{ topic: "MongoDB & R" }

{ by: "Ryan Nitz" }

Overview

- Installation and setup

- What is MongoDB

- Schema Overview

- Using the shell

- Using R

- Who is 10gen

Installation and Setup

Review installation guide

Guide is missing (run in R shell):

install.packages("rJava",repos='http://cran.stat.ucla.edu')

install.packages("RUnit",repos='http://cran.stat.ucla.edu')

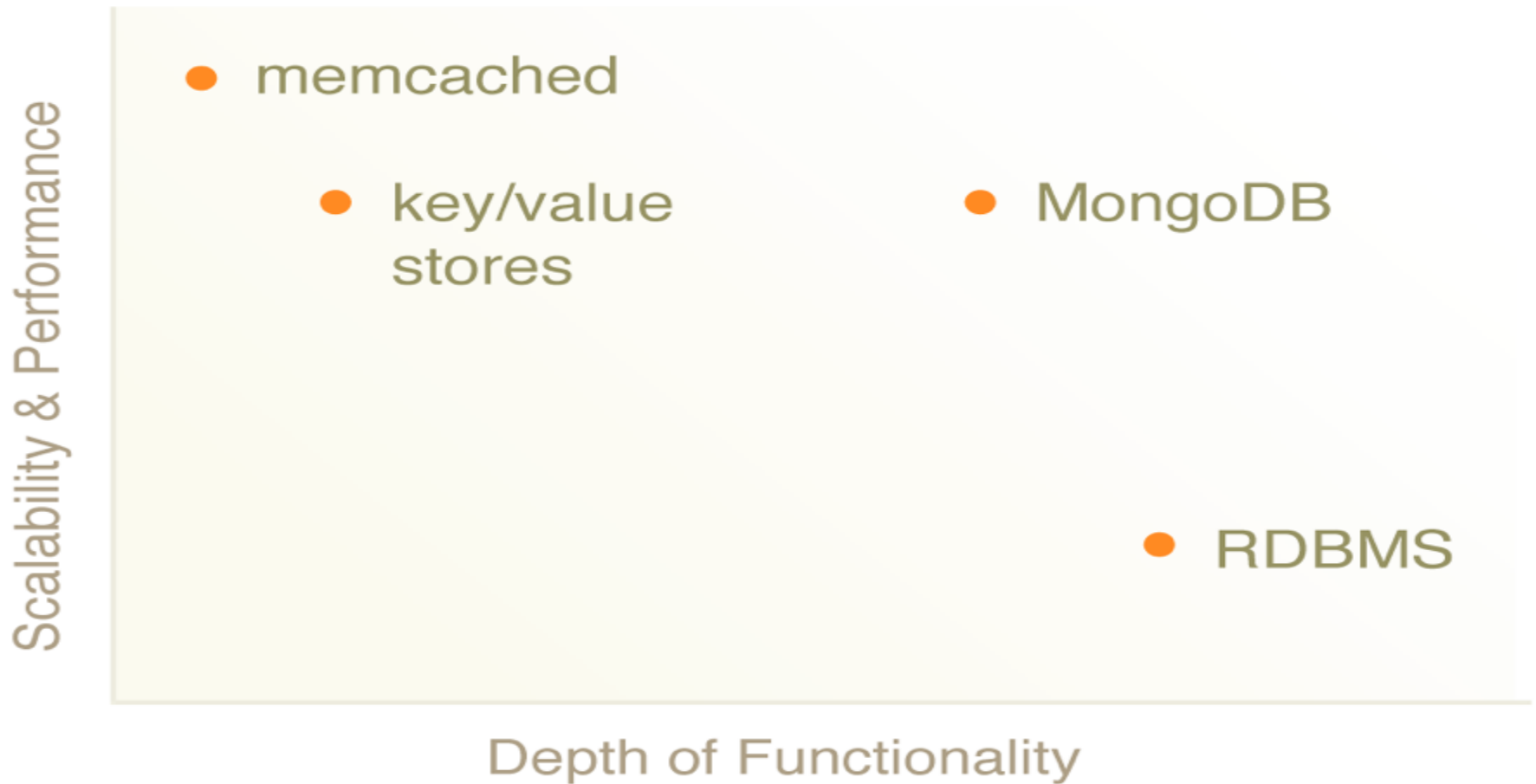install.packages("rjson",repos='http://cran.stat.ucla.edu')

