

MongoDB – Overview

Lesson 01

Version Sheet: MANDATORY (hidden in ‘slide show’ mode)

Version	Changes	Author
001	Redesign	Rohan Salvi



Copyright © Capgemini 2015. All Rights Reserved 2

Note to the SME:

Please read before you begin reviewing this module as SME

Dark Red box with white text

Note to the SME

The SME must provide missing info.

Amber box with black text

Note to the SME

The SME must validate the re-design/
the modification/ addition.

Note to the SME:

This is a temporary slide and will not be part of the final upload at all. It is to help the SME understand how we will communicate changes to them.

To Review and Navigate the comments in the presentation Click on the “Review” Tab and then Click “Next” to review all the comments . Also you can add your comments using the “New comment” button

The comments in the review section have also been updated in a green textbox with black text.



Copyright © Capgemini 2015. All Rights Reserved 3

Ground Rules for Face-to-face Classrooms

	Everyone participates	Respect individual opinions and diversities	
	Be open and honest	Give headlines, be concise	
	One speaker at a time	Make language a non-issue	
	Stick to time contracts	Seek first to understand, and then to be understood	
	Clean desk / room policy	No mobile phone, No computer (except for surveys, polls etc)	
	Clients & Leadership are often around, please remember that	Maintain a spirit of fun and enthusiasm	

 Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 4

Best to print on an A3 sheet and display in class. The idea is to co-create, to connect to learn, to do-learn-do (knowing is not enough, the real purpose is to apply what we learn on the job). It would also be great to have the class share their learning with their teams. You are the facilitator, you will help all share knowledge and learn new concepts or skills, there will be no lectures, everyone is a teacher.

Ground Rules for Virtual Classrooms

Participate actively in each session

- Share experiences and best practices
- Bring up challenges, ask questions
- Discuss successes
- Respond to whiteboards, polls, quizzes, chat boxes
- Hang up if you need to take an urgent phone call, don't put this call on hold

Communicate professionally with others

- Mute when you're not speaking
- Wait for others to finish speaking before you speak
- Each time you speak, state your name
- Build on others' ideas and thoughts
- Disagreeing is OK –with respect and courtesy

Be on time for each virtual session

As a best practice...be just a few minutes early!



Copyright © Capgemini 2015. All Rights Reserved 5

Show this slide and hide the one for onsite classrooms, if this is a virtual session.

ONLY for Virtual classrooms

Module at a Glance

SME to provide the details required in the table.

Target Audience:	
Course Level:	<i>Basic</i>
Duration (in hours):	30 mins
Pre-requisites, if any:	NA
Post-requisites, if any:	<i>Submit Session Feedback</i>
Relevant Certifications:	None



Copyright © Capgemini 2015. All Rights Reserved 6

Introductions (for Virtual Classrooms)

Business Photo

SME to provide the photos and names of the facilitators.

Business Photo

Facilitator
Name
Role

Moderator
Name
Role



Copyright © Capgemini 2015. All Rights Reserved

7

Add pics and Capgemini roles for the Facilitator and moderator. Establish their credibility, what qualifies them to facilitate this session.

Agenda

- 1 NOSQL Introduction & Types
- 2 Introduction & Basics
- 3 CRUD
- 4 Aggregation
- 5 Indexes
- 6 Performance
- 7 Replication & Sharding



Copyright © Capgemini 2015. All Rights Reserved 8

Module Objectives



What you will learn

At the end of this module, you will learn:

- What is NOSQL

Note to the SME : Please provide the module Objectives or validate the partially updated content

What you will be able to do



At the end of this module, you be able to:

- Understand what is NOSQL
- Describe CRUD
- State the types of NOSQL
- Explain what is Aggregation
- Describe Replication & Sharding

Lets Get Social

Let's get Social

10gen | mongoDB

Capgemini

Copyright © Capgemini 2015. All Rights Reserved 10

More Technologies

More Technologies and Requirements Than Ever

NoSQL Document Data Stores
Analytics

Big Data JSON
Key-value Datawarehouse

Hadoop ODS

MongoDB
Graph Wide-column

Cloud Computing

Agile Development

Internet of Things

Mobile Gamification

Social networking

Consumerization

4



Copyright © Capgemini 2015. All Rights Reserved 11

Boom of NOSQL

HOW TO WRITE A CV

The cartoon consists of three panels. In the first panel, two people are at a computer; one says 'NO' and the other asks 'DO YOU HAVE ANY EXPERTISE IN SQL?'. In the second panel, the person on the right says 'geek & pole'. In the third panel, they say 'DOESN'T MATTER. WRITE: "EXPERT IN NO SQL"'. Below the third panel is the text 'Leverage the NoSQL boom'.

geek & pole

DO YOU HAVE ANY EXPERTISE IN SQL?

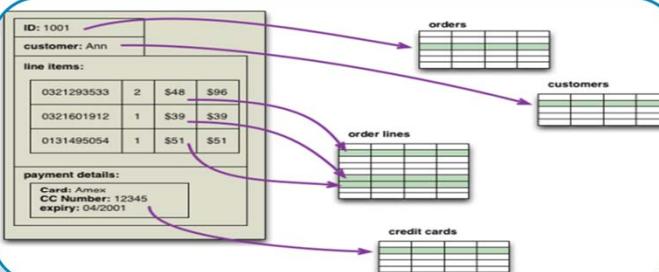
DOESN'T MATTER. WRITE: "EXPERT IN NO SQL"

Leverage the NoSQL boom

Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 12

Why NoSQL



- Handles Schema Changes Well (easy development)
- Solves Impedance Mismatch problem
- Rise of JSON
 - python module: simplejson

NO SQL (or) BIG Data Envt

The diagram shows three overlapping circles. The leftmost circle is orange and labeled 'VELOCITY'. The middle circle is light gray and labeled 'VARIETY'. The rightmost circle is light gray and labeled 'VOLUME'.

