MongoDB Overview

- NOSQL Introduction & Types
- Introduction & Basics
- CRUD
- Aggregation
- Indexes
- Performance Replication & Sharding

Module Objectives

What you will learn

At the end of this module, you will learn:

- What is NOSQL

What you will be able to do

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

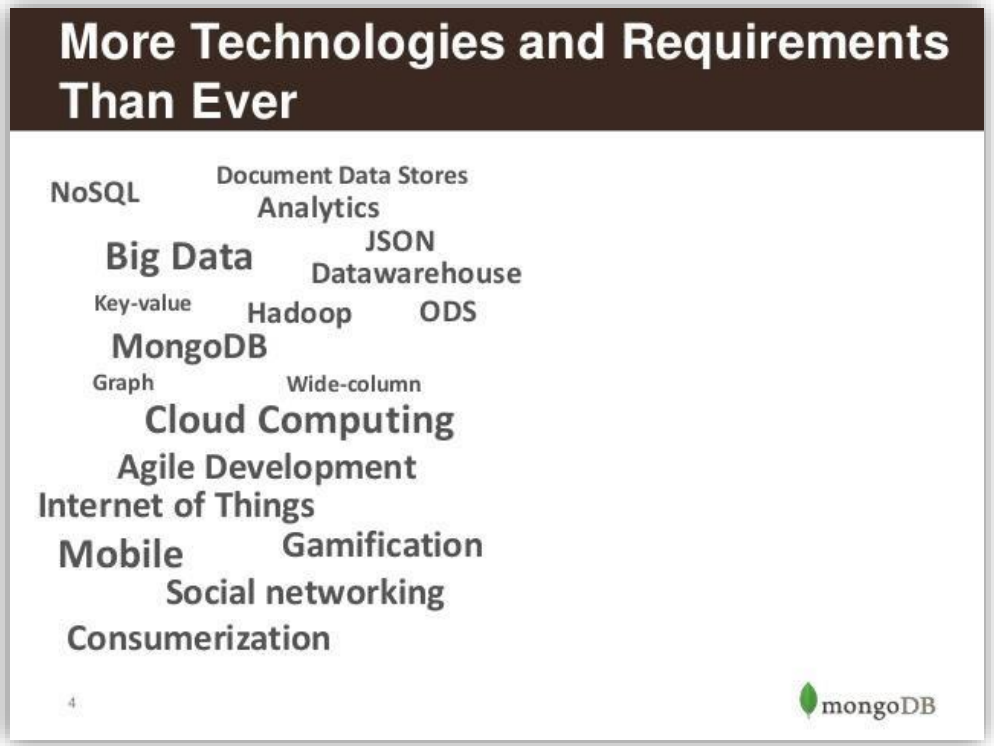
- Understand what is NOSQL
- Describe CRUD
- State the types of NOSQL
- Explain what is Aggregation
- Describe Replication & Sharding