Installation and Setup

- Start MongoDB

- Run the example script

What is MongoDB

- Document-oriented database

- NoSQL
  - (scalable) non-relational next generation operational datastores and databases

  - no joins + light transactional semantics = horizontally scalable architectures

- Schemaless

- Open Source - AGPL (drivers are Apache 2.0)

What is MongoDB

What is MongoDB

RDBMS	Mongo
Table, View	Collection
Row(s)	JSON Document
Index	Index
Join	Embedded Document
Partition	Shard
Partition Key	Shard Key

What is MongoDB

Agility

- Complex data often easier to model as **documents**

- **Schemaless** data model enables faster development cycle

Scale

- Relaxed transactional/query semantics enable easy scale out

Cost

- Cost effectively **manage** abundant data (clickstreams, logs, etc.)

What is MongoDB

Quick questions about MongoDB?

Schema Overview

Data Types

- int
- long
- double
- String (UTF-8)
- binary
- boolean
- regexp
- Date
- ObjectId
- null

Schema Overview

Data Types

- UUID
- MD5
- List/array
- Document (nested)

For more information, see: http://bsonspec.org/

Schema Overview

A quick overview of JavaScript/JSON/BSON
- array == [ ]
- document == { }
- fields in an doc: { enabled : true }
- doc in a doc:

```
{
  name : "Ryan",
  address : { street: '124 Some', zip : 10280, state: 'NY' }
  favoriteNumbers : [ 42, 3 ]
}
```

Schema Overview

- Complex representation of data in a single document

```
{
  historicalValues : [ 1, 2, 3, 4, 5, 6 ]
  features : {
    ironFeatures : [
      { name: 'xyz', value: 132.234243432, type: 3 },
      { name: 'xyz1', value: 424242424242, type : 1 }
    ]
    experimentalFeatures : [
      { name: 'notSure', value: [ 10, 21, 35, 10 ], type: 2 },
      { name: 'likelyGood', value: 132.42424242, type: 3 }
    ]
  }
}
```

Using The Shell

- Enter the shell: ./mongo

- Specific server: ./mongo localhost:27018

Using The Shell

- List database: show dbs

- Select a database: use DATABASE_NAME

- Show collections: show collections
  - show tables works too :-)

- Find one document
  - db.COLLECTION.findOne()

- Find one with query params
  - db.COLLECTION.findOne({ user: 'Ryan' })

Using The Shell

- Find multiple documents
  - db.foo.find()

- Find multiple documents (with query)
  - db.foo.find({ type : 1 })

- Find multiple documents (query & limit)
  - db.foo.find({type: 1}).limit(10)

- Find multiple documents (query, limit, sort)
  - db.foo.find({type: 1}).sort({name: -1}).limit(10)

Using The Shell

- Find multiple documents (clean display)
  - db.foo.find({type: 1}).limit(10).forEach(printjson)

- Inserting a document
  - db.foo.insert({name: 'Ryan', location: 'NY' })

- Updating a document
  - db.foo.update({name: 'Ryan'},
                       {location: { $set: 'CA' }, false, false}

Using R

- library("RMongo")

- Open the connection
  - mongo <- mongoDbConnect("dataGotham", "localhost", 27017)

- Close the connect
  - dbDisconnect(mongo)

- Query:
  - resultDoc <- dbGetQuery(mongo, 'gorillaDataResults', toJSON(list('_id'='testid')))

Who is 10gen

- Founded by Dwight Merriman and Eliot Horowitz in 2007

- Owns the copyrights, IP and sponsors the MongoDB project

- Currently ~ 160 employees
  - Adding another 100 in the next 12 months

- Used by all - startups, governments, enterprises, financial institutions

Who is 10gen

- Here to stay - recently raised $42mm ($73mm to date)

- Dedicated to growing the open source community

- Office hours every Wednesday from 16:30 - 18:30

Thanks!

- ryan@10gen

- https://github.com/rgnitz

- https://github.com/deftlabs