

Lecture 17: NoSQL database

BADM/ACCY 352

Spring 2017

Instructor: Yi Yang, PhD

Last lecture

- Database warehouse

This lecture

- NoSQL database

What is NoSQL?

- NoSQL: Not Only SQL
 - A NoSQL database provides a mechanism for storage and retrieval of data which is modeled in means other than the tabular relations used in relational database.
 - It does not use SQL as querying language. It has its own query language.
 - No fixed schema
 - No joins

It's not a replacement for a RDBMS but compliments it

NoSQL databases

- NoSQL classification based on data model
 - Key-Value: Amazon DynamoDB, Redis, ...
 - Document: MongoDB, CouchDB, ...
 - Graph: Neo4J
 - Column: HBase, Google BigTable, Facebook Cassandra,

Relational Databases



ORACLE

PostgreSQL



SYBASE

NoSQL Databases

key-value

Amazon
DynamoDB (Beta)

ORACLE
BERKELEY DB **11g**



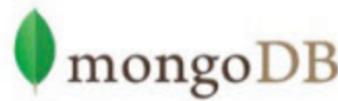
graph



column



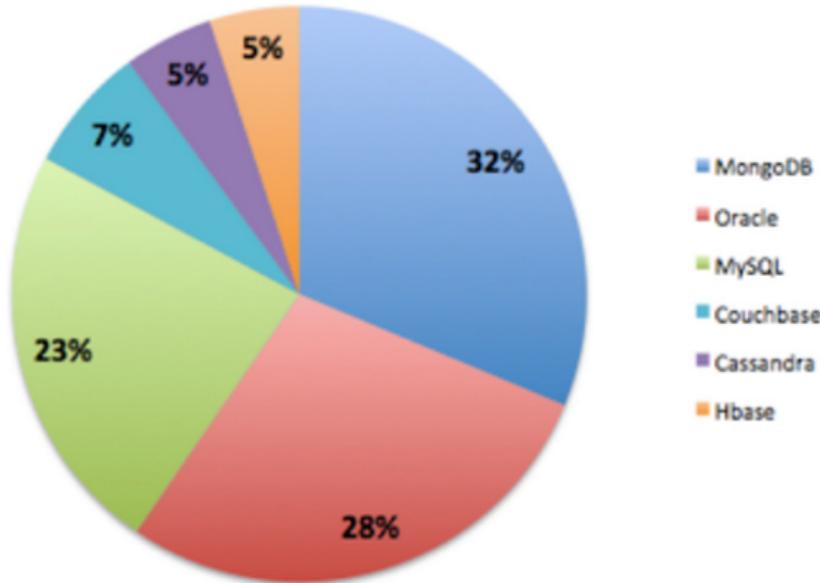
document



The rise of NoSQL

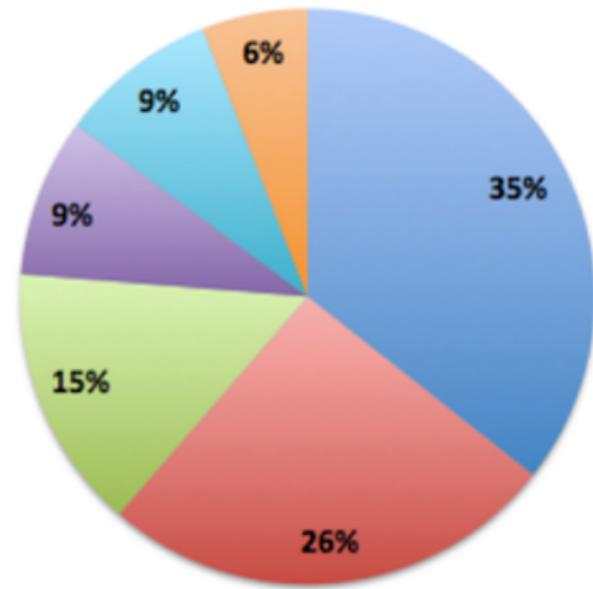
Database Popularity

March 2013-2014



Database Popularity

March 2014-2015



Database popularity.

