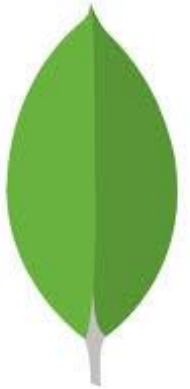
MongoDB & Neo4j

CHETAN PANGAM

SAI LALITHA SREE VUDDAMARRY

Professor: Abhishek Gupta

MongoDB



- Developed by MongoDB Inc. and is free and open source.
- Document Oriented Database
- Three main terms in MongoDB are
  - Database
    - Physical Container for collections
  - Collections
    - Contains Documents
      - Different fields
  - Document
    - Field-value pairs
    - The values of fields may include other documents, arrays, and arrays of documents.
    - _id is the Unique identifier for each document

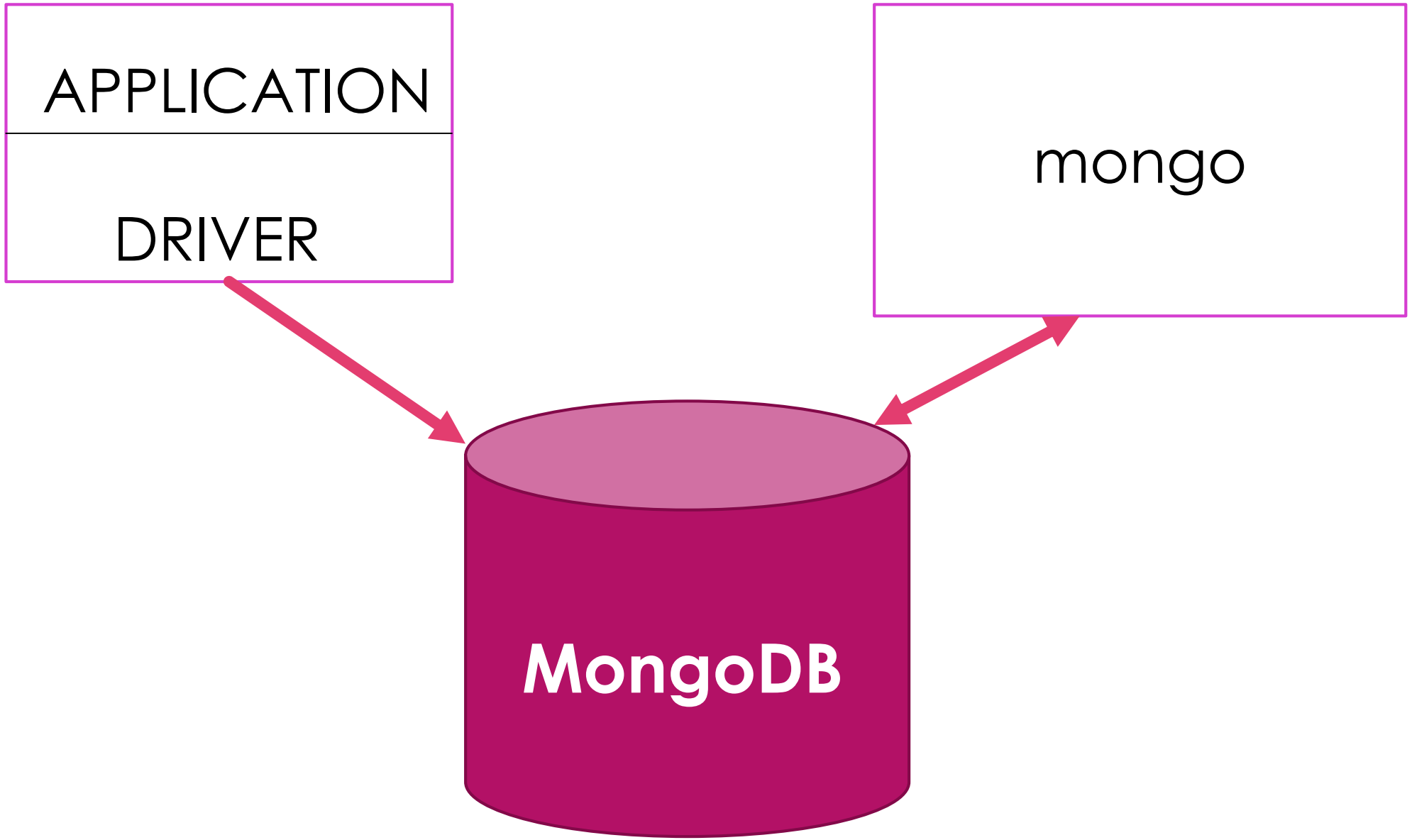

```
{
  "_id" : ObjectId("54c955492b7c8eb21818bd09"),
  "address" : {
     "street" : "2 Avenue",
     "zipcode" : "10075",
     "building" : "1480",
     "coord" : [ -73.9557413, 40.7720266 ]
  },
  "borough" : "Manhattan",
  "cuisine" : "Italian",
  "grades" : [
     {
        "date" : ISODate("2014-10-01T00:00:00Z"),
        "grade" : "A",
        "score" : 11
     },
      "restaurant_id" : "41704620"
}
```

