



mongoDB

What is MongoDB?

Why do we have MongoDB?

Why is MongoDB open source?

What it's like to work at an open source database company?

Tech Dive

(First, a little about me)



Raymond Jacobson
Class of 2016 CS/CSE

Interned at MongoDB, summer '15

So... What is MongoDB?

What is a database?

People have always used
databases!



AUDITE

Dominum eum
propterea et laudab
et exultab deo

Exulta in dñe

Propterea et laudab
et exultab deo





Store Information



1970s

The Relational Database

Data is inherently relational

Professors

Name	Favorite Software
Goldschmidt	Notepad
Moorthy	Observatory
Magdon-Ismail	LaTeX
Cutler	Dr. Memory

Software

Name	Platforms	Release Date
Notepad	Windows	1983
Dr. Memory	Cross-platform	2011
Observatory	Web	2010
LaTeX	Cross-platform	1985



```
SELECT FAVORITE_SOFTWARE FROM PROFESSORS WHERE NAME='MOORTHY'
```

```
>> "Observatory"
```

Joins

Professors

Name	Favorite Software
Goldschmidt	Notepad
Moorthy	Observatory
Magdon-Ismail	LaTeX
Cutler	Dr. Memory

Software

Name	Platforms	Release Date
Notepad	Windows	1983
Dr. Memory	Cross-platform	2011
Observatory	Web	2010
LaTeX	Cross-platform	1985

```
SELECT PROFESSORS.NAME  
FROM PROFESSORS JOIN SOFTWARE ON  
PROFESSORS.FAVORITE_SOFTWARE=SOFTWARE.NAME  
WHERE SOFTWARE.PLATFORMS="Cross-platform"
```

>> "Magdon-Ismail", "Cutler"

Professors

Software

Name	Favorite Software		Name	Platforms	Release Date
Goldschmidt	Notepad	→	Notepad	Windows	1983
Moorthy	Observatory	↙ ↘	Dr. Memory	Cross-platform	2011
Magdon-Ismail	LaTeX		Observatory	Web	2010
Cutler	Dr. Memory		LaTeX	Cross-platform	1985

```
SELECT PROFESSORS.NAME
FROM PROFESSORS JOIN SOFTWARE ON
PROFESSORS.FAVORITE_SOFTWARE=SOFTWARE.NAME
WHERE SOFTWARE.PLATFORMS="Cross-platform"
```

>> "Magdon-Ismail", "Cutler"

Name	Favorite Software	Platforms	Release Date
Goldschmidt	Notepad	Windows	1983
Moorthy	Observatory	Web	2010
Magdon-Ismail	LaTeX	Cross-platform	1985
Cutler	Dr. Memory	Cross-platform	2011

```
SELECT PROFESSORS.NAME
FROM PROFESSORS JOIN SOFTWARE ON
PROFESSORS.FAVORITE_SOFTWARE=SOFTWARE.NAME
WHERE SOFTWARE.PLATFORMS="Cross-platform"
```

>> "Magdon-Ismail", "Cutler"

1970/80/90s

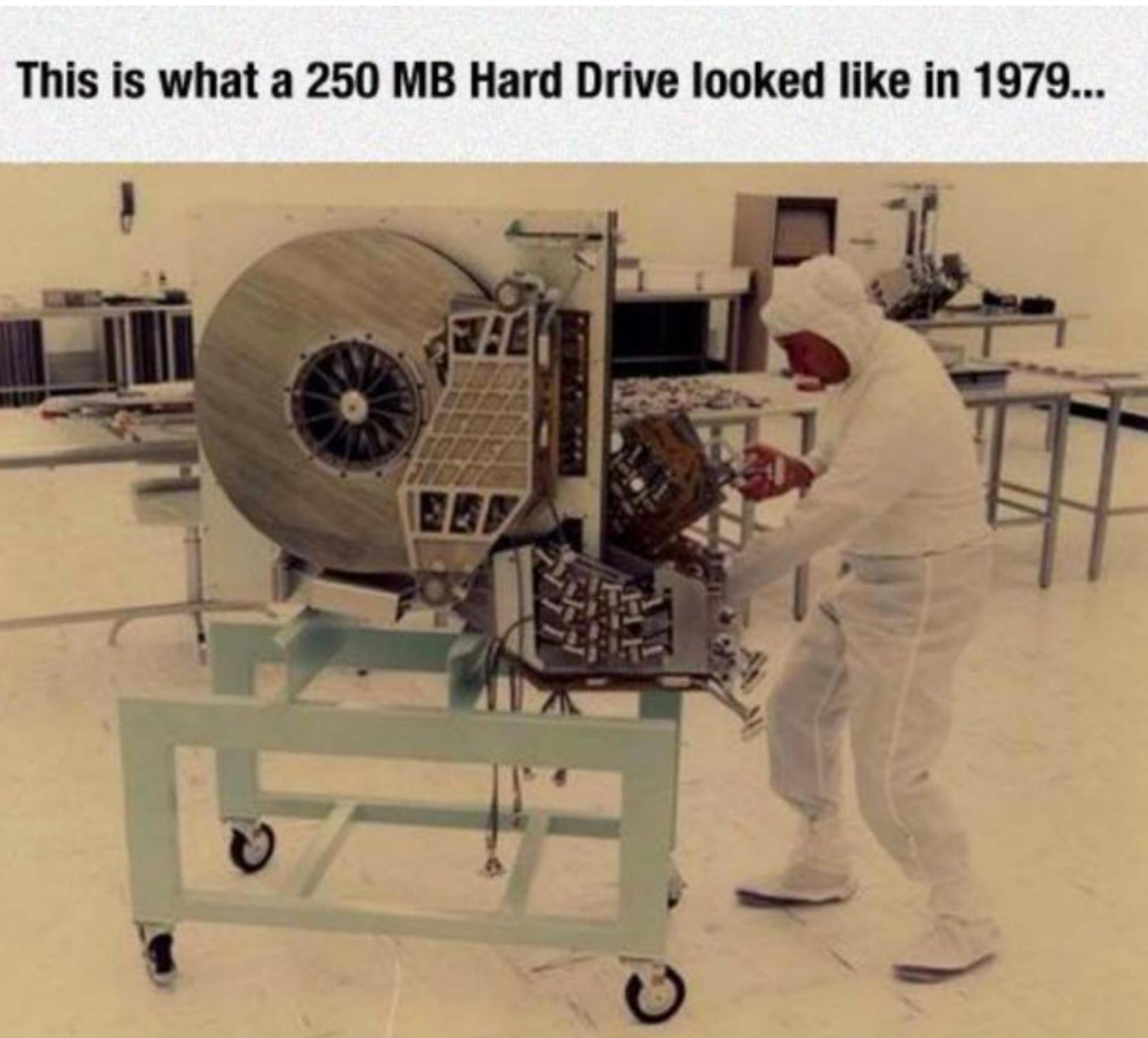
Keep things separate (normalize)

