureka!

Features	MongoDB	MongoDB Enterprise
JSON Data Model with Dynamic Schemas	•	•
Auto-Sharding for Horizontal Scalability	•	•
Built-In Replication and High Availability	•	•
Full, Flexible Index Support	•	•
Rich Document Queries	•	•
Fast In-Place Updates	•	•
Aggregation Framework and MapReduce	•	•
Large Media Storage with GridFS	•	•
Text Search	•	•
Cloud, On-Premise and Hybrid Deployments	•	•
Role-Based Privileges	•	•
Advanced Security with Kerberos		•
On-Prem Management		•
SNMP Support		•
OS Certifications		•
Private On-Demand Training		•

MongoDB Installation – Live Demo

MongoDB Installation – Live Demo

edureka!

- → Running MongoDB on Windows
- → Installation of MongoDB on Windows as a Service
- → Running of MongoDB on Linux (CentOS)
- → Installation of MongoDB on CentOS

QUESTIONS

Assignment

edureka!

→Write a JSON document which can have all data types supported by JSON?

→What all core differences are there in MongoDB, Hadoop, HBase and Cassandra?

→How can you define Horizontal & Vertical Scalability?

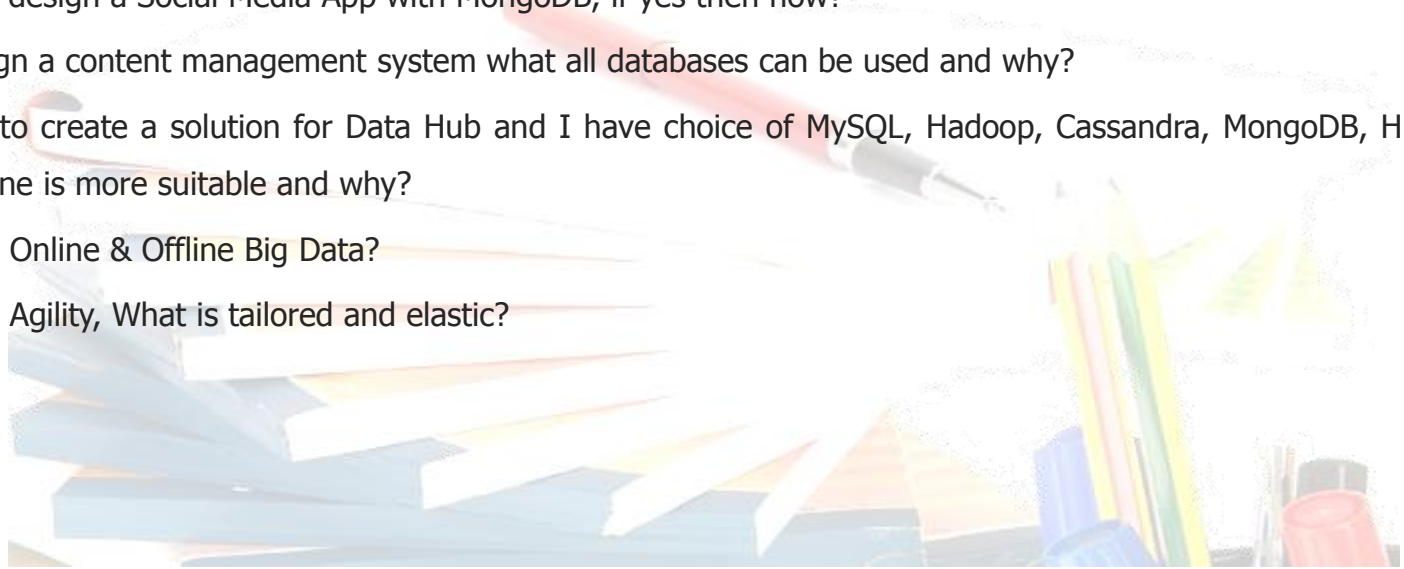
→Can we design a Social Media App with MongoDB, if yes then how?

→To design a content management system what all databases can be used and why?

→I want to create a solution for Data Hub and I have choice of MySQL, Hadoop, Cassandra, MongoDB, HBase, which one is more suitable and why?

→What is Online & Offline Big Data?

→What is Agility, What is tailored and elastic?



Pre-work for Next Module

edureka!

→MongoDB Installation on Windows or MongoDB Installation on Centos

→Generate Test Data on MongoDB Database

→Execute all Module1 Script present in LMS

→Read FAQ Module1 in LMS

→Read FAQ Module1 in LMS

→Take Quiz in LMS

→Complete Assignment

Agenda for Next Module

edureka!

→MongoDB Development Overview

→MongoDB Production Overview

→MongoDB CRUD Introduction

→MongoDB CRUD Concepts

→MongoDB CRUD Syntax & Queries

edureka!