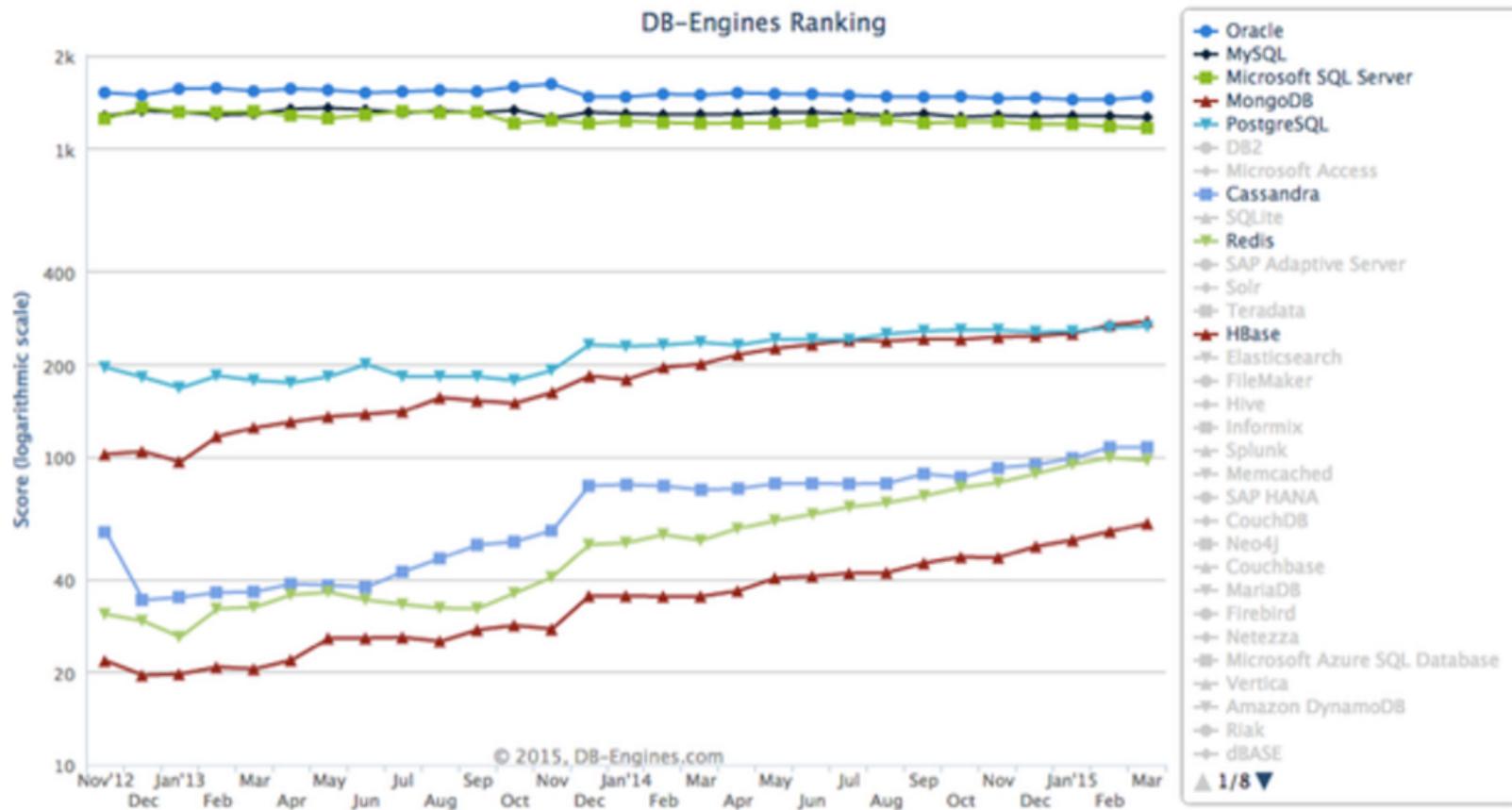
measured in media mentions

The rise of NoSQL

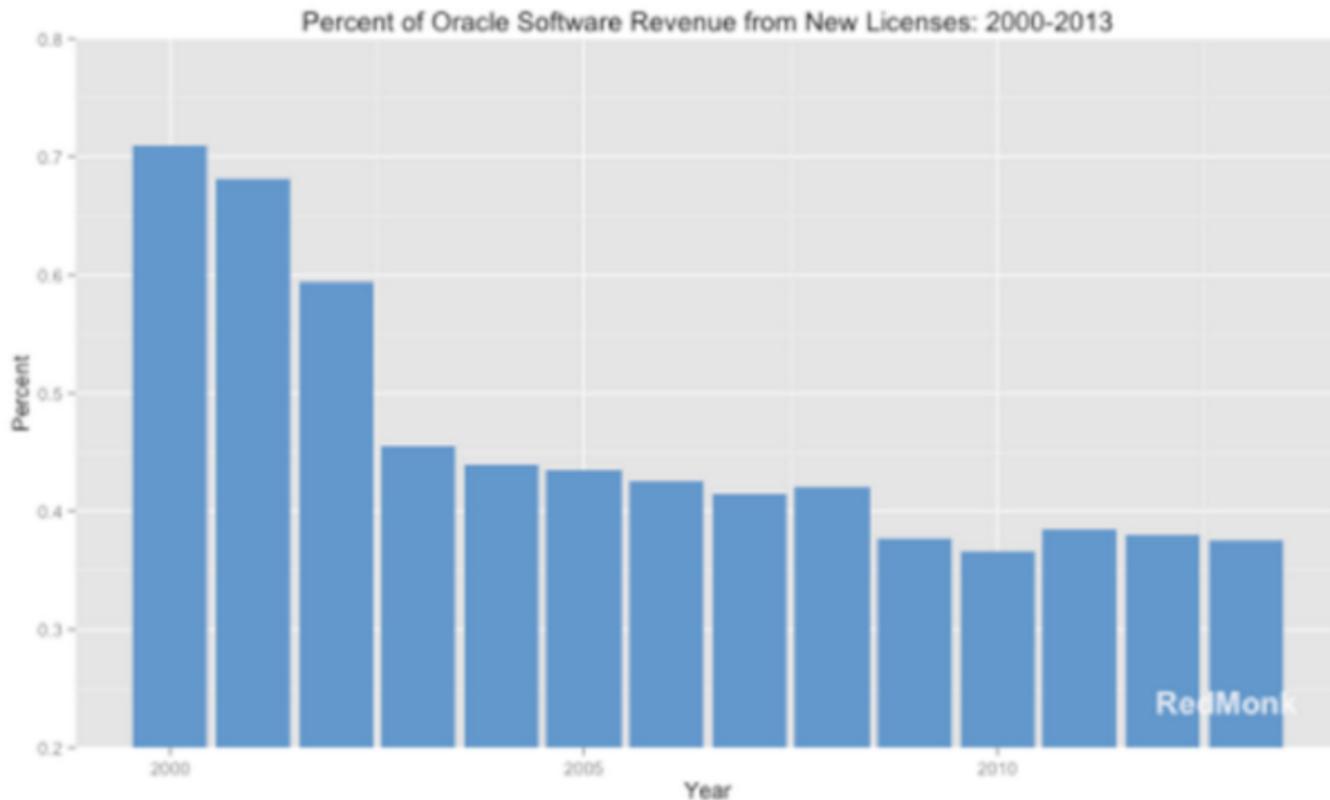
Rank			DBMS	Database Model	Score		
Nov 2015	Oct 2015	Nov 2014			Nov 2015	Oct 2015	Nov 2014
1.	1.	1.	Oracle	Relational DBMS	1480.95	+13.99	+28.82
2.	2.	2.	MySQL	Relational DBMS	1286.84	+7.88	+7.77
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1122.33	-0.90	-97.87
4.	4.	↑ 5.	MongoDB +	Document store	304.61	+11.34	+59.87
5.	5.	↓ 4.	PostgreSQL	Relational DBMS	285.69	+3.56	+28.33
6.	6.	6.	DB2	Relational DBMS	202.52	-4.28	-3.71
7.	7.	7.	Microsoft Access	Relational DBMS	140.96	-0.87	+2.12
8.	8.	↑ 9.	Cassandra +	Wide column store	132.92	+3.91	+40.93
9.	9.	↓ 8.	SQLite	Relational DBMS	103.45	+0.78	+8.17
10.	10.	↑ 11.	Redis +	Key-value store	102.41	+3.61	+20.06