9

Lets Get Social

10

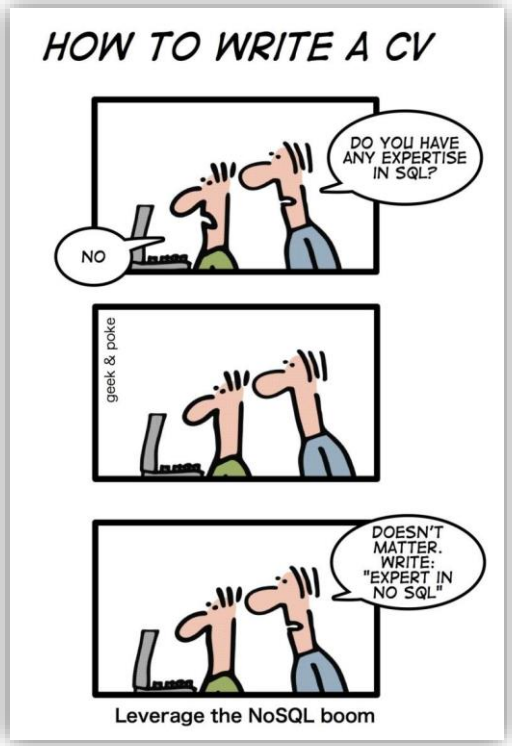
More Technologies



11

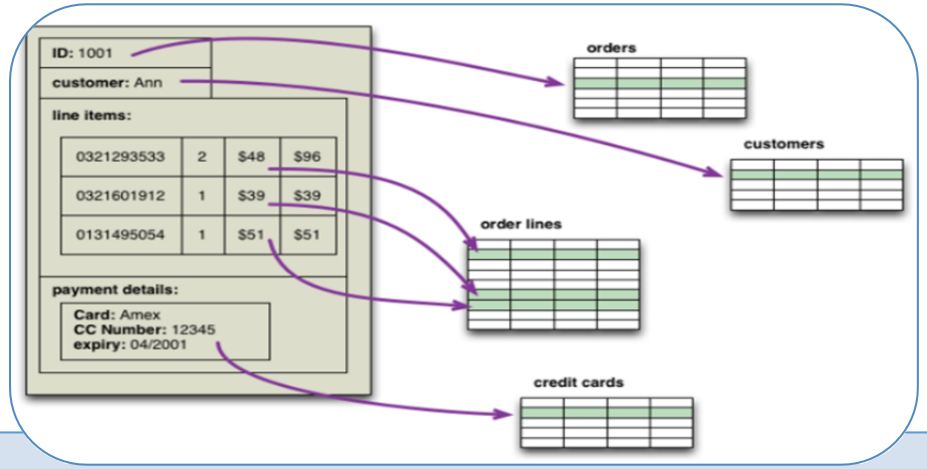
5

Boom of NOSQL



12

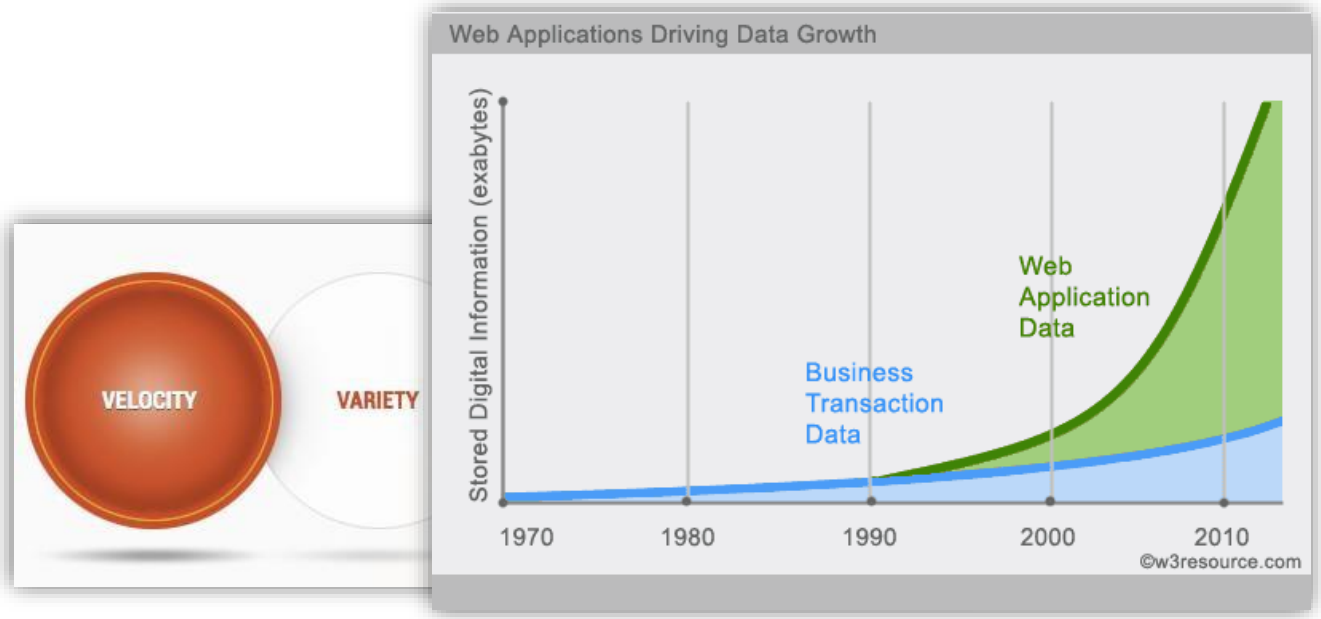
Why NoSQL



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

13

NO SQL (or) BIG Data Envt



14

Example 1

Social Network Graph

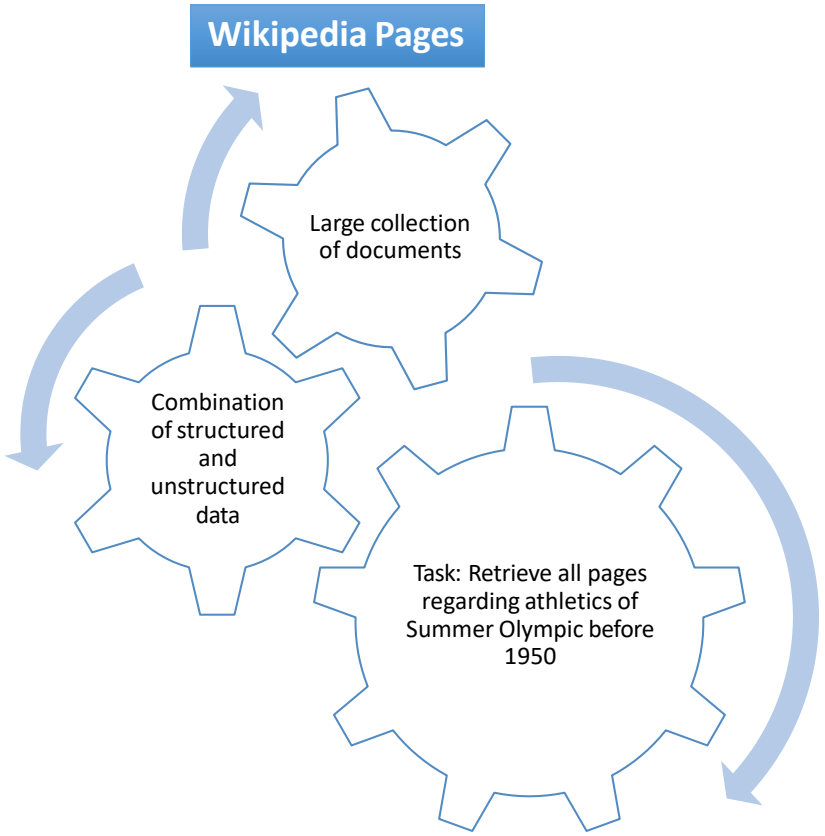
Each record: UserID1, UserID2

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

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

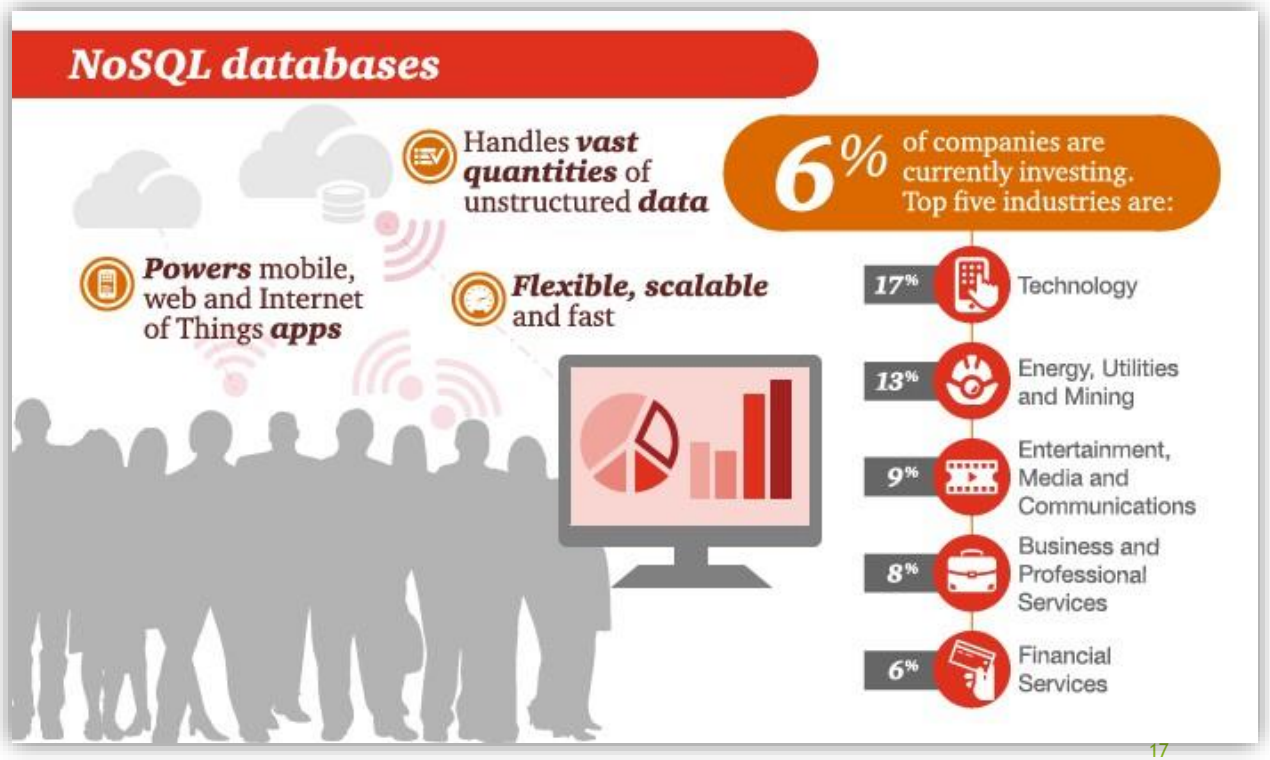