Web Applications Driving Data Growth

The graph plots 'Stored Digital Information (exabytes)' against time. It shows two main trends: a blue line for 'Business Transaction Data' and a green line for 'Web Application Data'. Both lines show a sharp increase starting around 2000, with the green line rising more steeply after 2005.

Year	Business Transaction Data (Exabytes)	Web Application Data (Exabytes)
1970	~0.1	~0.1
1980	~0.2	~0.2
1990	~0.5	~0.5
2000	~1.0	~1.0
2010	~10.0	~100.0

©w3resource.com

Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 14

Example 1

Social Network Graph

Each record: UserID1, UserID2

Separate records: UserID, first_name, last_name, age, gender, ...

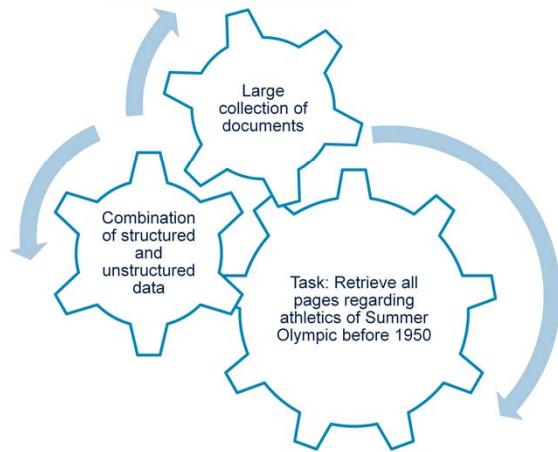
Task: Find all friends of friends of friends of ... friends of a given user

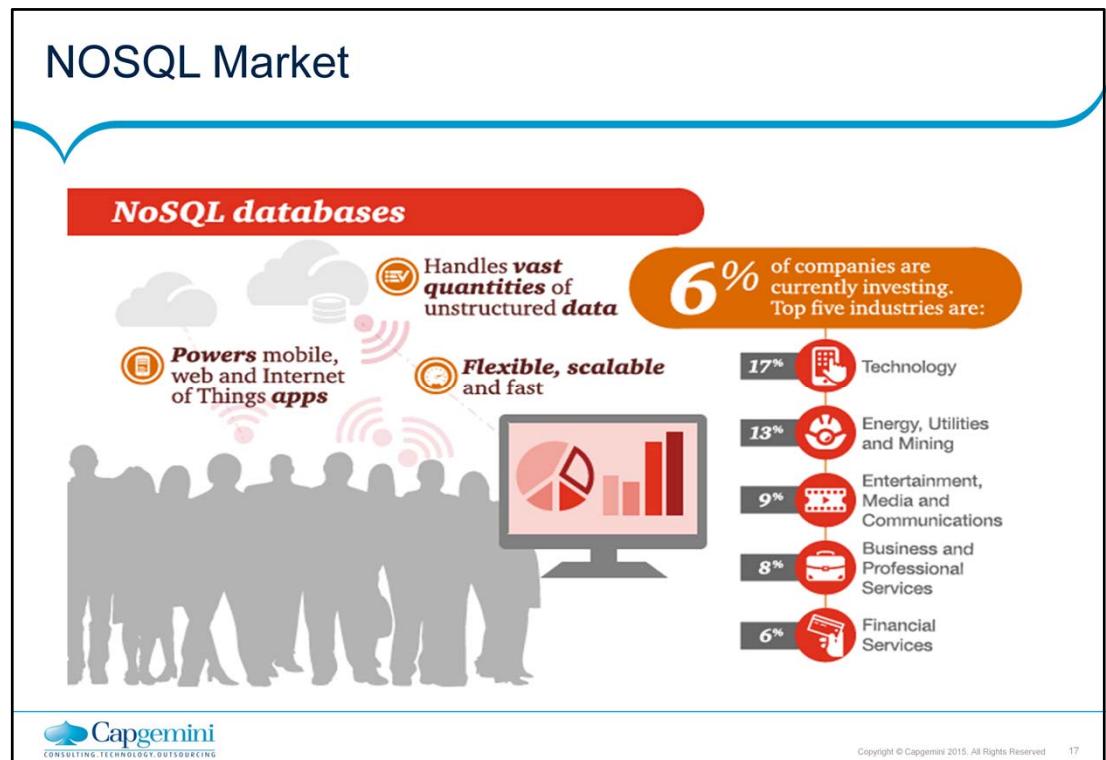


Copyright © Capgemini 2015. All Rights Reserved 15

Example 2

Wikipedia Pages





What is NoSQL?

Stands for ‘Not Only SQL’.

Originally refers to “non SQL” or “non Relational” database.

Term coined by Carlo Strozzi in 1998.

Open Source.

No Rows-Columns / Tables.

No Predefined schema.

Eventually consistency rather than ACID property.

Distributed computing.

Unstructured and unpredictable data.