Why use MongoDB?

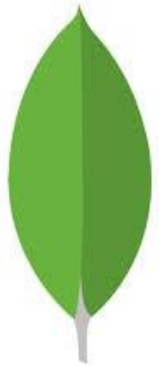
- Document-oriented
- Ad hoc queries
- Indexing
- Replication
- Load balancing

Where to use MongoDB ?

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


APPLICATION
DRIVER
mongo
MongoDB

Installation of MongoDB (on Mac OS X)

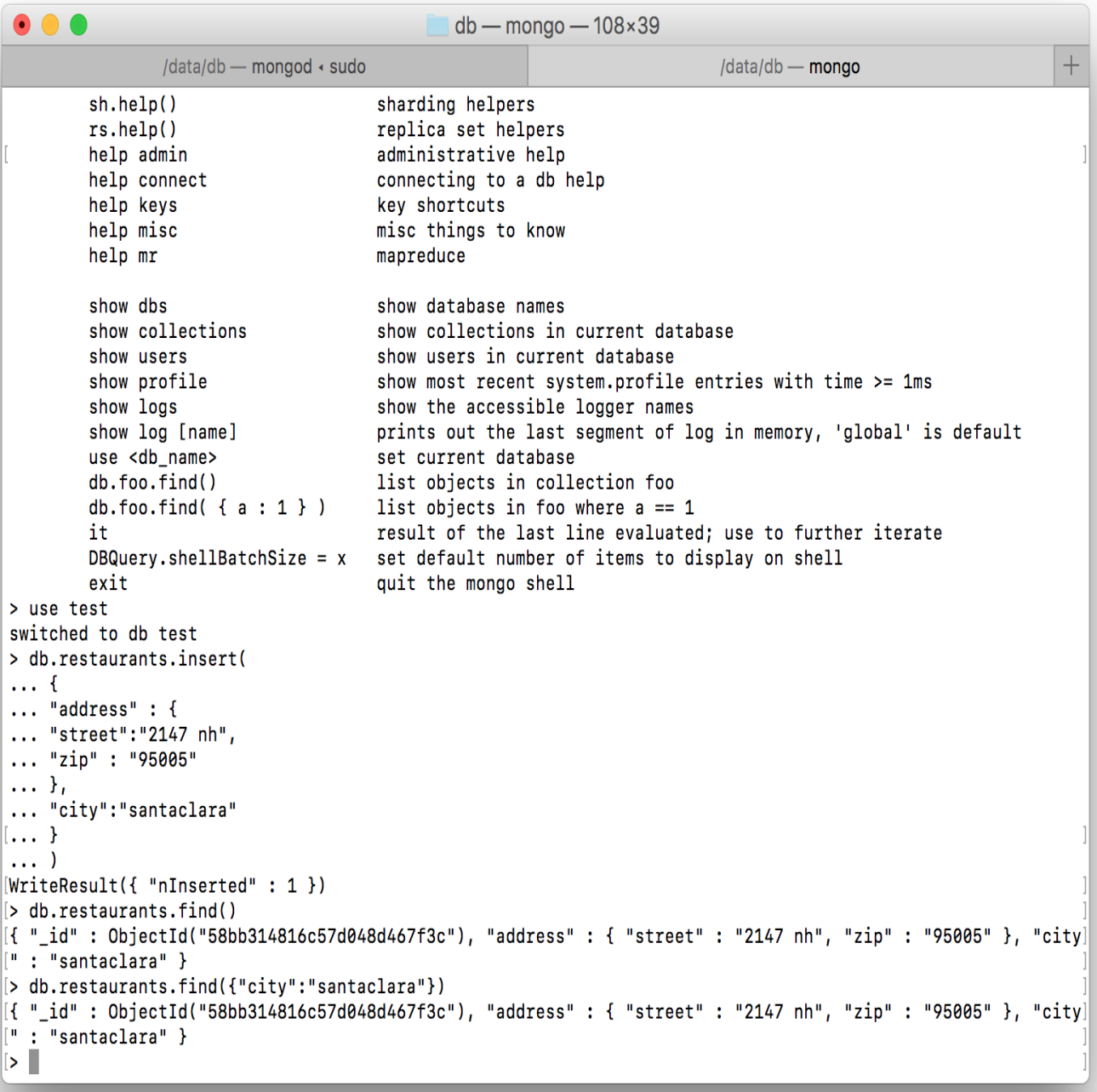


- To install the MongoDB binaries, issue the following command in a system shell:
  brew install mongodb
- Then make a new directory using the following command
  mkdir -p /data/db
- Change the path to that directory
- Run mongo command. mongo is an interactive JavaScript shell interface to MongoDB,
- In the mongo shell connected to a running mongod instance, switch to the test database.
  use test
- To insert a document, run the following.

```
db.restaurants.insert(
{
        "address" : {
        "street":"2147 nh",
        "zip" : "95005"
