

R user conference Korea 2015

# USing R with MongoDB

김민준

SK c&c

# Tables

1. MongoDB?
2. Rmong & RMongodb
3. Why R + MongoDB
4. Tips for R + MongoDB

I am...

7월 10일 (SK c&c)

cloud computing

Infra Service Team 1

MongoDB Korea, R Korea



**Table 1. MongoDB...?**

# MongoDB 첫 질문?

- Mongoi 사람들이 만들었나요?
- MongoDB 만든 곳의 대주주가 Mongoi 사람...?
- 알고보니 MongoDB 만든 사람이 Mongoi 후손...?

# What is MongoDB?

“humongous” =  $\pi\tau H\dot{\iota}\tau$



huMongousDB

# What is MongoDB?

- Document Oriented Database
- Replication / Sharding
- Multiple Transaction(cRUD)
- Map-Reduce also possible
- Latest **Stable** version is 3.0.3(May 2015, bug hunting)



# Question?

“몽고DB 쓰지 마세요”...왜?

 좋아요 220  트윗 80  +1 23  미투 2

NoSQL을 대표하는 오픈소스로 각광받고 있는 '몽고DB'를 사용하지 말라는 얘기가 고기  
말 엔터프라이즈 소프트웨어 개발자 모임인 인포메이션큐(InfoQ)를 시작으로 몽고DB  
해커뉴스를 비롯한 다양한 사이트에서 몽고DB 성능을 문제 삼는 개발자들이 늘어나는  
고DB의 데이터 '안정성'을 문제 삼으며, 몽고DB가 명성에 비해 기능이 떨어진다고 지적

몽고DB는 스키마 제약 없이 키와 밸류로 이뤄진 도큐먼트를 저장하는 오픈소스 NoSQL  
약이 없기 때문에 좀 더 유연하게 데이터를 처리할 수 있다는 장점을 갖고 있다. 1천만 개  
포스퀘어가 몽고DB를 도입해 데이터를 처리하기 시작하면서 널리 알려지기 시작했다.

 **mongoDB**  
(name: "mongo", type: "DB")

 **MONGO DOCS**  
MongoDB Documentation »

 **TRY IT OUT**  
Try The Online Shell »

 **DOWNLOADS**  
Download MongoDB »

 **DRIVERS**  
Get The Latest Drivers »

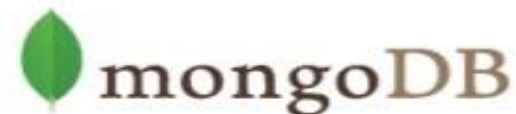


# Answer!

## MongoDB, 1억5천만달러 투자유치로 텀블러보다 높은 12억달러 가치 입증

2013년 10월 7일 by Hazel CATEGORY: NEWS&ISSUE, 해외 | 1 Comment

Bloomberg가 온라인 데이터베이스 관리 회사 MongoDB가 알려지지 않은 여러 금융 서비스사로부터 1억5천만달러(1,600억원)의 투자를 유치한 사실을 전했다. 이는 12억달러(1조3천억원)가치의 면모를 내세움과 동시에 뉴욕시 출신 스타트업으로써 탑 클래스에 진출했음을 알리는 계기이다.



공개된 회사 가치 금액이 그대로 유지된다면, MongoDB는 야후가 금년 10억달러(1조700억원)에 인수한 블로그 서비스 Tumblr, 그리고 최근 1억5천만달러(1,600억원)의 투자유치에 성공한 10억달러 가치의 쇼핑 사이트 Fab보다 더 큰 가치를 지닌셈이다.

MongoDB는 Goldman Sachs, MetLife 등 600개가 넘는 기업의 데이터를 관리중인 빅데이터 데이터베이스 제공사이다. MongoDB의 소프트웨어는 대량의 클라우드 서비스의 백본의 역할을 감당한다. MongoDB는 서버 당 5,000달러(530만원)이라는 비용을 받고 초기 스타트업들이 파산하지 않을 수 있도록 데이터 베이스 문제를 해결해준다.