Prioritizes high scalability ,high availability and scalability.

Replication support .



Copyright © Capgemini 2015. All Rights Reserved 18

NoSQL Database Types

- Key-value
- Graph database
- Document-oriented
- Column family

redis
riak
Neo4j
HyperGraphDB
mongoDB
CouchDB
Cassandra
HBASE

Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 19

- Huge quantity of data => Distributed systems => expensive joins =>
- New fields, new demands (graphs) =>

Different data structures:
Simpler or more specific

CAP Theorem

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.

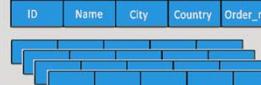


Copyright © Capgemini 2015. All Rights Reserved 20

NO SQL Databases Types

Key Value Stores

row-store



- + easy to add/modify a record
- might read in unnecessary data

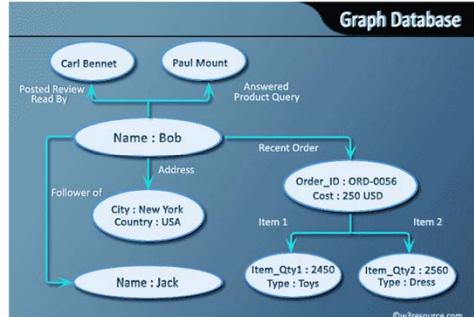
column-store



- + only need to read in relevant data
- tuple writes require multiple accesses

=>suitable for read-mostly, read-intensive, large data repositories

Graph Database



Graph Database

Carl Bennet, Paul Mount, Name : Bob, Name : Jack, City : New York, Country : USA, Order_ID : ORD-0056

Posted Review, Read By, Answered Product Query, Recent Order, Follower of, Address

Item 1, Item 2, Item_Qty1, Item_Qty2, Type

©w3resource.com



Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 21

Types of NoSQL Databases

Wide Column Store

Column Store

Database	Table : supplier	Table : order
T/SCF	ID : 1 C: Name : Bob CF/SC: Address C: City : New York C: Country : USA CF/SC: Order C: Order_no : ORD-0056	Order_ID : ORD-0056 CF/SC: Price: C: Cost : 250 USD CF/SC: Item: C: Item_Qty1 : 2450 C: Item_Qty2 : 2560
Row	ID : 2 C: Name : Jack CF/SC: Address C: City : Paris C: Country : France CF/SC: Order C: Order_no : ORD-0057	Order_ID : ORD-0057 CF/SC: Price: C: Cost : 400 USD CF/SC: Item: C: Item_Qty1 : 3000 C: Item_Qty2 : 3530

©w3resource.com

Document Store

Document Store

Database	Table : supplier	Table : order
Document	ID : 1 Name : Bob Address : City : New York Country : USA Order : Order_no : ORD-0056	Order_ID : ORD-0056 Cost : 250 USD Item_Qty1 : 2450 Item_Qty2 : 2560
Document	ID : 2 Name : Jack Address : City : Paris Country : France Order : Order_no : ORD-0057	Order_ID : ORD-0057 Cost : 400 USD Item_Qty1 : 3000 Item_Qty2 : 3530

©w3resource.com

 Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 22

Production Deployment



Google



Facebook



Mozilla



Adobe



Foursquare



LinkedIn



Digg



McGraw-Hill
Education



Vermont
Public Radio



Copyright © Capgemini 2015. All Rights Reserved 23

MongoDB Introduction



mongoDB®



Copyright © Capgemini 2015. All Rights Reserved 24

MongoDB Philosophy

MongoDB and Enterprise IT Stack

The diagram illustrates the MongoDB's place within the enterprise IT stack. It shows a layered architecture:

- Management & Monitoring:** On the left.
- Applications:** CRM, ERP, Collaboration, Mobile, BI.
- Data Management:** A central layer.
- Online Data:** mongoDB, RDBMS.
- Offline Data:** Hadoop, EDW.
- Infrastructure:** OS & Virtualization, Compute, Storage, Network.

A vertical bar on the right indicates "Scalability & Auditing".

Operational Database Landscape

This graph plots database systems against two axes:

- Scalability & Performance (Y-axis):** An upward arrow.
- Depth of Functionality (X-axis):** A horizontal arrow pointing right.

Legend:

- key/value stores (grey circle)
- wide column (grey circle)
- MongoDB (green circle)
- RDBMS (grey circle)

MongoDB is positioned in the upper-right quadrant, indicating high scalability and depth of functionality.

 Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 25

Page 01-25

What is MongoDB

MongoDB (from *humongous*) is a cross-platform document oriented database

Has driver to all most every popular language programming

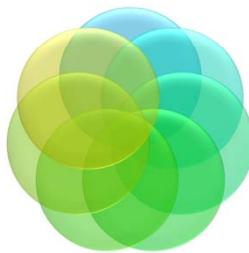
10gen Inc. also offers professional services around MongoDB

Open-source project that mainly driven by 10gen Inc.