15

Example 2


Wikipedia Pages
Large collection of documents
Combination of structured and unstructured data
Task: Retrieve all pages regarding athletics of Summer Olympic before 1950

16

NOSQL Market



17

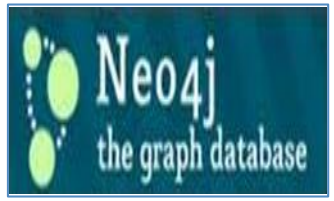
What is NoSQL?

- 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 .

18

NoSQL Database Types

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



19

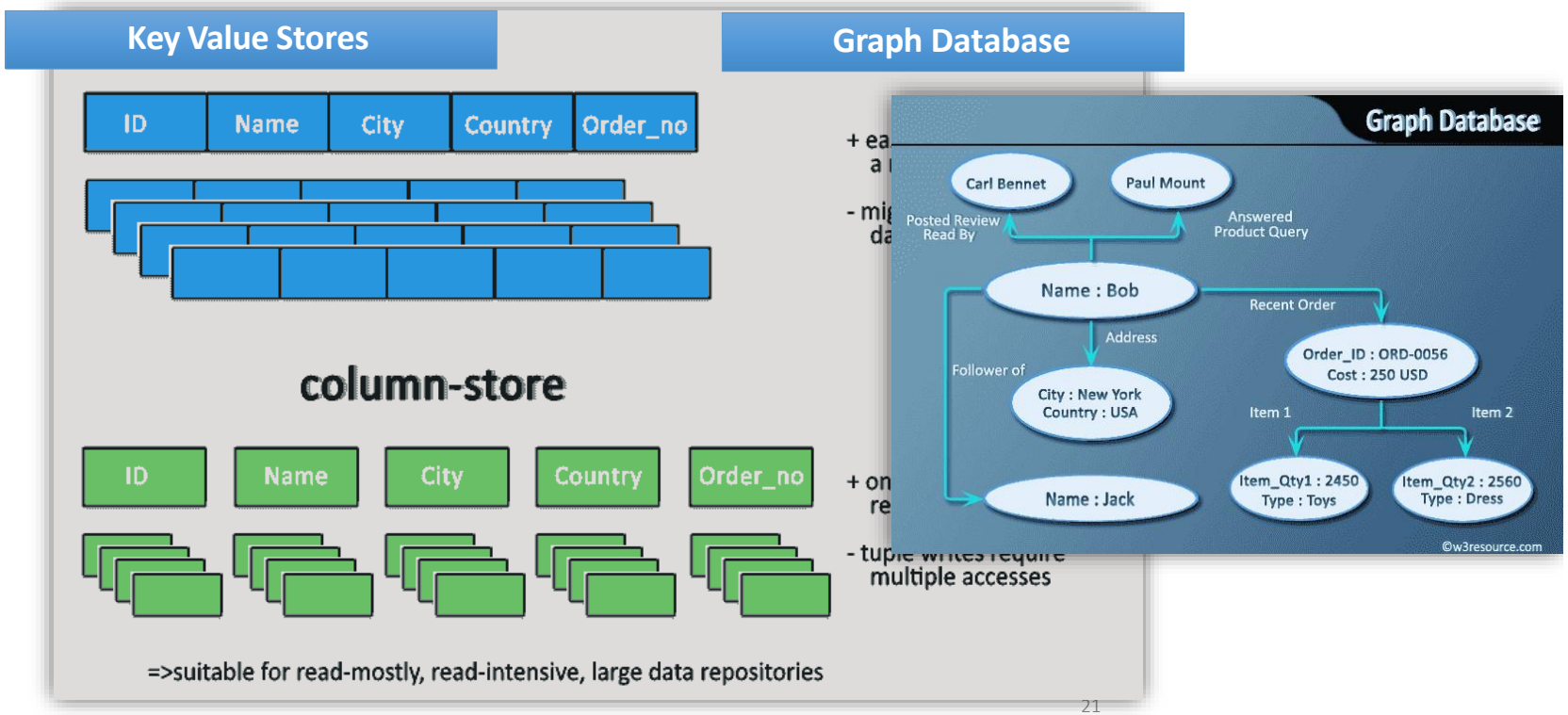
CAP Theorem

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.