# Growing UP!

## DB-Engines Ranking

The DB-Engines Ranking ranks database management systems according to their popularity. The ranking is updated monthly.

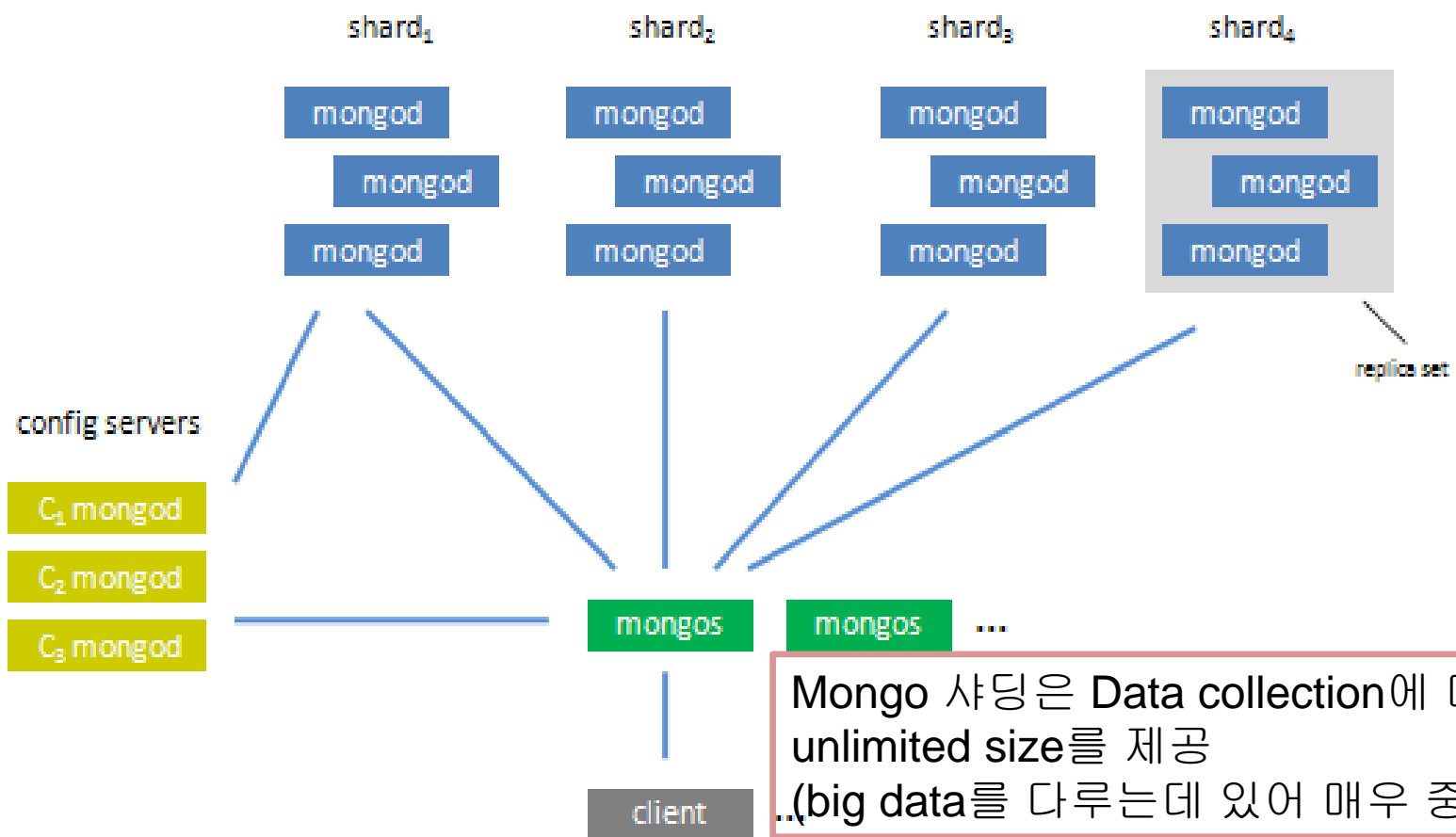
Read more about the [method](#) of calculating the scores.