Written in C++

Schema-free document database

Nosql database



Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

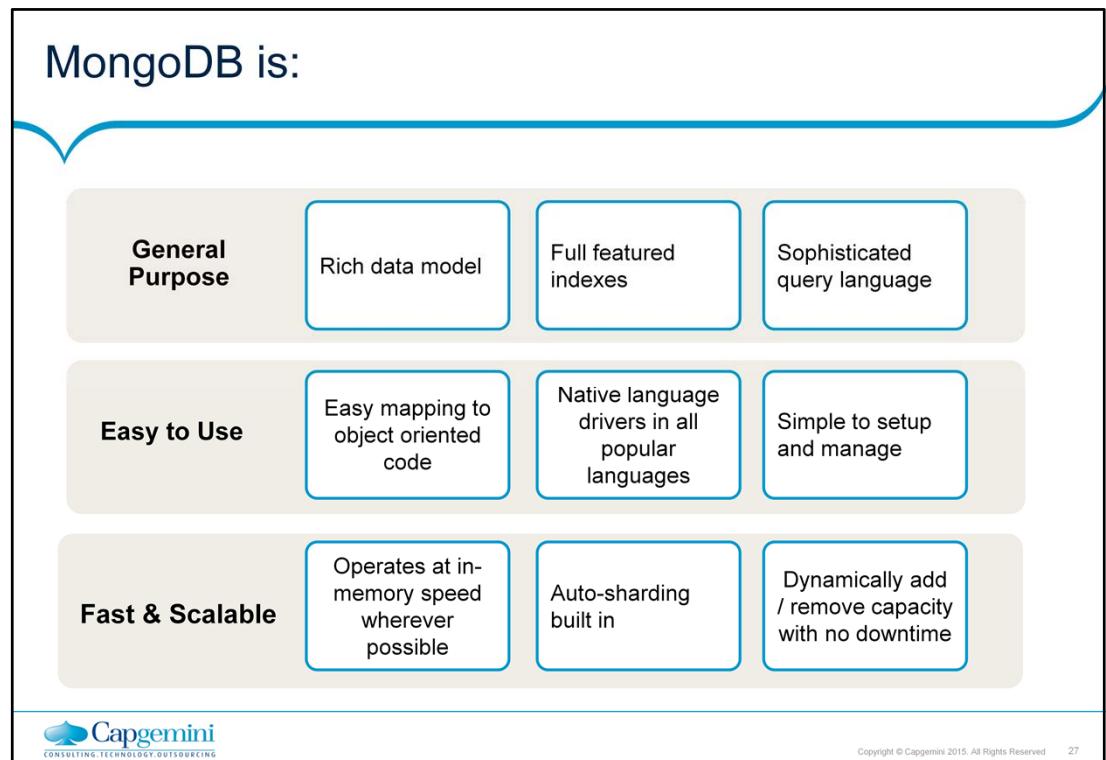
Copyright © Capgemini 2015. All Rights Reserved 26

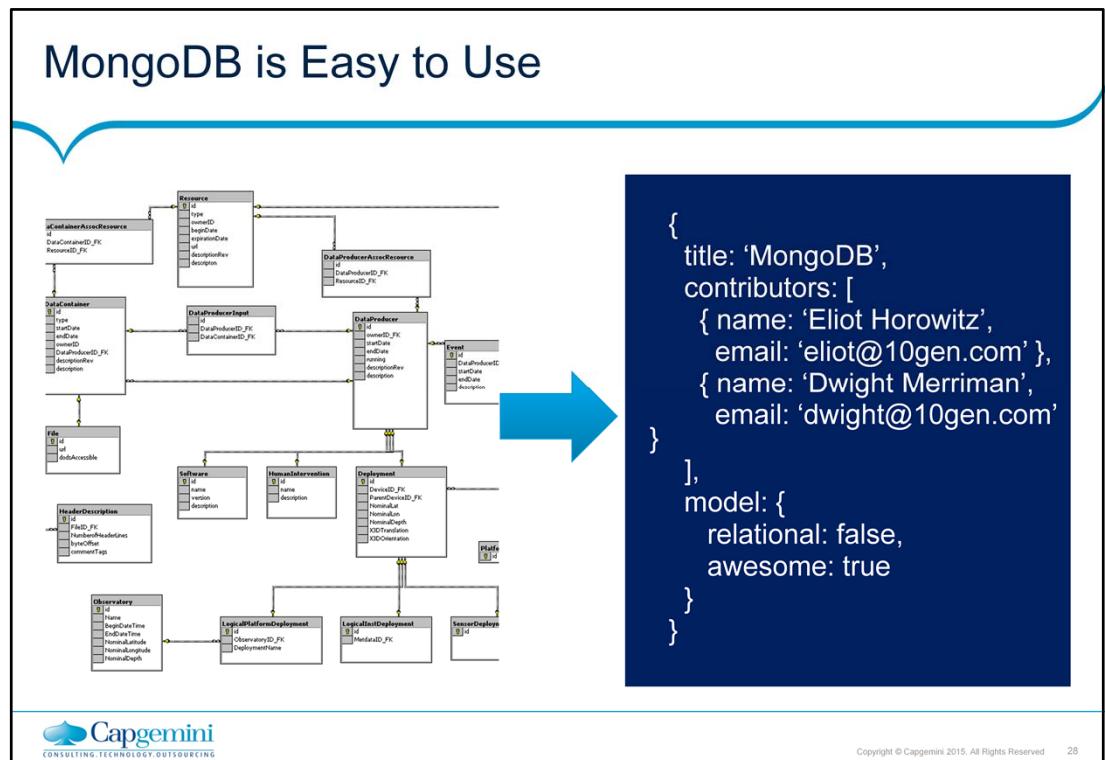
MongoDB (from *humongous*) is a cross-platform document oriented database . Classified as a NoSQL database, MongoDB eschews the traditional table-based relational database structure in favor of JSON -like documents with dynamic schemas (MongoDB calls the format BSON), making the integration of data in certain types of applications easier and faster. Released under a combination of the GNU Affero General Public License and the Apache License, MongoDB is free and open-source software

MongoDB stores data in the form of documents, which are JSON-like field and value pairs.