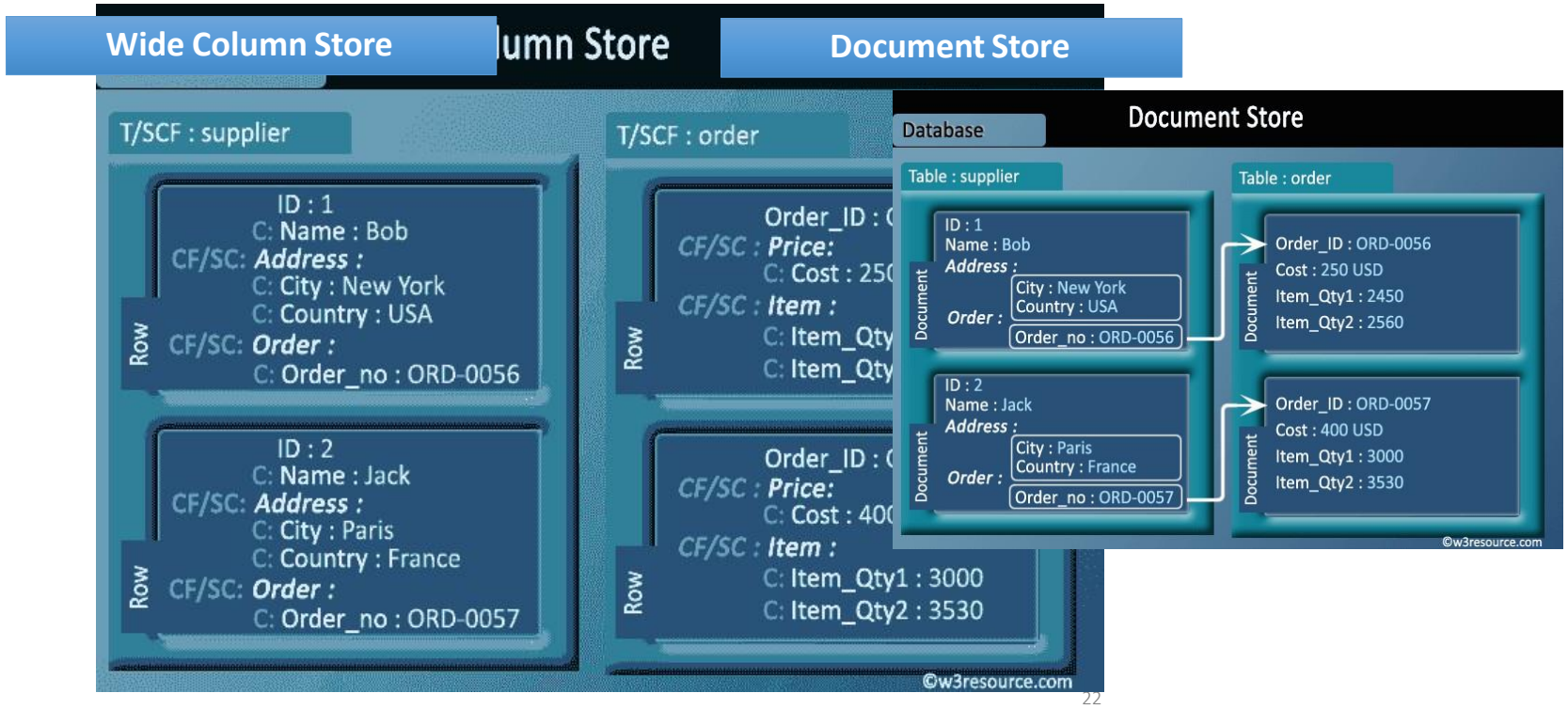
20

14

NO SQL Databases Types



Types of NoSQL Databases



22

16

Production Deployment

Google

Facebook

Mozilla

Adobe

Foursquare

LinkedIn

Digg

McGraw-Hill Education

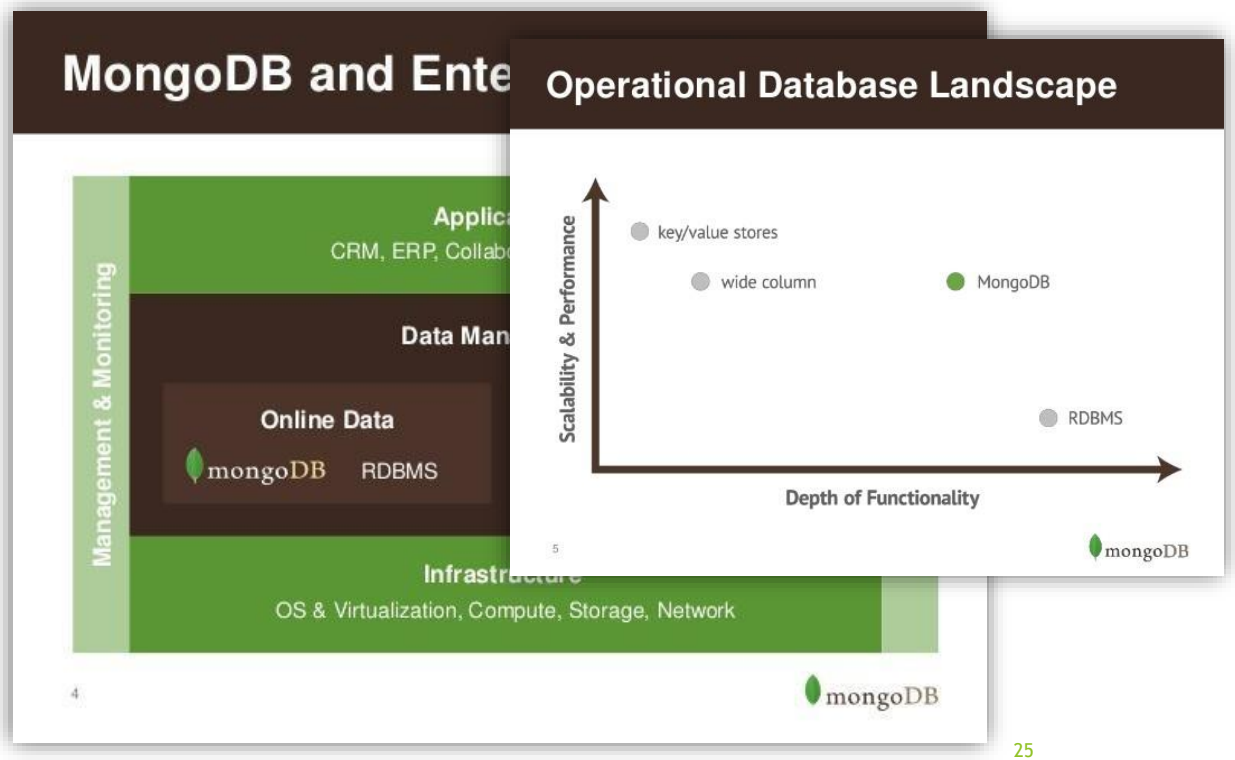
Vermont Public Radio

23

mongoDB®

MongoDB Introduction

MongoDB Philosophy

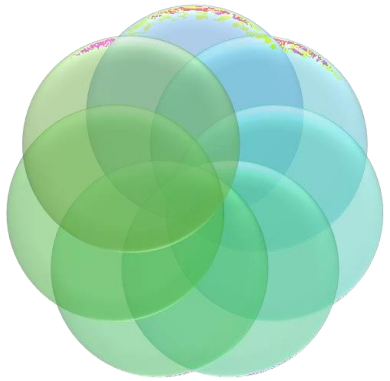


25

What is MongoDB

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

Has driver to all most every popular language programming



Schema-free document database