DB popularity ranking, based on a more complete measure

The rise of NoSQL

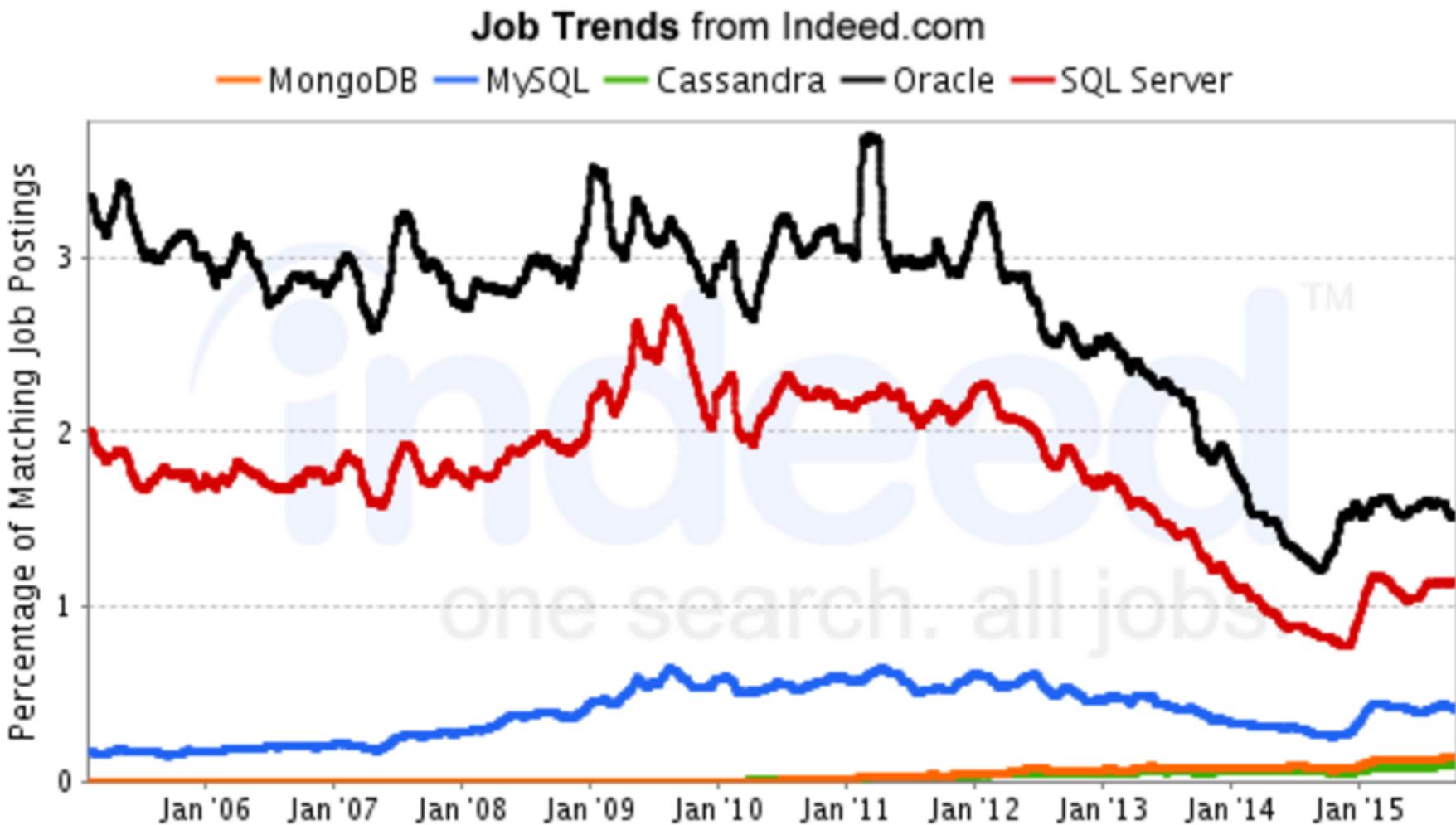


Oracle is selling less

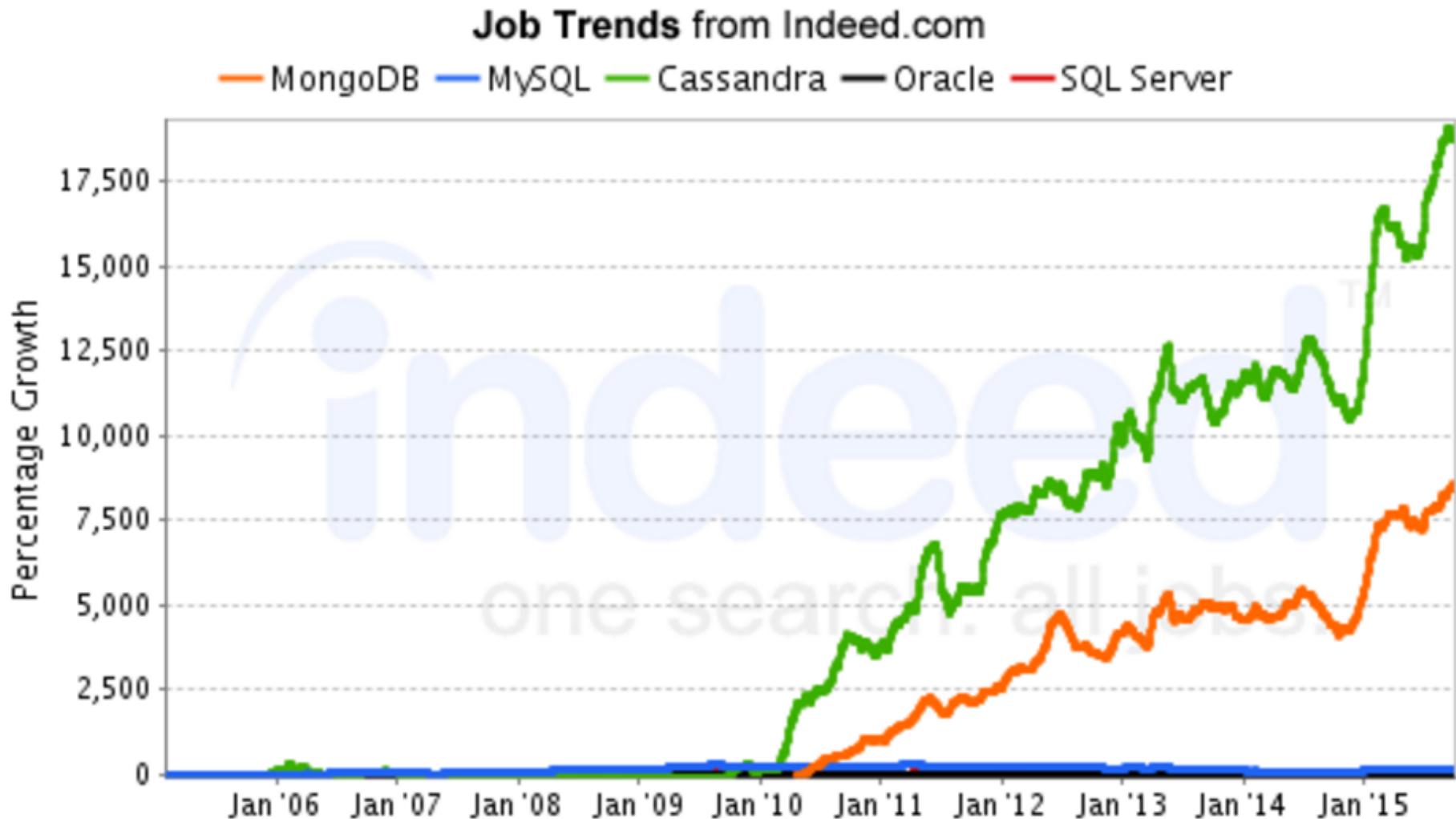


Percent of Oracle software revenue from new licenses.