Documents are analogous to structures in programming languages that associate keys with values (e.g. dictionaries, hashes, maps, and associative arrays).

Formally, MongoDB documents are BSON documents. BSON is a binary representation of JSON with additional type information. In the documents, the value of a field can be any of the BSON data types, including other documents, arrays, and arrays of documents





Schema Free

MongoDB does not need any pre-defined data schema.

Every document could have different data!

```
{name: "will", eyes: "blue", birthplace: "NY", aliases: ["bill", "la ciacco"], loc: [32.7, 63.4], boss: "ben"}  
{name: "jeff", eyes: "blue", loc: [40.7, 73.4], boss: "ben"}  
{name: "brendan", aliases: ["el diablo"]}  
{name: "matt", pizza: "DiGiorno", height: 72, loc: [44.6, 71.3]}
```

mongoDB

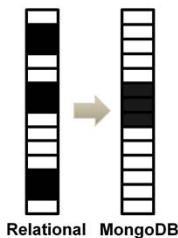
Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 29

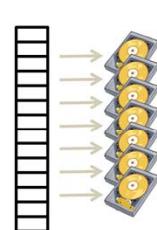
Applications enforce the data “schema” and integrity, much like MUMPS does in VistA

MongoDB is Fast and Scalable

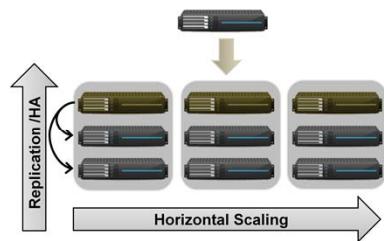
Better data locality

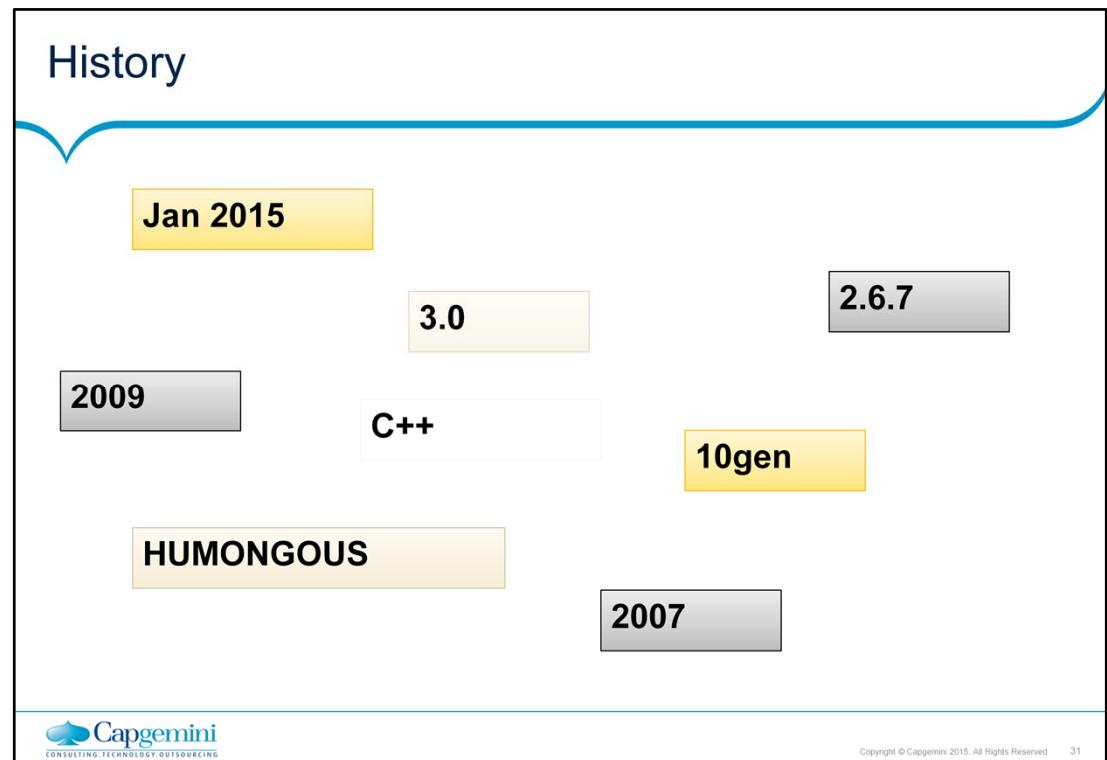


In-Memory Caching



Distributed Architecture





Features of MongoDB

Document Oriented Database

Adhoc queries

Indexing

High Performance

High Availability

Sharding

Easy Scalability

File Storage

Rich Query Language

Load Balancing

Replication



Copyright © Capgemini 2015. All Rights Reserved 32

Document-oriented

Documents (objects) map nicely to programming language data types

Embedded documents and arrays reduce need for joins

Dynamically-typed (schemaless) for easy schema evolution

No joins and no multi-document transactions for high performance and easy scalability

High performance

No joins and embedding makes reads and writes fast

Indexes including indexing of keys from embedded documents and arrays

Optional streaming writes (no acknowledgements)