        },
        "city":"santaclara"
}
)
```

This return a WriteResult() object.
A wrapper that contains the result status of the mongo shell write methods

To return all documents in a collection, call the find() method
db.restaurants.find()



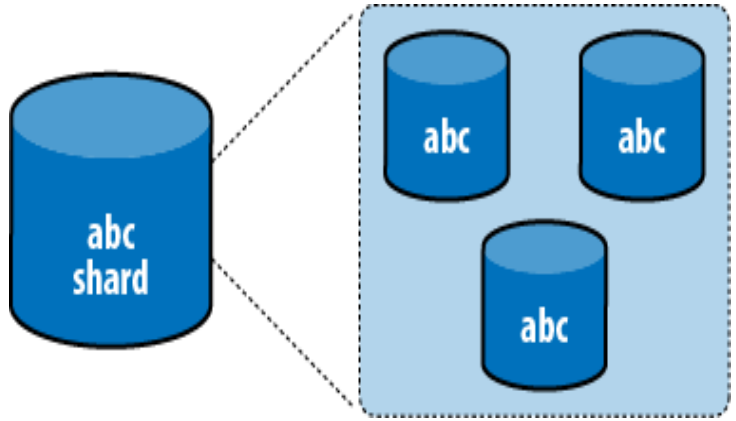
Similarly, we can Update and remove data

```
db.restaurants.update(
   { “zip" : “95050" },
   {
     $set: { “city": “New York" },
     $currentDate: { "lastModified": true }
   }
)
```

```
db.restaurants.remove( { “city": “New York" }
```

SHARDING

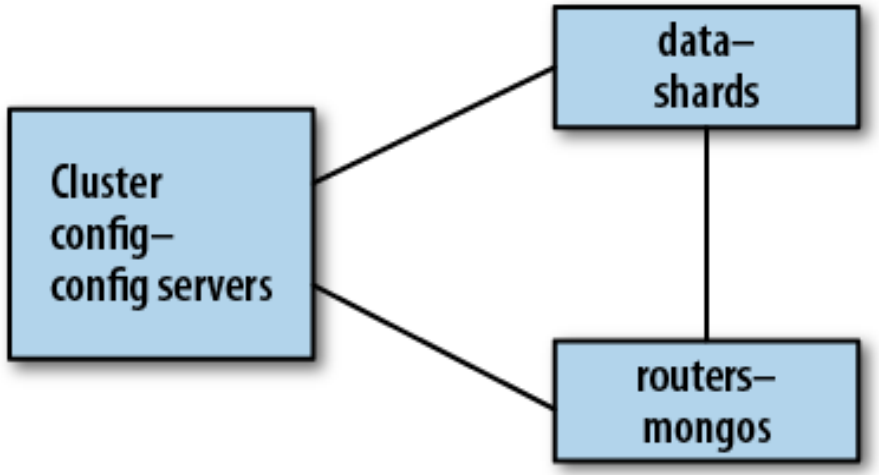
- Method MongoDB uses to split large collections of data across many servers – Clusters.
- Shard - servers in a cluster that are responsible for some subset of data
  - Consists of many servers – each server has an identical copy of data
  - Moves subsets of data from shard to shard –Eg: username field.