Absolute job trends



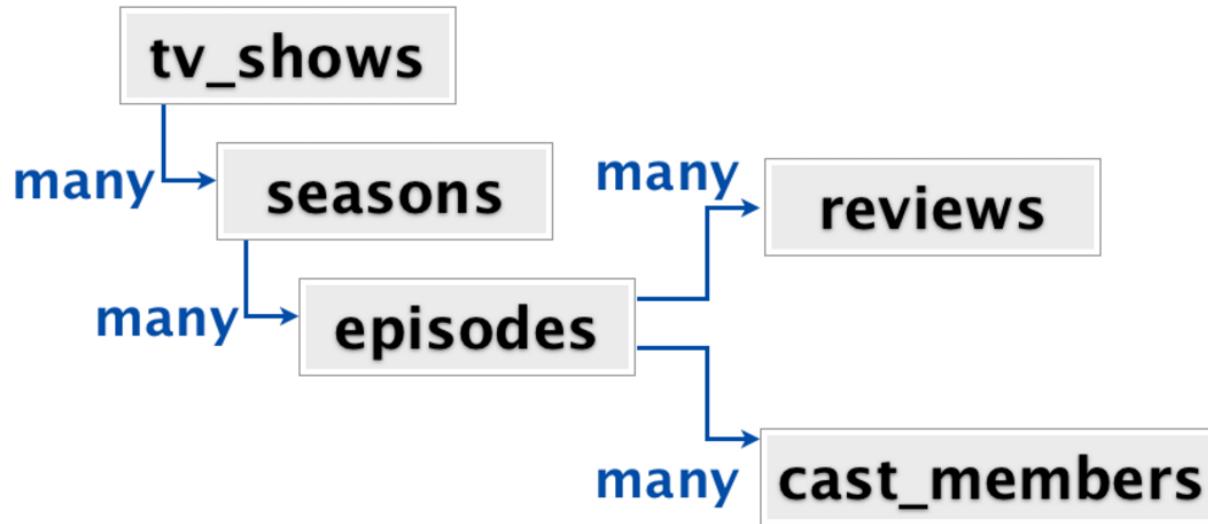
Relative job trends



Relational Data Model Revisit

- Relational Database contains a set of related tables.
 - Tables
 - Row, Column
 - PK, FK
- Entity-Relationship Modeling (ERD)
- Normalization
- SQL: structured query language
- RDBMS: MySQL, Oracle, etc..

Relational model



- In relational database, how many table-joins do you need to get **ALL** information about a TV show?
- The performance of such design is a little slow
 - “General Hospital” has aired over 12,000 episodes over the course of 50+ seasons

Alternative way to store data; Document Store data model

- Instead of storing data records in **pre-defined** table structure

employees

	Firstname	Lastname
	John	Doe

- Each document is in JSON format

```
{"firstName": "John", "lastName": "Doe"}
```

Each data record is stored
inside curly braces

A name/value pair consists of a field name (in double quotes), followed by a colon, followed by a value

employees

Firstname	Lastname
John	Doe
Anna	Smith
Peter	Jones

```
{"employees": [  
    {"firstName": "John", "lastName": "Doe"},  
    {"firstName": "Anna", "lastName": "Smith"},  
    {"firstName": "Peter", "lastName": "Jones"}  
]}
```

If a field has multiple values,
the set of value is stored
inside square brackets