277 systems in ranking, June 2015

Rank			DBMS	Database Model	Score		
Jun 2015	May 2015	Jun 2014			Jun 2015	May 2015	Jun 2014
1.	1.	1.	Oracle	Relational DBMS	1466.36	+24.26	-34.56
2.	2.	2.	MySQL	Relational DBMS	1278.36	-15.91	-31.20
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1118.05	-12.98	-105.74
4.	↑ 5.	4.	PostgreSQL	Relational DBMS	280.90	+7.39	+40.92
5.	↓ 4.	5.	MongoDB +	Document store	279.05	+1.73	+47.61
6.	6.	6.	DB2	Relational DBMS	198.70	-2.35	+0.67
7.	7.	7.	Microsoft Access	Relational DBMS	146.49	+0.91	+4.13
8.	8.	↑ 9.	Cassandra +	Wide column store	108.91	+2.36	+27.06
9.	9.	↓ 8.	SQLite	Relational DBMS	107.97	+2.81	+18.79
10.	10.	↑ 12.	Redis	Key-value store	95.49	+0.76	+30.36

# MongoDB 구조 간단 review -1



# MongoDB 구조 간단 review - 2

## ■ mongod

- 데이터를 저장, 관리하는 서버(복제 정책 적용 가능)

## ■ mongos

- client의 요청 받아 환경 설정 서버의 partitioning 정보를 참고해 적절한 데이터 서버로 요청을 포워딩

## ■ config Server

- sharding에 대한 환경 설정 서버
- partitioning에 대한 정보를 관리

# MongoDB Benefit

Elements of MongoDB fit for cloud Environment

- Open Source
- Easy Scalability(Scale out), Easy Install, Easy use
- Scheme-less
- High Availability

# MongoDB use (이후)

- facebook, google, twitter
- MetLife, foursquare, NewYork Times, Forbes
- 이외에도...

<http://www.mongodb.com/who-uses-mongodb>



해와 사르게 불타는다면...



# MongoDB use (국내)

- FIFA Online 2
- Daumkakao 'My Agora'
- LG U+ 그룹웨어, 패턴 분석 처리 시스템
- 인터파크, 상품 검색 및 채팅 서비스 시스템

# MongoDB use (국내)

 아고라







 축구를 즐기는 젊.대.기.준  
**FIFA ONLINE 2**

 **LG U+**

 Shopper's Heaven  
**INTERPARK**



시간이 없어서 나머지는 다음 기회에

...

직접 해보시려면...

직접 해보시려면...

## Download and Run MongoDB Yourself

[Current Release](#)


[Previous Releases](#)

[Development Releases](#)


### Production Release (3.0.0)

3/3/2015 [Release Notes](#) [Changelog](#)

Download Source: [tgz](#) | [zip](#)

 Windows

 Linux

 Mac OS X

[Solaris](#)

VERSION:

Windows 64-bit 2008 R2+ ▼

INSTALLATION PACKAGE: [Installation Instructions](#) [View Build Archive](#)

Ⓢ DOWNLOAD (MSI)

MongoDB Download link



귀찮음의 시작..?

귀찮은시작..?



[Docs](#) [Try It Out](#) [Downloads](#) [Community](#)

## A MongoDB Shell in your browser Just enough to scratch the surface.

Welcome to a self guided tour of the MongoDB shell.