10gen Inc. also offers professional services around MongoDB

Nosql database

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

Written in C++

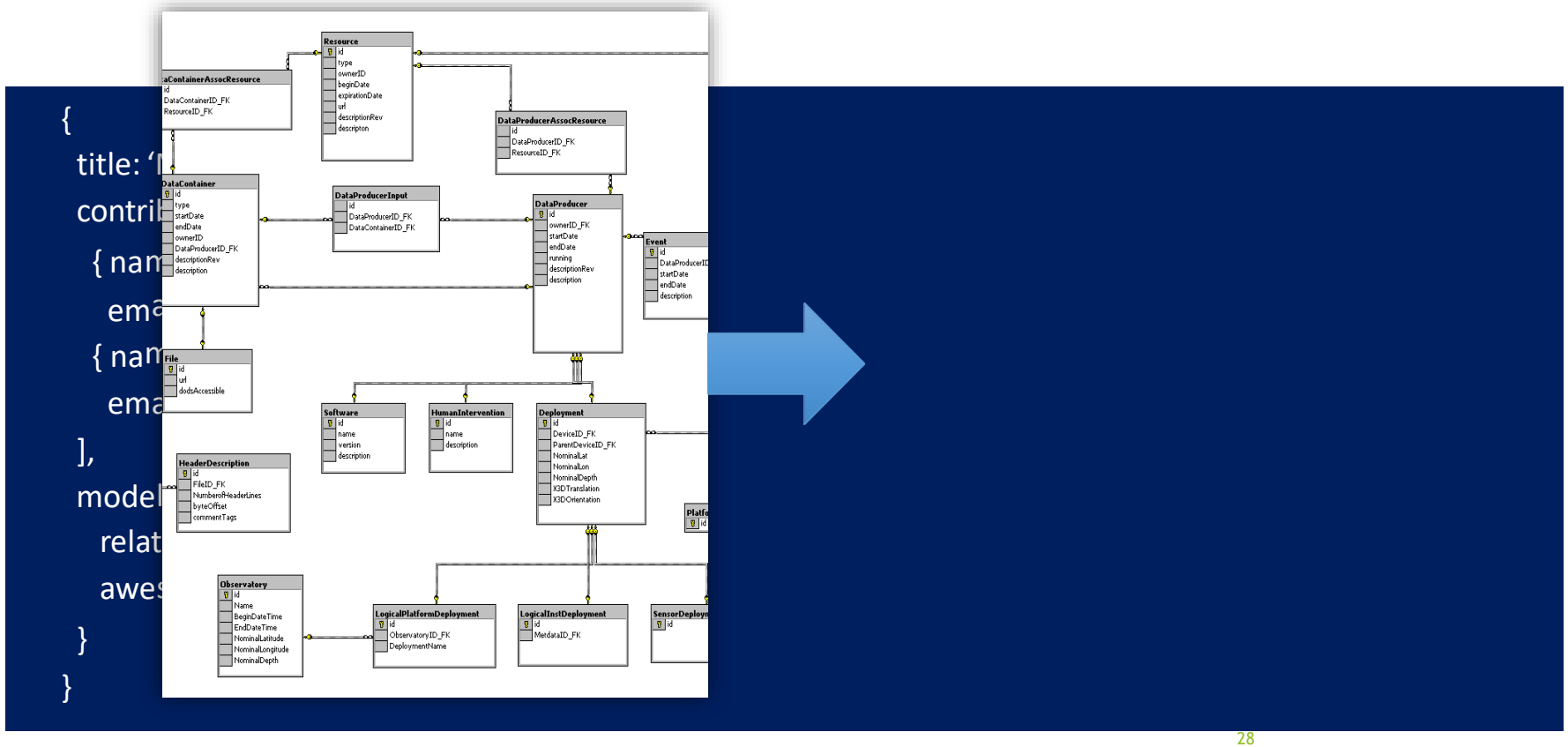
26

MongoDB is:

General Purpose	Rich data model	Full featured indexes	Sophisticated query language
Easy to Use	Easy mapping to object oriented code	Native language drivers in all popular languages	Simple to setup and manage
Fast & Scalable	Operates at in-memory speed wherever possible	Auto-sharding built in	Dynamically add / remove capacity with no downtime