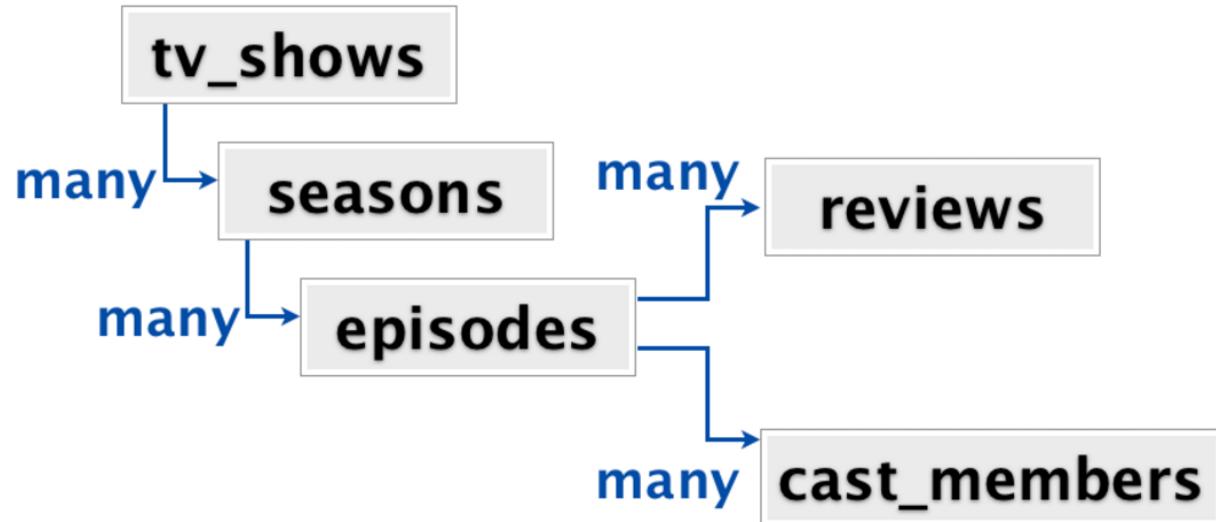
employees

ID	Firstname	Lastname
E1	John	Doe
E2	Anna	Smith
E3	Peter	Jones

customers

ID	name	EMP_ID
C1	Ford	E1
C2	GE	E1

```
{"employees": [
    {"id": E1, "firstName": "John", "lastName": "Doe", "customers": [
        {"id": C1, "name": "Ford"}, 
        {"id": C2, "name": "GE"} 
    ]}, 
    {"id": E2, "firstName": "Anna", "lastName": "Smith"}, | 
    {"id": E3, "firstName": "Peter", "lastName": "Jones"} 
]}|
```



```

{
  title: 'Babylon 5',
  seasons: [
    {season_number: '1',
      episodes: [
        {ordinal_within_season: '1',
          title: 'Midnight on the Firing Line',
          reviews: [...],
          cast_members: [...]
        }
      ]
    }
  ]
}
  
```

All of the data we need for a TV show is under one document, so it's very fast to retrieve all this information at once, even if the document is very large.

Document Store Model

Document store: MongoDB

- MongoDB uses **document store** data model.



Relational data model

Highly-structured table organization with rigidly-defined data formats and record structure.



Document data model

Collection of complex documents with arbitrary, nested data formats and varying “record” format.

MySQL MongoDB

Table Collection

Row Document

Column Field

Advantages

- Schema free
(more flexible)
- No complex joins
(more fast?)
- MongoDB is
easier to scale
- Open source

Disadvantages

- Schema free
- ACID is not supported, which means there is no guarantee of data consistency.
(eventual consistency)
- No common standard.

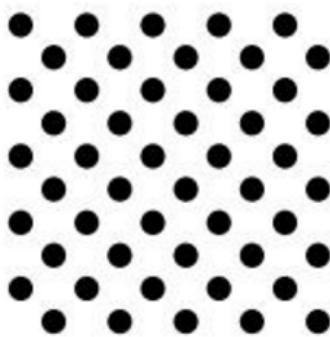
iClicker question

What is a better database option for a banking database system that records customer transactions?

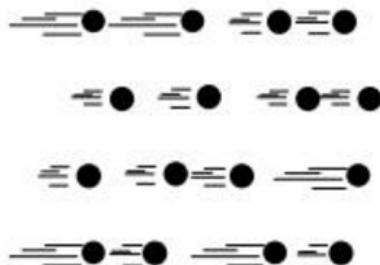
- a. Relational database
- b. NoSQL database