Anatomy of a Cluster

Three types of processes

- Shards - storing the data
- mongos processes – Routing requests to the correct data
- config servers – keeps track of the cluster's state



MongoDB Atlas

- Database-As-A-Service
- Makes deployment and management of MongoDB easier, with managed hardware provisioning, failure recovery and backups.
- Easily provision the MongoDB clusters across the public cloud – AWS
- Building a Cluster
  - Specify the AWS region, instance size, storage, payment details.
- Pricing
  - Depends on the Instance size, Replication factor (i.e. number of nodes required), backup.
- Connection
  - Setup IP whitelists for outside connectivity into the cluster.
- Data Import
  - mongodump can export data from either mongod or mongos instances.

Build Your New Cluster

Cluster Name
Name your cluster wisely, you cannot change this later.

Cluster0

MongoDB Version
All clusters launch with the WiredTiger™ storage engine.

MongoDB 3.2 with WiredTiger™
Change version

AWS Region
Select an AWS region for your cluster. Cross-region clusters are not currently supported. Once your cluster has been deployed, you cannot change its region. Any future clusters created in this group will be deployed to this region as well.

- us-east-1 (N. Virginia)
- us-east-2 (Ohio)
- us-west-2 (Oregon)
- eu-west-1 (Ireland)
- ap-southeast-2 (Sydney)

Instance Size
Select the size of your servers. You can easily upgrade your servers at any time after deploying, with no down-time.

Great for **development environments** *and* **low traffic websites.**

M0	*included:*	**shared** RAM	512 MB storage	FREE	SELECT
M10	*included:*	**2 GB** RAM	10 GB storage	$0.03/HOUR	SELECT
M20	*included:*	**4 GB** RAM	20 GB storage	$0.07/HOUR	SELECT

Perfect for **production environments,** *supporting* **high-traffic applications** *or* **large datasets.**

M30	*included:*	**8 GB** RAM	40 GB storage	$0.18/HOUR	✓ SELECTED
M40	*included:*	**16 GB** RAM	80 GB storage	$0.34/HOUR	SELECT
M50	*included:*	**32 GB** RAM	160 GB storage	$0.66/HOUR	SELECT
M60	*included:*	**64 GB** RAM	320 GB storage	$1.30/HOUR	SELECT
M100	*included:*	**160 GB** RAM	1000 GB storage	$3.01/HOUR	SELECT

CUSTOM STORAGE CAPACITY

40 GB

40 GB INCLUDED | 160 GB | 640 GB | 2 TB | 8 TB | 16 TB

CUSTOM STORAGE SPEED

- **Standard** *120 max/IOPS*
- **Fast** *660 max/IOPS*
- **Fastest** *1200 max/IOPS*

☐ Use encrypted storage volumes

Replication Factor
Select how many copies of your data should exist in your cluster. You can easily change your cluster's replication factor after deploying, with no down-time.

Cluster Overview

Region: N. Virginia	Instance Size: M30
RAM: 8 GB	Disk Storage: 40 GB
Disk Speed: 120 IOPS	Replication Factor: 3
Backup: Enabled	Shards: 1

Pricing
Hourly Cost: $0.54/hr*

Pay-as-you-go! You will be billed hourly and you can terminate your cluster at any time.

* Price estimate above excludes variable data transfer, backup, and taxes.