High availability

Replicated servers with automatic master failover

Easy scalability

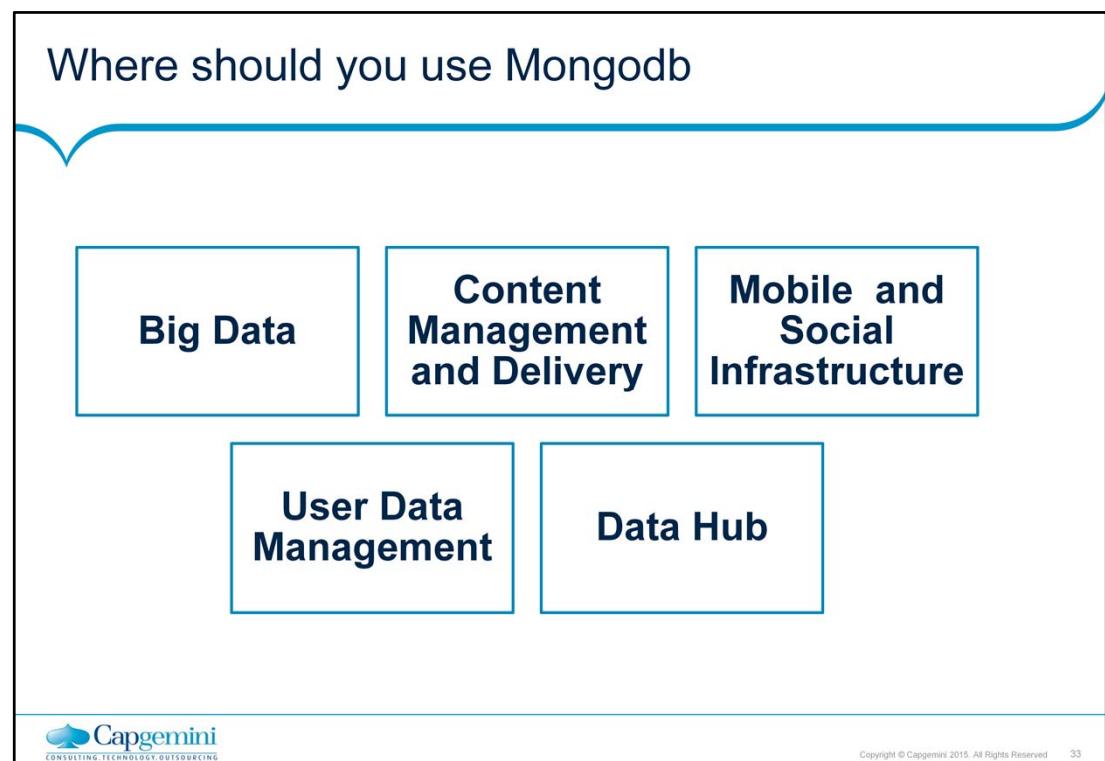
Automatic sharding (auto-partitioning of data across servers)

Reads and writes are distributed over shards

No joins or multi-document transactions make distributed queries easy and fast

Eventually-consistent reads can be distributed over replicated servers

Rich query language



Document-oriented

Documents (objects) map nicely to programming language data types

Embedded documents and arrays reduce need for joins

Dynamically-typed (schemaless) for easy schema evolution

No joins and no multi-document transactions for high performance and easy scalability

High performance

No joins and embedding makes reads and writes fast

Indexes including indexing of keys from embedded documents and arrays

Optional streaming writes (no acknowledgements)

High availability

Replicated servers with automatic master failover

Easy scalability

Automatic sharding (auto-partitioning of data across servers)

Reads and writes are distributed over shards

No joins or multi-document transactions make distributed queries easy and fast

Eventually-consistent reads can be distributed over replicated servers

Rich query language

MongoDB to SQL Terminology

MongoDB	SQL
database	database
collection	table
document	record (row)
field	column
linking/embedded documents	join
primary key (_id field)	primary key (user designated)
index	index



Copyright © Capgemini 2015. All Rights Reserved 34

Important Terminology

Database

- 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.