27

MongoDB is Easy to Use

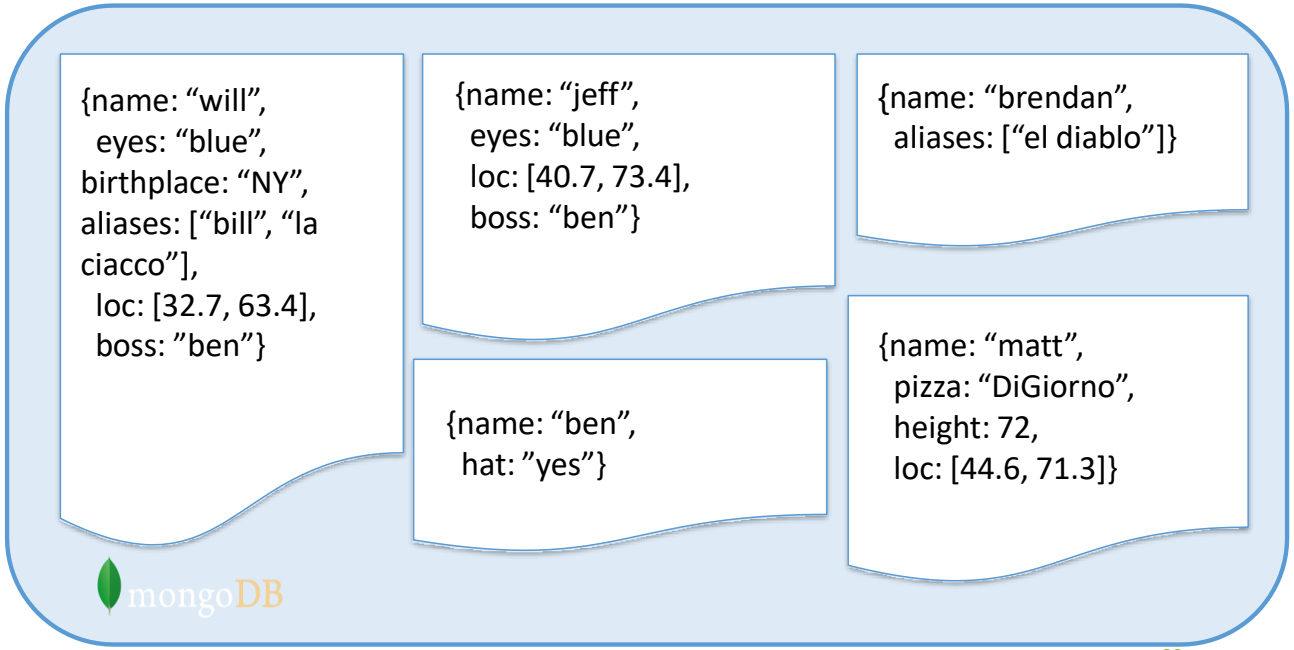


28

Schema Free

MongoDB does not need any pre-defined data schema.

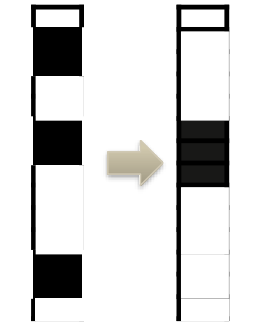
Every document could have different data!



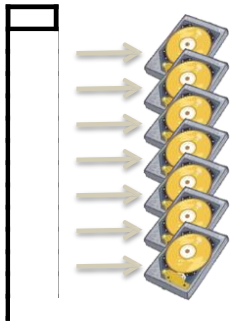
29

MongoDB is Fast and Scalable

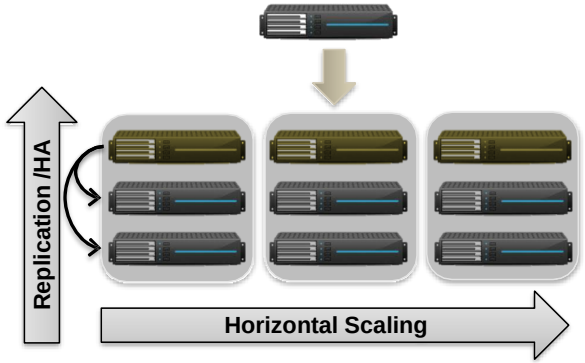
Better data locality



In-Memory Caching

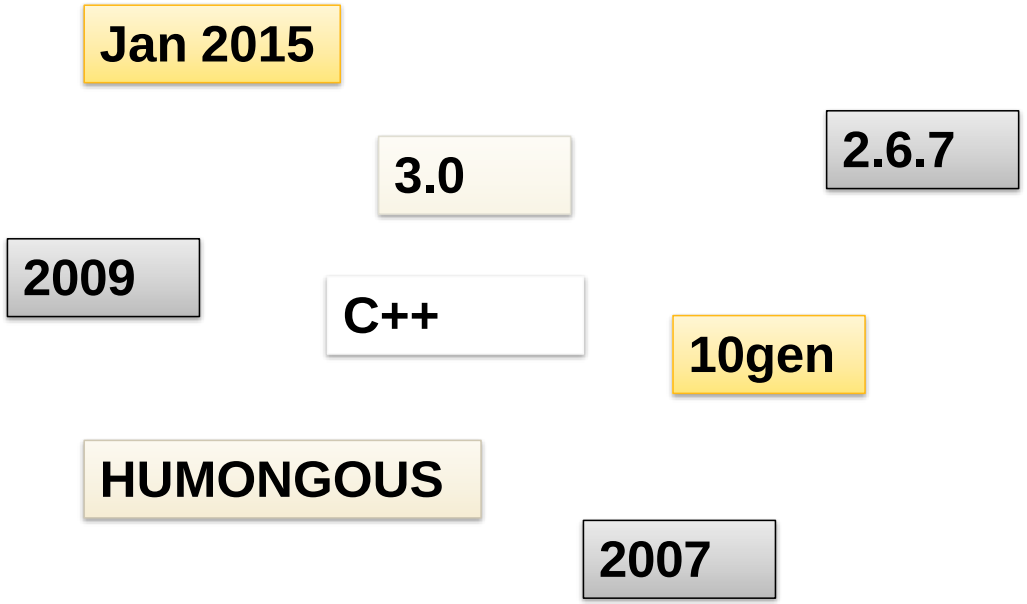


Distributed Architecture



30

Histor
y



31

Features of MongoDB

- Document Oriented Database
- Adhoc queries
- Indexing
- High Performance
- High Availability
- Sharding
- Easy Scalability
- File Storage
- Rich Query Language
- Load Balancing
- Replication