Get started with ``help``

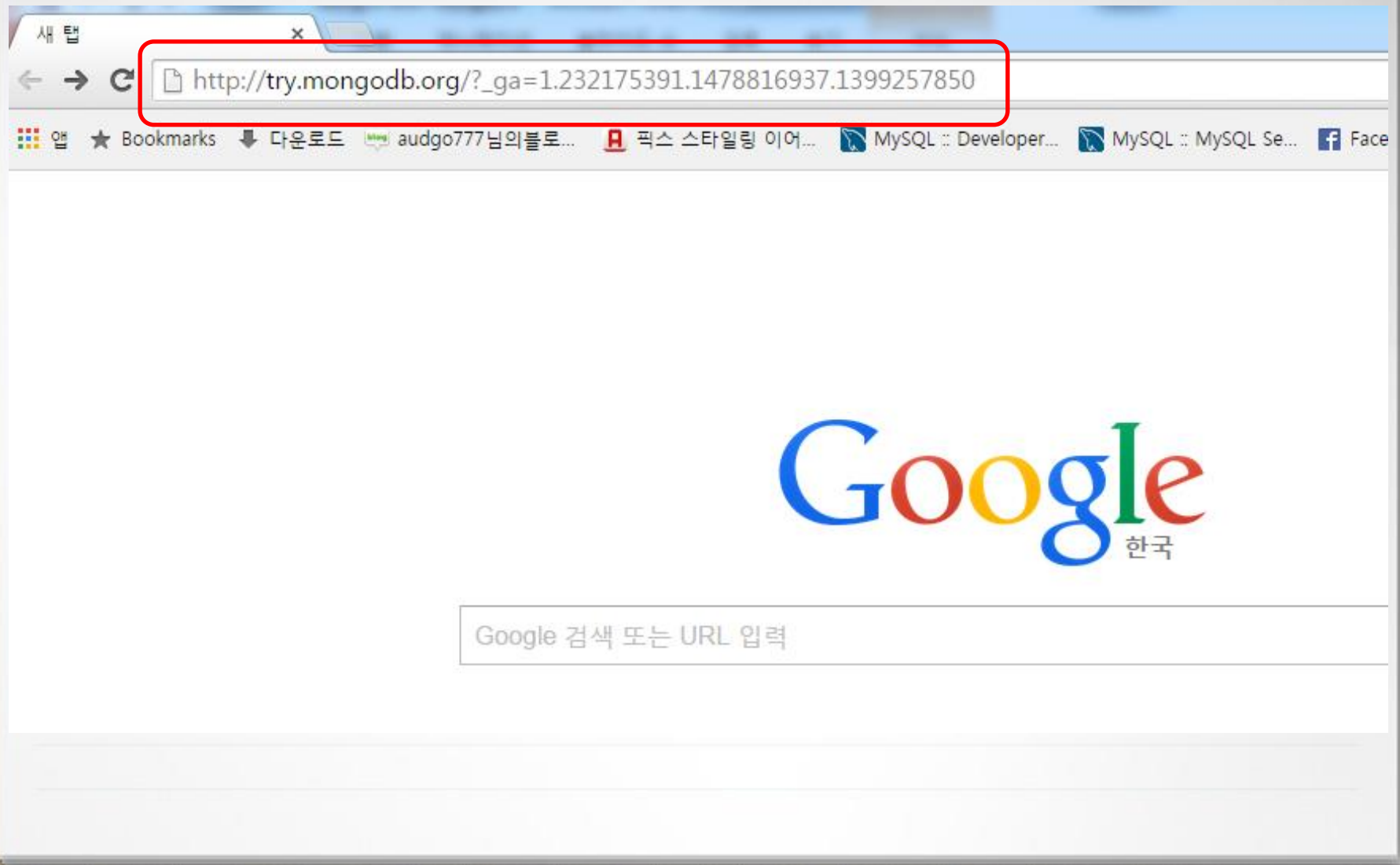
To try out an interactive tutorial type ``tutorial``