Collection

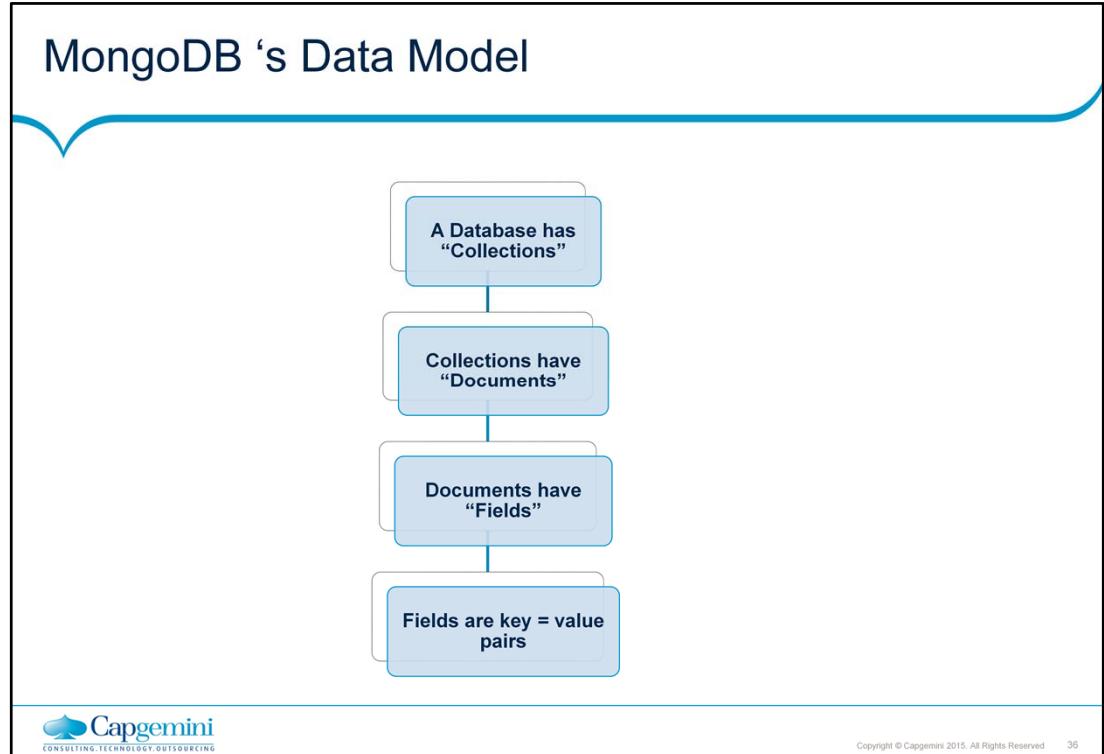
- Collection is a group of MongoDB documents. It is the equivalent of an RDBMS table. A collection exists within a single database. Collections do not enforce a schema. Documents within a collection can have different fields. Typically, all documents in a collection are of similar or related purpose.

Document

- A document is a set of key-value pairs. Documents have dynamic schema. Dynamic schema means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection's documents may hold different types of data.



Copyright © Capgemini 2015. All Rights Reserved 35



Data Model

Document based (max 16 MB).

Documents are in BSON formats consisting of field / value pairs.

Each document stored in a collection.

Schema less.



Copyright © Capgemini 2015. All Rights Reserved 37

Data Model

Document Data Model

Relational

7

Documents are Rich Data Structures

MongoDB

```
{
  "first_name": "Paul",
  "surname": "Miller",
  "city": "London",
  "location": [45.123, 47.232],
  "Profession": ["banking", "finance", "trader"],
  "cars": [
    { "model": "Bentley", "year": 1973, "value": 100000, ... },
    { "model": "Rolls Royce", "year": 1965, "value": 330000, ... }
  ]
}
```

8

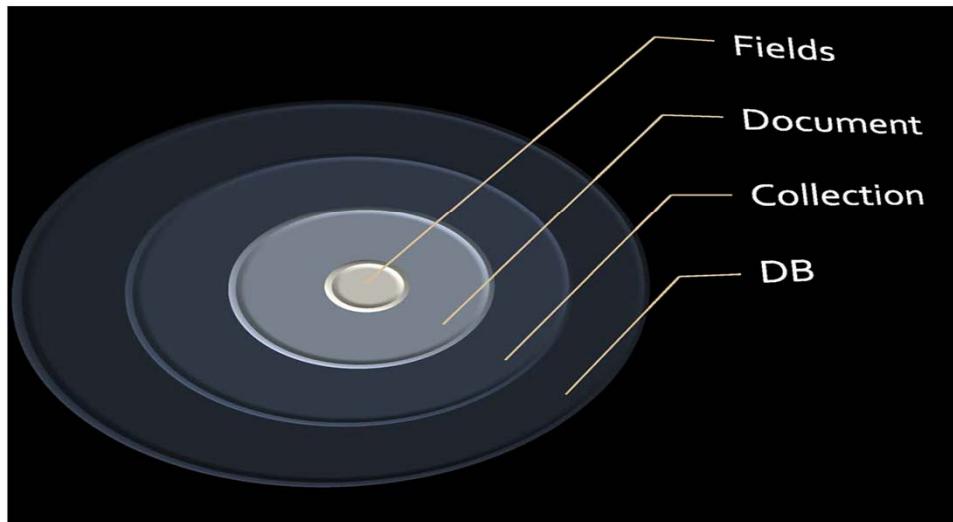
Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 38

Fields

Fields can contain arrays
Fields can contain an array of sub-documents

MongoDB Data Model



The Basics of MongoDB

A MongoDB instance may have one or more Databases.

A database may have one or more Collections.