CONTINUE TO PAYMENT

Cluster Overview

Region: N. Virginia	Instance Size: M30
RAM: 8 GB	Disk Storage: 40 GB
Disk Speed: 120 IOPS	Replication Factor: 3
Backup: Enabled	Shards: 1

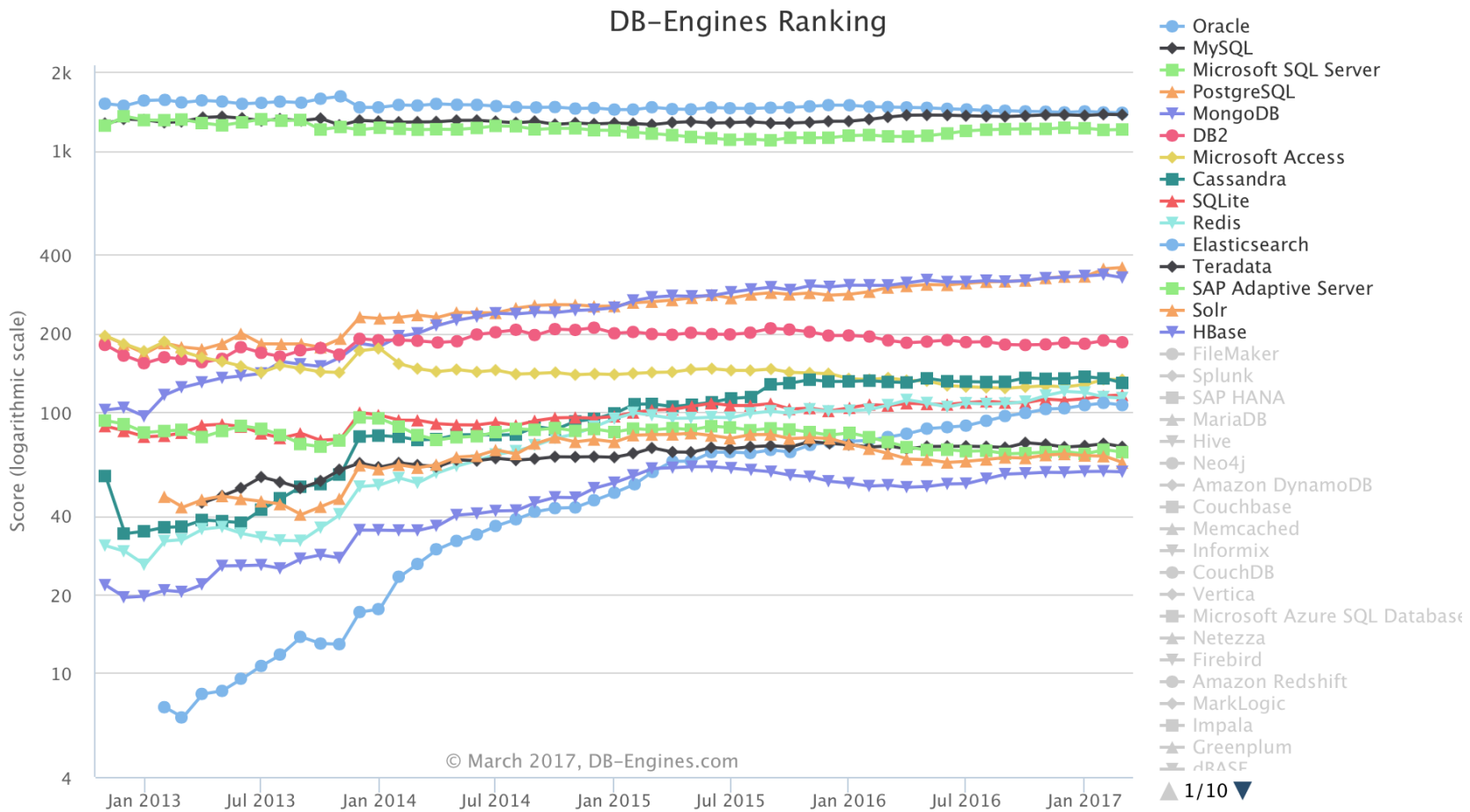
Pricing
Hourly Cost: $0.54/hr*

Pay-as-you-go! You will be billed hourly and you can terminate your cluster at any time.

* Price estimate above excludes variable data transfer, backup, and taxes.