>

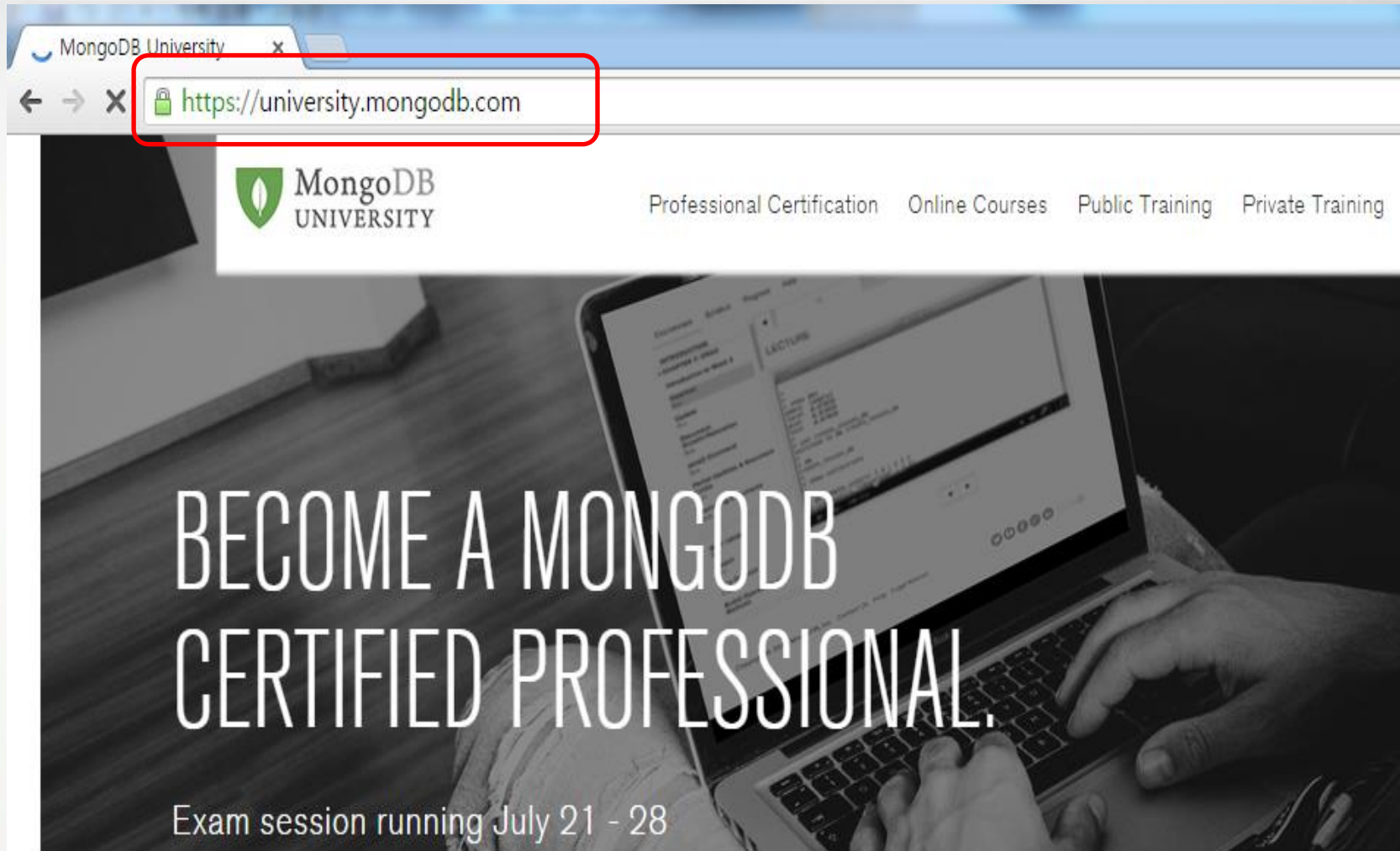
MongoDB Shell link

[http://try.mongodb.org/?\\_ga=1.232175391.1478816937.1399257850](http://try.mongodb.org/?_ga=1.232175391.1478816937.1399257850)

귀찮은시작..?




귀찮은시작..?



MongoDB University

← → × <https://university.mongodb.com>

 **MongoDB**  
UNIVERSITY

Professional Certification   Online Courses   Public Training   Private Training

**BECOME A MONGODB  
CERTIFIED PROFESSIONAL.**

Exam session running July 21 - 28

이런 사...

직접 공부해 보려면...



공부해보고싶다면...