32

Where should you use Mongodb

- **Big Data**
- **Content Management and Delivery**
- **Mobile and Social Infrastructure**
- **User Data Management**
- **Data Hub**

33

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

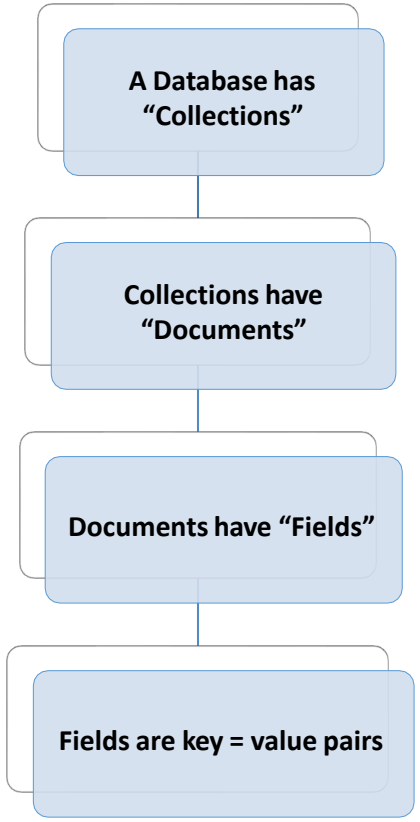
MongoDB to SQL Terminology

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.

35

MongoDB 's Data Model



36

Data Model

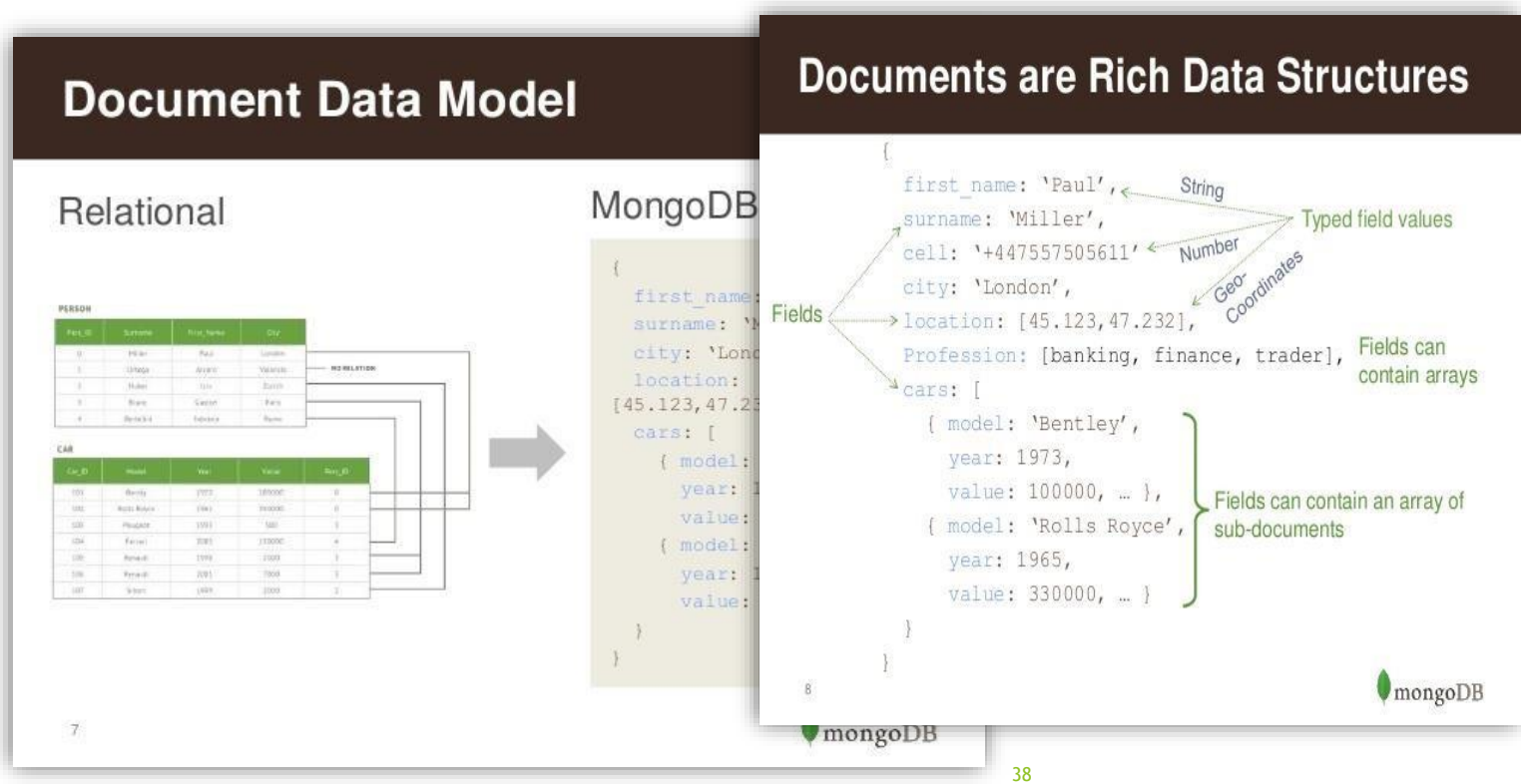
Document based (max 16 MB).

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

Each document stored in a collection.