CONTINUE TO PAYMENT

Popularity of MongoDB



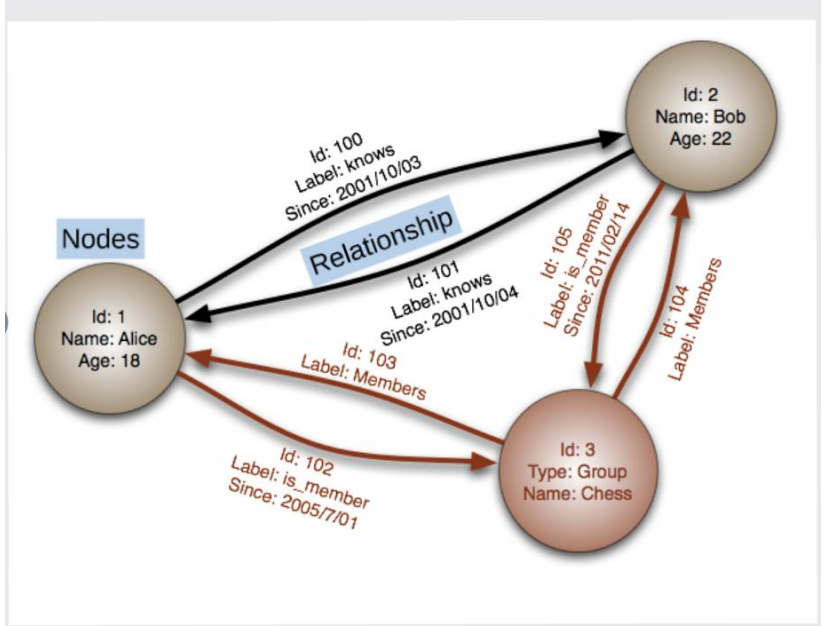
Neo4j

What is Neo4j

- Neo4j is an open source and world's leading graph database management system developed by Neo Technology, Inc.
- Designed for optimizing fast management, storage, and traversal.highly scalable, native graph database.
- Delivers constant real-time performance.
- Build applications to meet today’s evolving data challenges.

GRAPH DATABASE

- Graph Database-Today's world is driven by the connections between Data.
- Big data is dealing with large volume data, but enterprise leaders need more than volume.
- Need a database that control those connections and solution is Graph database.



- It uses graph structures for semantic queries with nodes, edges and properties.
- Nodes linked together directly through edgesnon-relational storage engines with tags.
- Cypher query Language

Features

1. Flexible schema
2. Scaling and Performance
3. Drivers for popular languages and frameworks
4. Cloud ready
5. Powerful Cypher Query language
6. Data Import
7. Hot Backups

Neo4j

Three main primitives in Neo4j:

1. Nodes
2. Relationships
3. Properties.

- Node and relationship identifiers are 35 bits in length
- Property record size is 41 bytes.
- Properties could be attached to both nodes and relationships.
- Relationship types have 2-byte identifiers.

Hosting Neo4j in Cloud

- Docker image and different Neo4j cloud-hosting partners.
- Neo4j Docker images, both for Neo4j Community as well as Enterprise Edition having more than 1M+ pulls.
- Neo4j partners that provide hosting of Neo4j instances in the cloud.
    - - GraphGrid
    - - GrapheneDB
    - - GraphStory

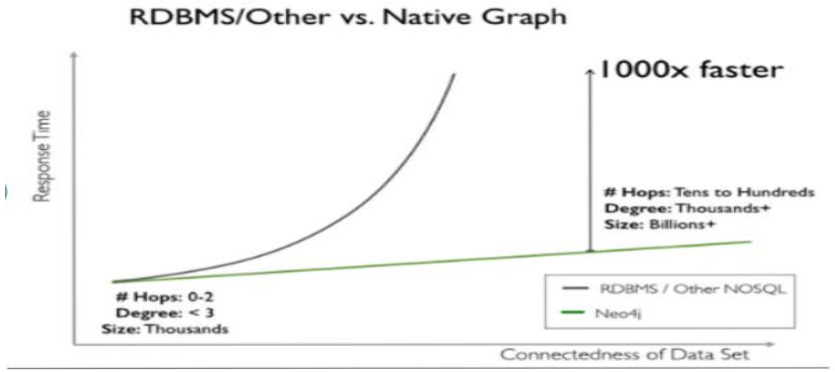
RDBMS vs Neo4j

- Data Storage
- Data Modeling
- Query performance

Performance comparison:
neo4j vs RDBMS

Supports index-free adjacency

Depth	RDBMS execution time (s)	Neo4j execution time (s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000