What makes it big data: 4V

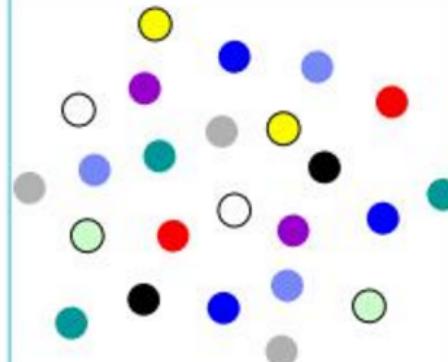
Volume



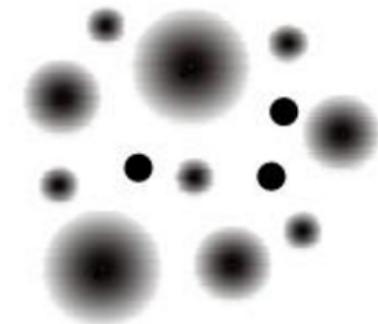
Velocity



Variety



Veracity*



Data at Rest

Terabytes to exabytes of existing data to process

Data in Motion

Streaming data, milliseconds to seconds to respond

Data in Many Forms

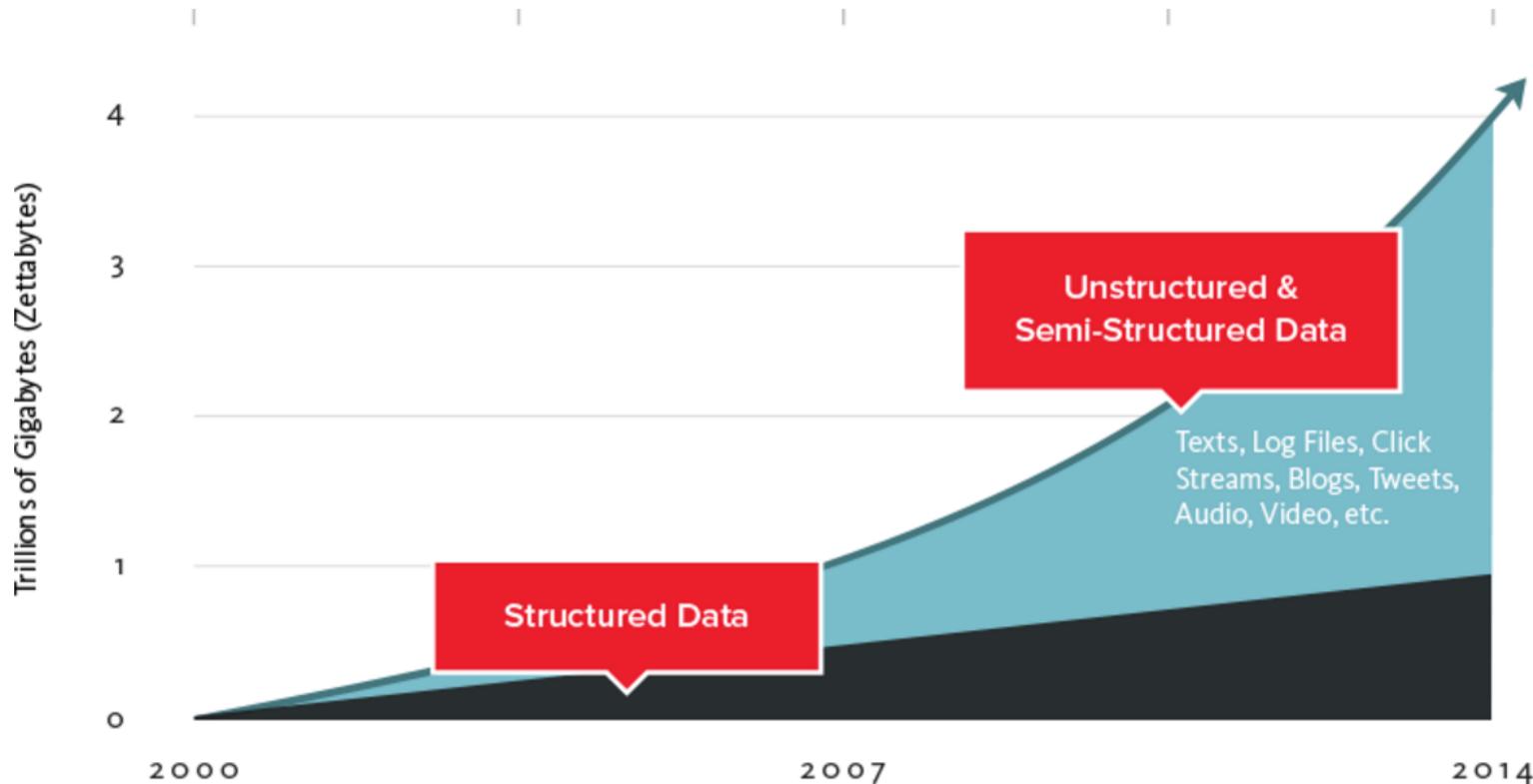
Structured, unstructured, text, multimedia

Data in Doubt

Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

Unstructured Data

Big Data



NoSQL databases are increasingly used in big data and real-time web applications

Unstructured Data

- **Unstructured Data** refers to information that either does not have a pre-defined data model or is not organized in a pre-defined manner.
- In other words, unstructured data does not fit well in relational database.
- Example
 - **Social media data:** data that is generated from the social media platforms such as YouTube, Facebook, Twitter, LinkedIn
 - **Mobile data:** This includes data such as text messages and location information.
 - **Photographs and video:** This includes security, surveillance, and traffic video.