<https://university.mongodb.com>

A banner for MongoDB University featuring a background image of hands typing on a laptop. The laptop screen displays a 'LECTURE' page with a table of contents. The text 'BECOME A MONGODB CERTIFIED PROFESSIONAL.' is overlaid in large white letters. Below this, it says 'Take the next exam starting April 21, 2015.' and a green button with the text 'REGISTER NOW »' is visible. The top navigation bar includes the MongoDB University logo, links for 'Professional Certification', 'Online Courses', 'Public Training', and 'Private Training', and a green button labeled 'MY COURSES'. The text 'InBum' is in the top right corner.

**MongoDB UNIVERSITY**

Professional Certification Online Courses Public Training Private Training **MY COURSES**

**BECOME A MONGODB CERTIFIED PROFESSIONAL.**

Take the next exam starting April 21, 2015.

**REGISTER NOW »**

MongoDB education link

<https://university.mongodb.com/>

공부할 수 있는 것들...

## ONLINE COURSES



M101P: MongoDB for  
Developers



M101JS: MongoDB for  
Node.js Developers



M101N: MongoDB for .NET  
Developers



M101J: MongoDB for Java  
Developers



M102: MongoDB for DBAs



M202: MongoDB Advanced  
Deployments and Operations

MongoDB education link

<https://university.mongodb.com/>

# 공부하다 얻을 수 있는 것...?

## MY COURSES

### COURSES

Free online seven week courses with fixed deadlines, periodically offered by MongoDB. In these classes, most of the help comes from fellow students in public forums.



#### M102: MongoDB for DBAs

*Class ended Sep 10, 2013*

Your final grade: **75%**.

[DOWNLOAD CERTIFICATE](#)[Unregister](#)

#### M101J: MongoDB for Java Developers

*Class ended Apr 23, 2013*

Your final grade: **73%**.

[DOWNLOAD CERTIFICATE](#)

# MongoDB 이걸 때 쓰지 마세요!

- 고객이 Oracle이랑 비교할 때(RDBMS와는 태생이 다름)
- 무조건라는 이야기 듣고 고객이 들이대릴 때
- 재정적으로 여유가 있을 때
- License에 대한 이해가 불충분한 경우

# 고객이 말합니다(feat. 불특정 다수의 감님)

- 오픈 소스인데 추가 비용이 왜 필요하죠? 라이선스....? 뭐래..
- 아니 왜 오라클 보다 성능이 안 나오는것 같죠?
- 오라클은 어저귀, MySQL은 저저귀, 나는 왕년에 #\$\$!@\$#\$...







# MONGODB 이럴 때 쓰세요!

- 유연함과 확장이 필요할 때
- log data, SNS data 등을 적재할 때
- 다양한 open source와 연계할 때 (ex. Hadoop)
- 서비스 구성 시 다양한 시도가 필요할 때
- 고객이 open source에 대한 이해가 충분할 때

**Table 2. Rmongo & rmongodb**

# "Rmongo" Package

- MongoDB interface for R
- 인터페이스는 mongo-java-driver의 java call을 통해 제공됨.
- 0.0.25, GPL-3
- R(  $\geq$  2.14.1), java(  $\geq$  1.6), MongoDB(  $\geq$  1.6)

# “Rmongo” : Advantage

- Rmongo는 매우 직관적인 스타일의 패키지
- 그로장기 때문에 이해하기 쉽고, 사용하기 쉬움(상대적으로)

# “Rmongo” : Dis-advantage

- 성능 상의 아쉬움
- 인터페이스가 mongo-java-driver이기 때문에 java에 대한 두려움이 있는 이들에게는 장벽으로 다가올 수 있음  
(단순 사용이 아닌 패키지 구조 확인 및 customizing에 있어서..)