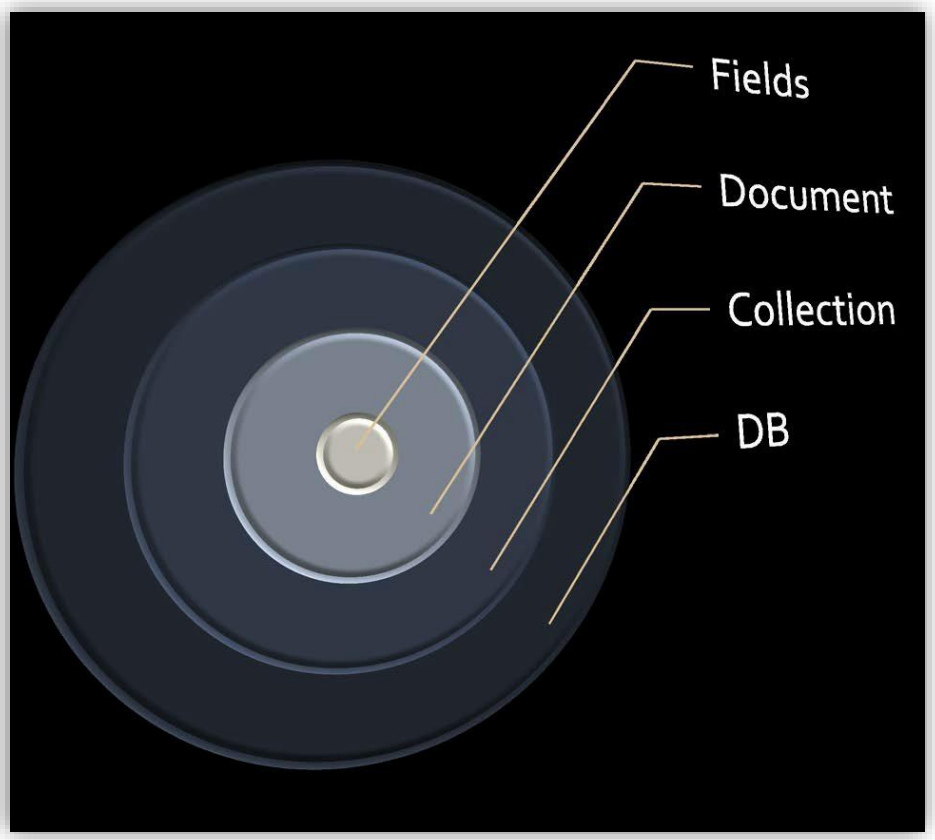
Schema less.

Data Model



38

MongoDB Data Model



39

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.

Database Server and Client

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.

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.

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:\setup\mongodb>mongod.exe --dbpath "d:\setup\mongodb\data"

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

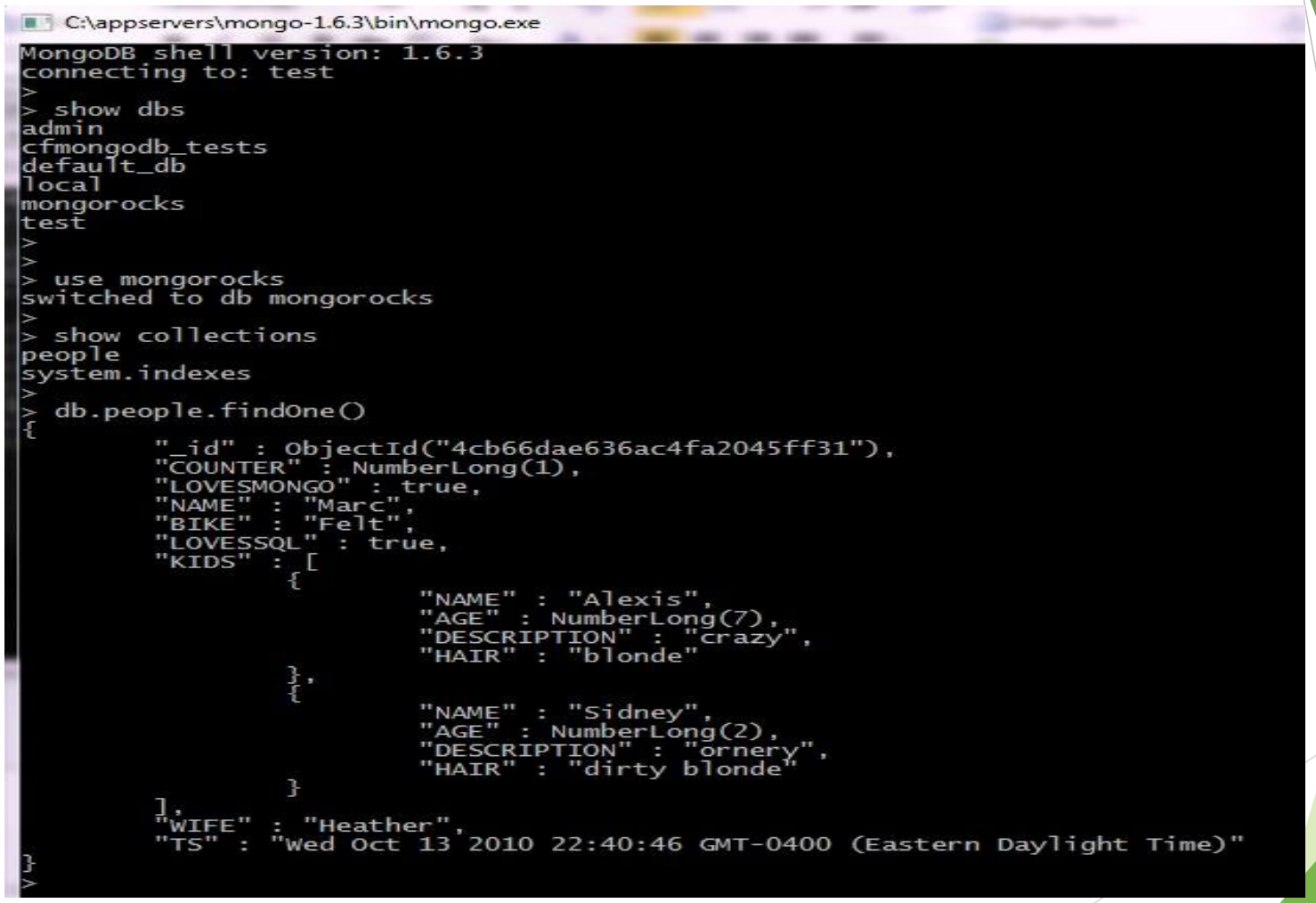
Now to run the mongodb you need to open another command prompt and issue the following command.

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



The Mongo Shell

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