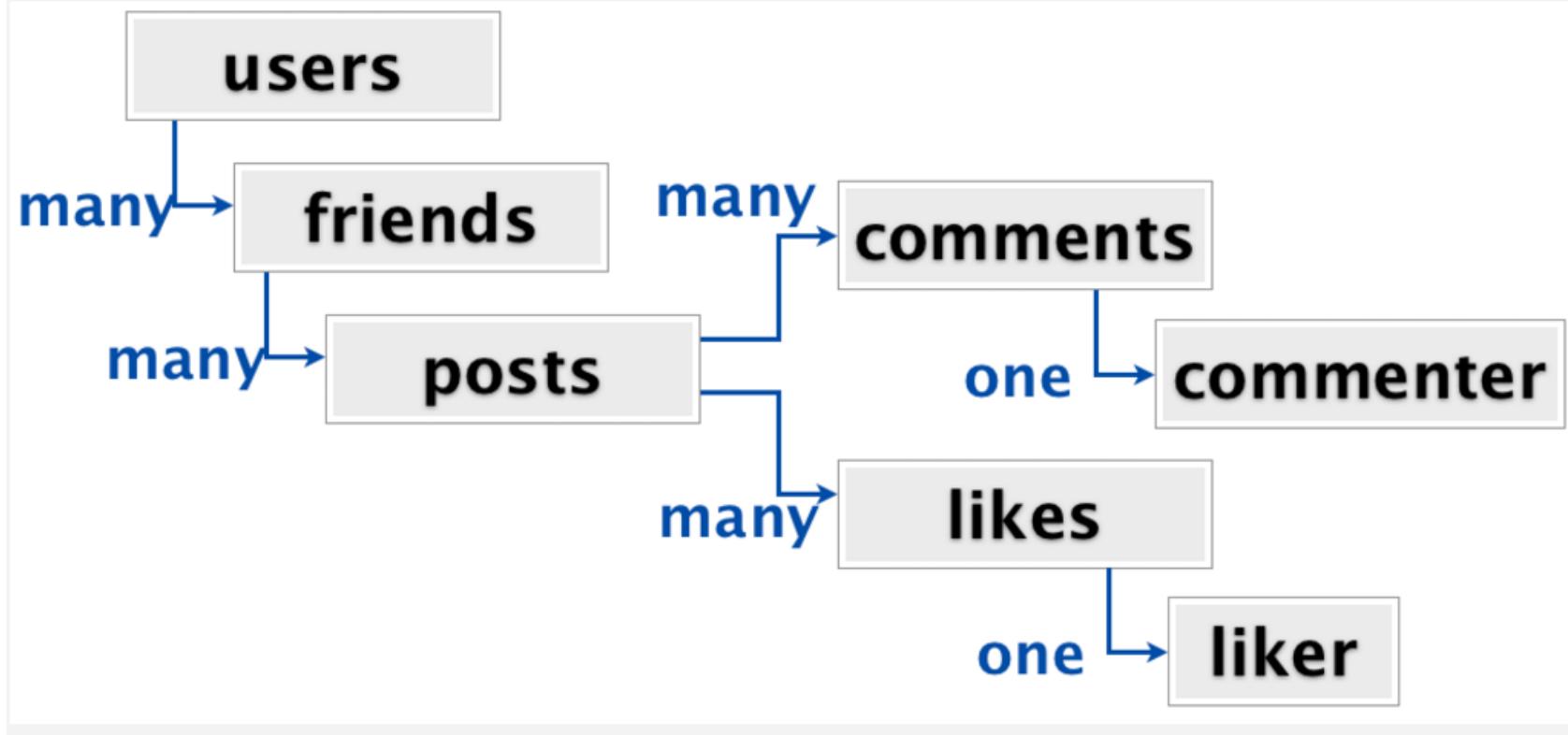


Chicago Uses MongoDB to Create a Smarter and Safer City

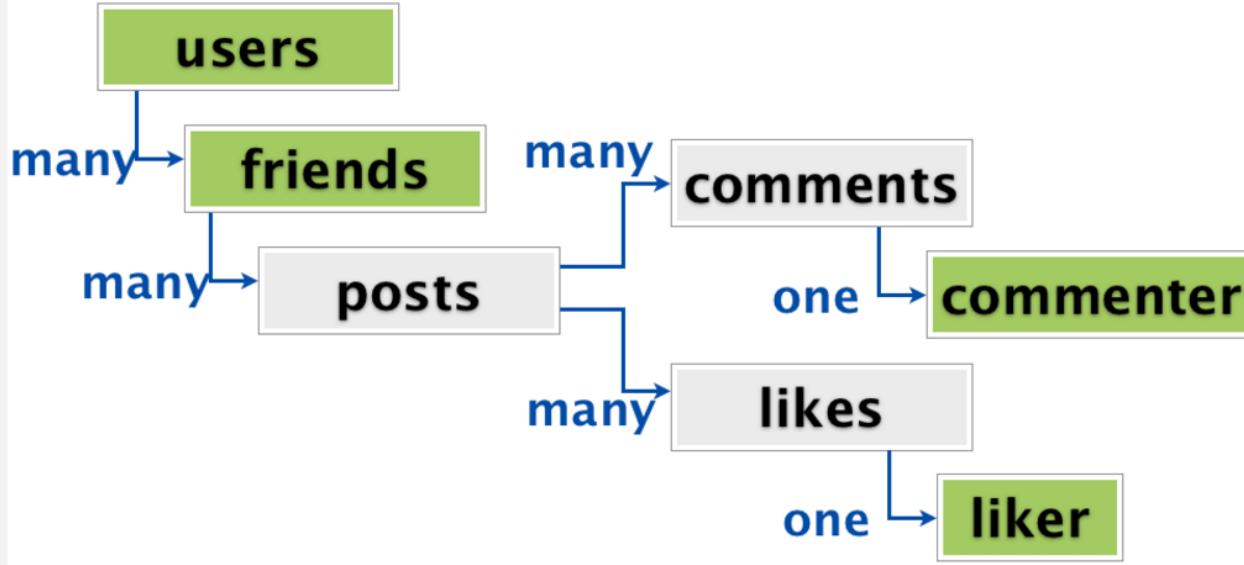
[Link](#)

- What's the data management problem that the City of Chicago is facing?
- Why do they choose NoSQL over relational database?

Is NoSQL always better? A social network example



Consider to build a database for a social network, do you think MongoDB, the document store database, is a right choice?



```
{  
  name: Joe,  
  url: '...',  
  stream:  
  [{  
    user: {name: Jane, url: '...'},  
    title: 'today',  
    body: 'go fly a kite',  
    likes: [  
      {user: {name: Lu, url: '...'}},  
      {user: {name: Joe, url: '...'}}  
    ],  
  }]  
}
```

Nested structure,
infinite loop

A true story

- Diaspora: four undergraduates from New York University made a Kickstarter video asking for \$10,000 (eventually, they raised \$200,000) to spend the summer building a distributed social network alternative to Facebook.
- Their team chose MongoDB.
- <https://diasporafoundation.org/>

Relational or NoSQL

- What's the nature of the data?
 - Structured data or unstructured data
- What's the volatility of the data?
 - Is the data model likely to change and evolve or is it most likely going to stay the same?
- Co-existence of RDBMS and NoSQL databases
- Recommended read (By Pinterest Engineering Team)

Learn to stop using shiny new things and love MySQL

Summary

- NoSQL database
- Learning objective
 - What are NoSQL databases
 - NoSQL vs. Relational database