# “rmongodb” Package

- MongoDB interface for R
- 인터페이스는 mongo-c-driver(ver 0.8) 를 이용하여 제공됨
- 1.8.0, Apache License
- R(  $\geq$  2.10), MongoDB, GNU Make



# “rmongodb” : Advantage

- MongoDB의 대부분의 기능을 반영함(Rmongo와 대조됨)
- Good Performance

# “rmongodb” Dis-Advantage

- MongoDB 3.0 이상에서부터는 인증 이슈 존재
- 패키지 유지 문제와 신규 기능에 대한 릴리즈의 어려움
  - Mongo-c-Driver 1.0+ 은 완전히 다른 API 가짐
  - 즉 0.8에서 1.0으로 업그레이드 하려면 거의 처음부터 모든 패키지를 다시 써야 함

# RMongo workflow

```
mongo <- mongoDbconnect("test", "localhost", 27017)
```

```
output <- dbInsertDocument(mongo, "test_data", '{"foo": "bar"}')
```

```
output <- dbGetQuery(mongo, "test_data", '{"foo": "bar"}')
```

```
dbDisconnect(mongo)
```

# rmongodb workflow

```
mongo <- mongo.create(host="localhost", db="test")
```

```
mongo.get.databases(mongo)
```

```
ns="test.test_data"
```

```
buf=mongo.bson.buffer.create()
```

```
Mongo.bson.buffer.append(buf,"foo","bar")
```

```
a=mongo.bson.from.buffer(buf)
```

```
Mongo.insert(mongo, ns, a)
```

```
mongo.find.all(mongo, ns)
```

```
(or mongo.find.one(mongo, "test.test_data", a) )
```

```
mongo.disconnect (mongo)
```

## Table 3. Why R + MongoDB

# R + MongoDB의 매력(1)

- 꾸준히 발전하는 openSource  
(release 주기가 안정적)

- 명확한 역할 구분

(r - visualization, mongodb - backend 저장소)



## R + MongoDB의 매력(2)

- 대규모의 데이터를 다루는 R에게 NoSQL은 매력적인 선택

- 명확한 역할 구분

(r - visualization, mongodb - backend 저장소)

## R + MongoDB의 매력(3)

- 이제 많은 Relational SQL 패키지가 존재.  
(RMySQL, RPostgreSQL, ROracle, RJDBC 등)
- 하지만 이들의 문제는 모든 결과를 R 메모리상에서 read 한다는 것!

## R + MongoDB의 매력(4)

- Rmongo 역시 같은 문제,  
하지만 rmongodb는 그러한 문제에서 자유로움
- 위와 더불어 MongoDB가 bulk insert에서  
탁월한 성능을 보임.

## Table 4. Tips for R + MongoDB

TIPS 1. **mongodb** 설치 서버와

**R** 설치 서버는 구분하는 것이 좋다.

TIPS 2. 예제의 문제가 있다면

config 서버에 설치하는 것도 좋다.



TIPS 3. CPU는 무조건 64bit~!

메모리는 Maximum 구성~!

TIPS 4. 굳이 MongoDB일 필요는 없다.

특징과 목적에 맞는 DB가 필요하다.

TIPS 5. MongoDB는 Schema에서 free...

하지만!!! 특정 주제에 맞춰 분석을 하려면

스키마에 대한 정의와 제한은 필요하다.

TIPS 6. MongoDB의 release에 따른  
R의 version 관리가 필요함.

deprecated 된 함수들이 종종  
발생할 수 있으며,

호환성 문제도 발생 가능

TIPS 7. MongoDB 상에 concurrent하게  
데이터를 read/write 하고 싶다면...  
mongodb.splitvector()를 써서  
chunk 단위로 collection을 분리하고  
각 chunk 상에 mclapply() 적용

맞음말

물론 여전히 분석에  
focusing을 맞춰야 하지만..



매일매일

MONGODB 를 비롯한 NOSQL등은  
확장성과 유연성을 제공할 수 있는 기반.

매트말

축구는 골을 넣어야 이기는 경기지만,  
수비가 잘되면 우승까지 노릴 수 있습니다.

맞음말

NOSQL 활용을 통해 R을 사용할 수 있다면,  
여러분은 든든한 리베로를 얻을 수 있을 것입니다.

맞춤말

그리고 그 리베로는 여러분들 좀 더  
높은 단계로 이끌어 줄 것 입니다.

감사합니다!

inbum85@gmail.com

Special Thanks to

Mr. Selivanov