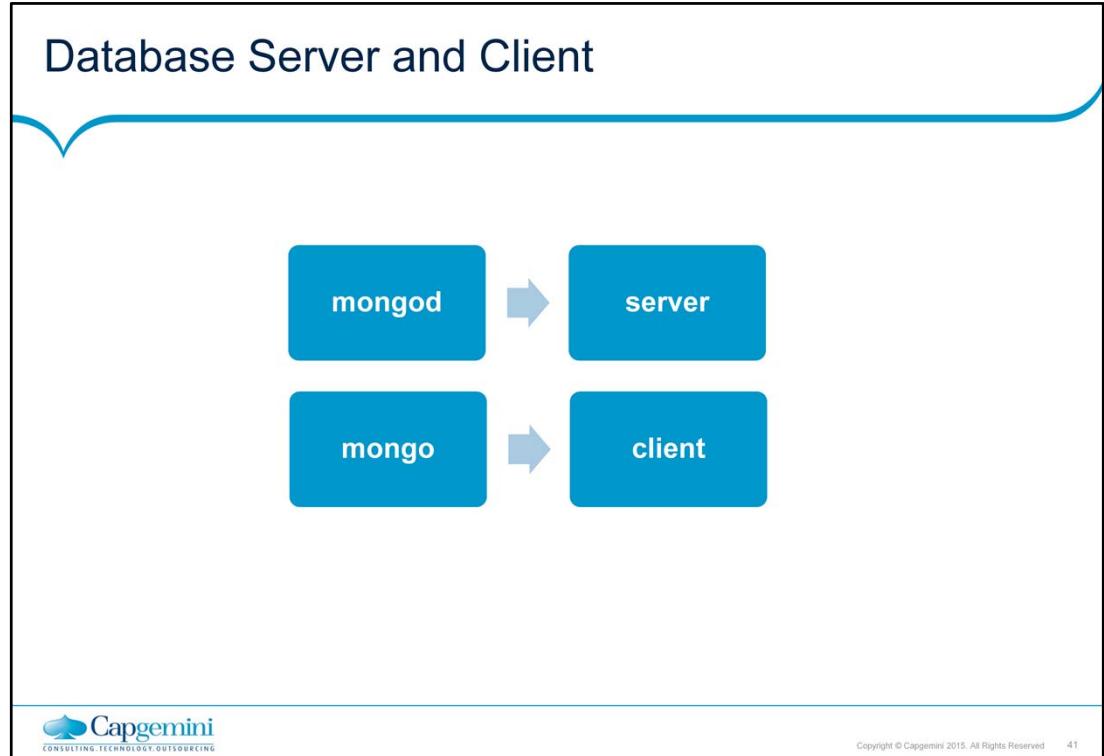
A collection may have zero or more Documents.

A document may have one or more Fields.

MongoDB indexes function much like their RDBMS counterparts.



Copyright © Capgemini 2015. All Rights Reserved 40



Getting Started

Install mongodb on windows from the link given below:

<http://www.mongodb.org/downloads>

Make sure you get correct version of MongoDB depending upon your windows version.

MongoDB for Windows 64-bit: This build type of MongoDB runs on any 64-bit version of Windows latest than Windows XP, involve Windows Server 2008 R2 and Windows 7 64-bit.



Copyright © Capgemini 2015. All Rights Reserved 42

What MongoDB does, How it works

MongoDB is a server process that runs on Windows/Linux , Os X.

It can be run both as a 32 or 64-bit application. We recommend running in 64-bit mode, since Mongo is limited to a total data size of about 2GB for all databases in 32-bit mode.

Clients connect to the MongoDB process, optionally authenticate themselves if security is turned on, and perform a sequence of actions, such as inserts, queries and updates.



Copyright © Capgemini 2015. All Rights Reserved 43

Starting the MongoDB Server

Create a directory where MongoDB stores all its data.

The MongoDB default data directory path is \data\db.

Create the data folder in D:\

Set the Path.

Run mongod.exe

To start MongoDB server, we need to run mongod.exe



Starting the MongoDB Server (contd.)

D:\set up\mongodb>mongod.exe --dbpath "d:\set up\mongodb\data"

This will show **waiting for connections** message on the console output indicates that the mongod.exe process is running successfully.

Now to run the mongodb you need to open another command prompt and issue the following command.



Copyright © Capgemini 2015. All Rights Reserved 45

Starting the MongoDB Server (contd.)

```
D:\set up\mongodb\bin>mongo.exe
```

```
MongoDB shell version: 2.2.0  
connecting to: test  
Welcome to the MongoDB shell
```



Copyright © Capgemini 2015. All Rights Reserved 46

The Mongo Shell

```
D>mongo  
  
>help()  
  
>show dbs  
  
>use <dbname>  
  
>show collections  
  
>db.collectionName.findOne()  
  
>db.collectionName.find()  
  
>db.help()  
  
>db.collectionName.help()
```

```
C:\appservers\mongo-1.6.3\bin\mongo.exe  
MongoDB shell version: 1.6.3  
connecting to: test  
> show dbs  
admin  
cfmongodb_tests  
default_db  
local  
mongorocks  
test  
>>  
> use mongorocks  
switched to db mongorocks  
> show collections  
people  
system.indexes  
>  
> db.people.findOne()  
{  
    "_id" : ObjectId("4cb66dae636ac4fa2045ff31"),  
    "COUNTER" : NumberLong(1),  
    "LOVEMONGO" : true,  
    "NAME" : "Bar",  
    "BIKE" : "Felt",  
    "LOVESSQL" : true,  
    "KIDS" : [  
        {"NAME" : "Alexis",  
         "AGE" : NumberLong(7),  
         "DESCRIPTION" : "crazy",  
         "HAIR" : "blonde"},  
        {"NAME" : "Sidney",  
         "AGE" : NumberLong(2),  
         "DESCRIPTION" : "crazy",  
         "HAIR" : "dirty blonde"},  
    ],  
    "WIFE" : "Heather",  
    "TS" : "Wed Oct 13 2010 22:40:46 GMT-0400 (Eastern Daylight Time)"  
}  
>
```



Summary

What is NOSQL database

Advantages of NOSQL

Why MongoDB

MongoDB Document database

MongoDB data model

Mongo Shell

Establishing Connection

Understand about Collection, document and fields



Copyright © Capgemini 2015. All Rights Reserved 48