RECENT TRENDS

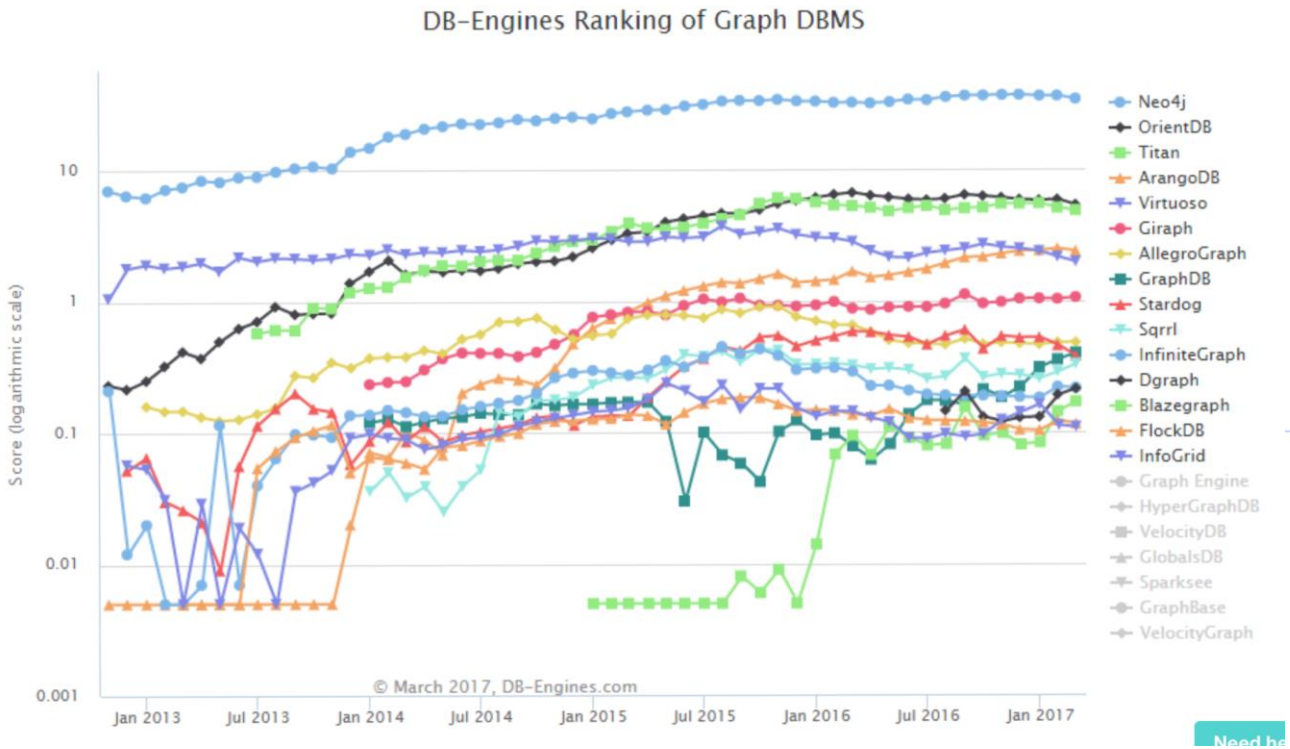
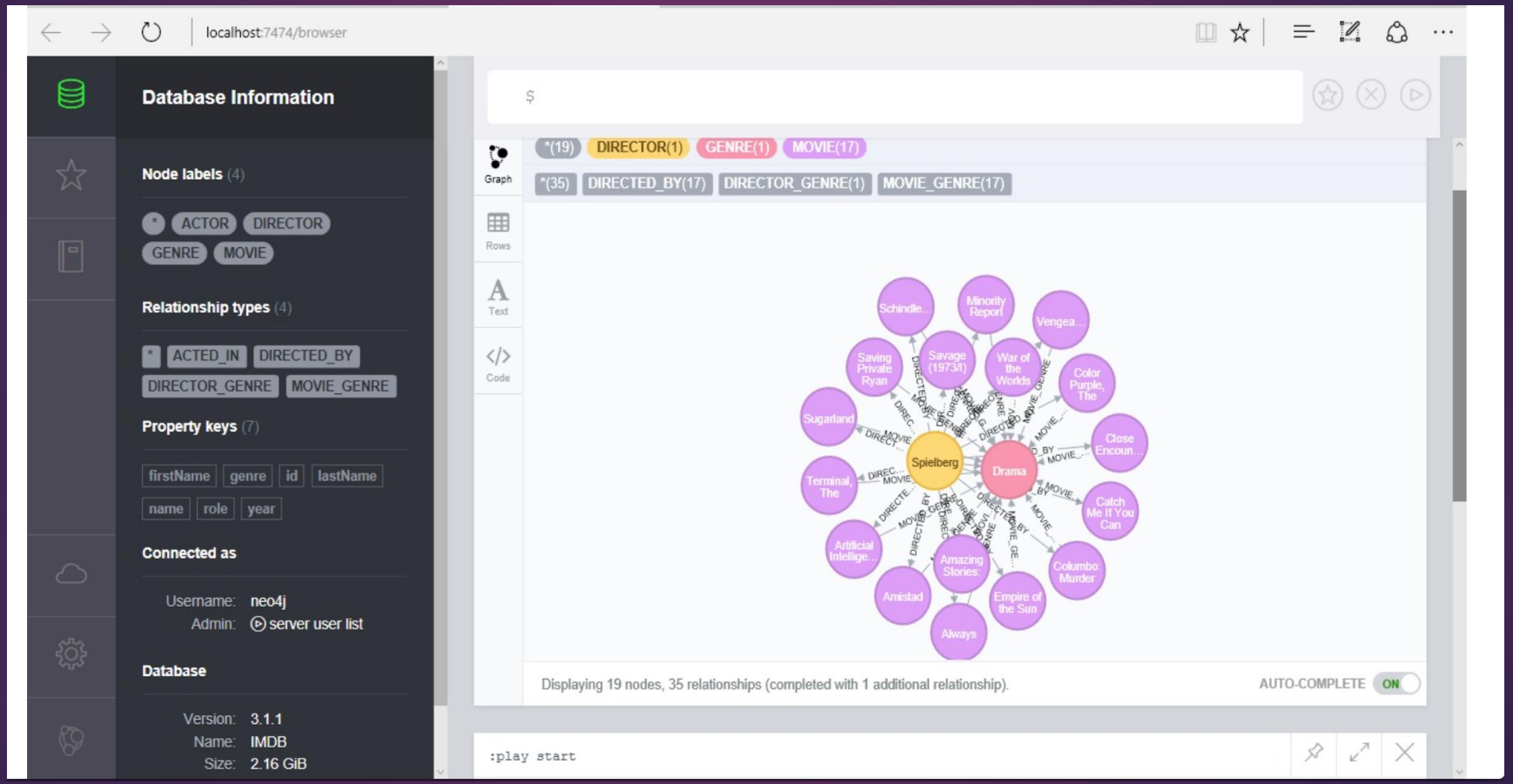


Figure 4 - Matrix of Neo4j Adoption, as of January 2015

DEMO


localhost:7474/browser
Database Information
Node labels (4)
* ACTOR DIRECTOR GENRE MOVIE
Relationship types (4)
* ACTED_IN DIRECTED_BY DIRECTOR_GENRE MOVIE_GENRE
Property keys (7)
firstName genre id lastName name role year
Connected as
Username: neo4j
Admin: server user list
Database
Version: 3.1.1
Name: IMDB
Size: 2.16 GiB
$
*(19) DIRECTOR(1) GENRE(1) MOVIE(17)
*(35) DIRECTED_BY(17) DIRECTOR_GENRE(1) MOVIE_GENRE(17)
Graph
Rows
Text
Code
Schindle...
Minority Report
Vengea...
Saving Private Ryan
Savage (1973/I)
War of the Worlds
Color Purple, The
Sugarland
Close Encoun...
Spielberg
Drama
Terminal, The
Catch Me If You Can
Artificial Intellige...
Amazing Stories:
Columbo: Murder
Amistad
Empire of the Sun
Always
Displaying 19 nodes, 35 relationships (completed with 1 additional relationship).
AUTO-COMPLETE ON
:play start

THANK YOU