No redundant data

Seems pretty sensible

Name	Developers/Owner	Initial Version
OracleDB	Oracle	1978
DB2	IBM	1983
MySQL	Oracle	1995
Microsoft SQL Server	Microsoft	1989
PostgreSQL	PostgreSQL Global Development Group	1996
Sybase	SAP	1987

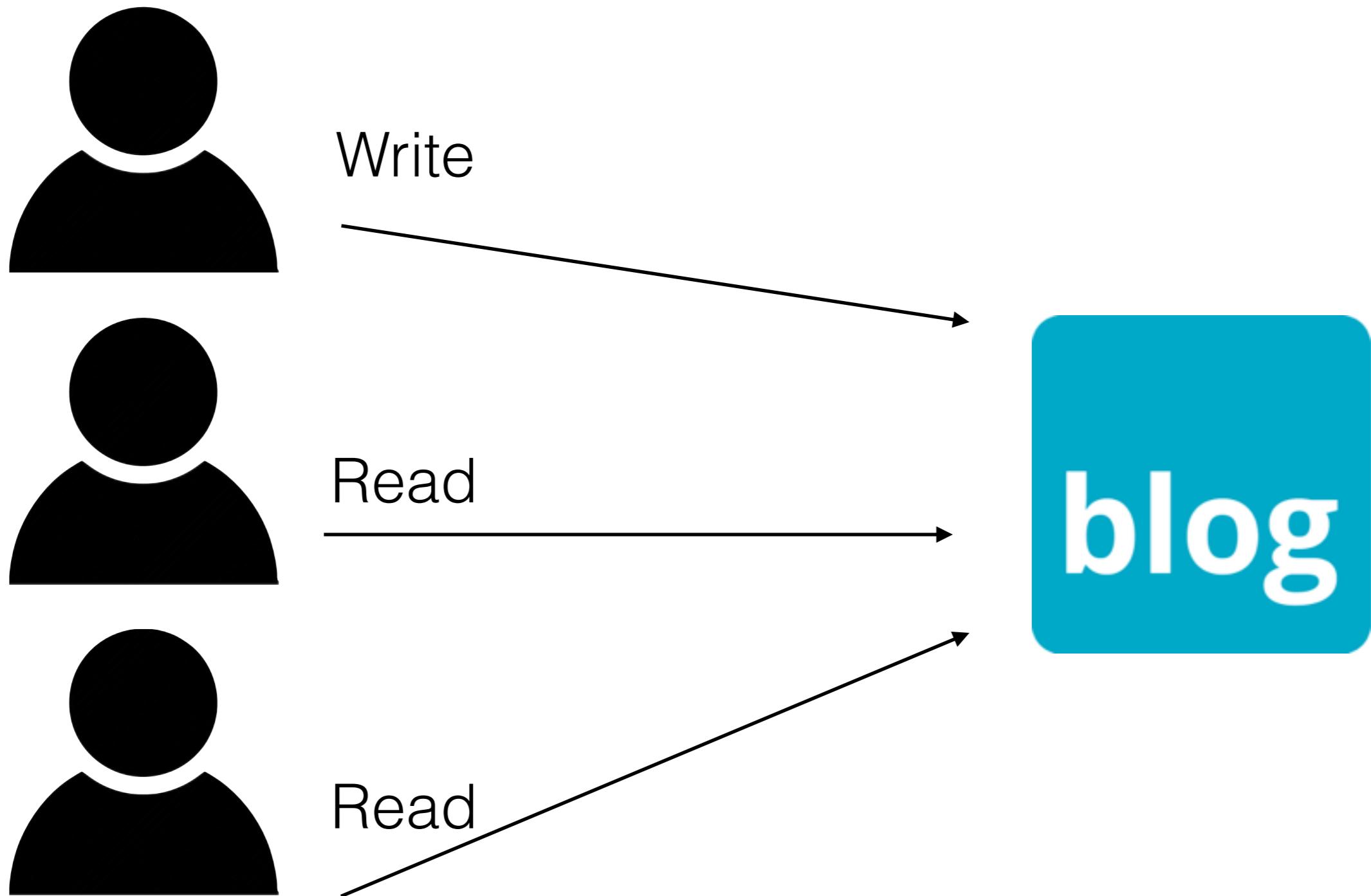
The world then



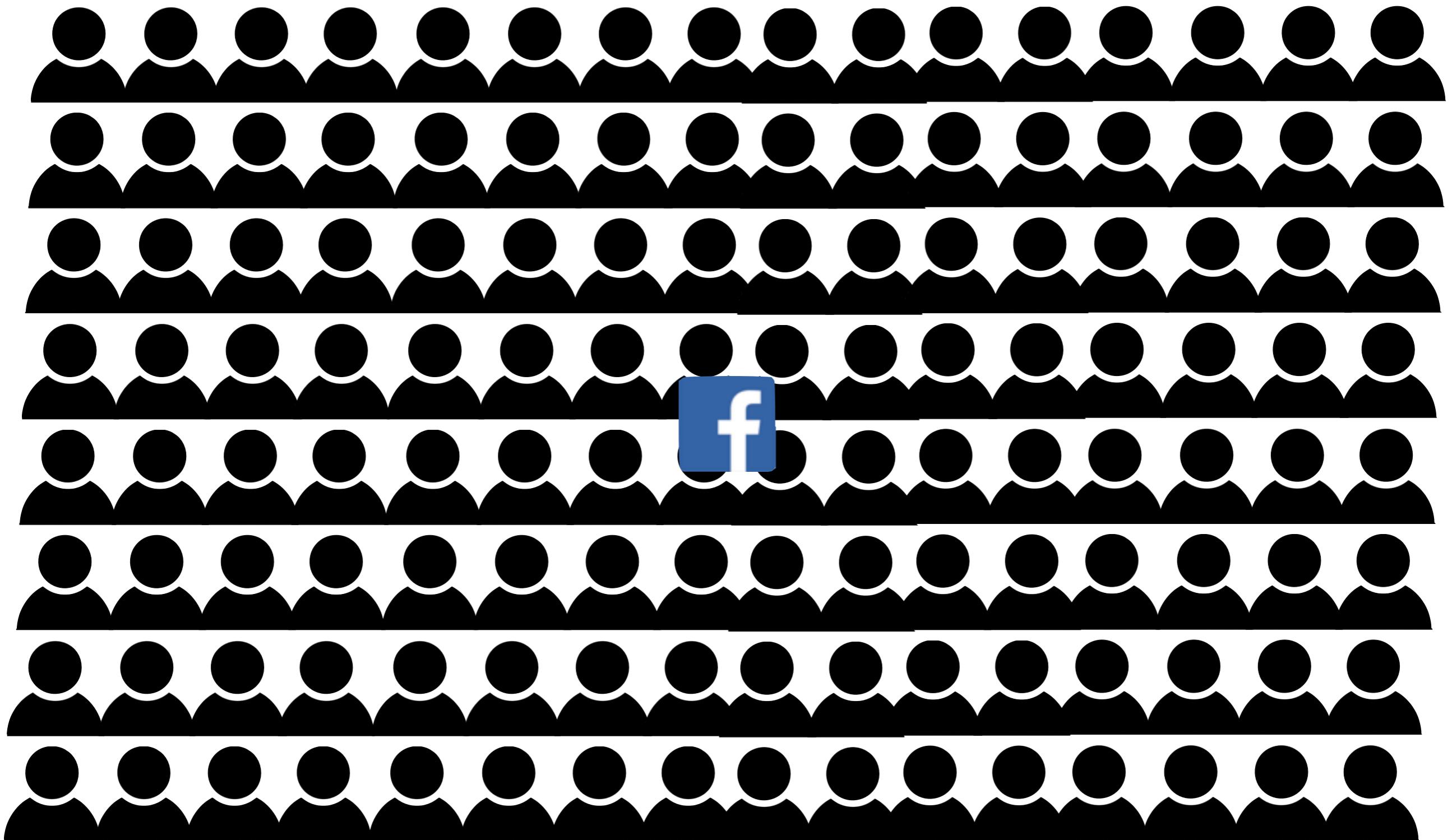
The world now



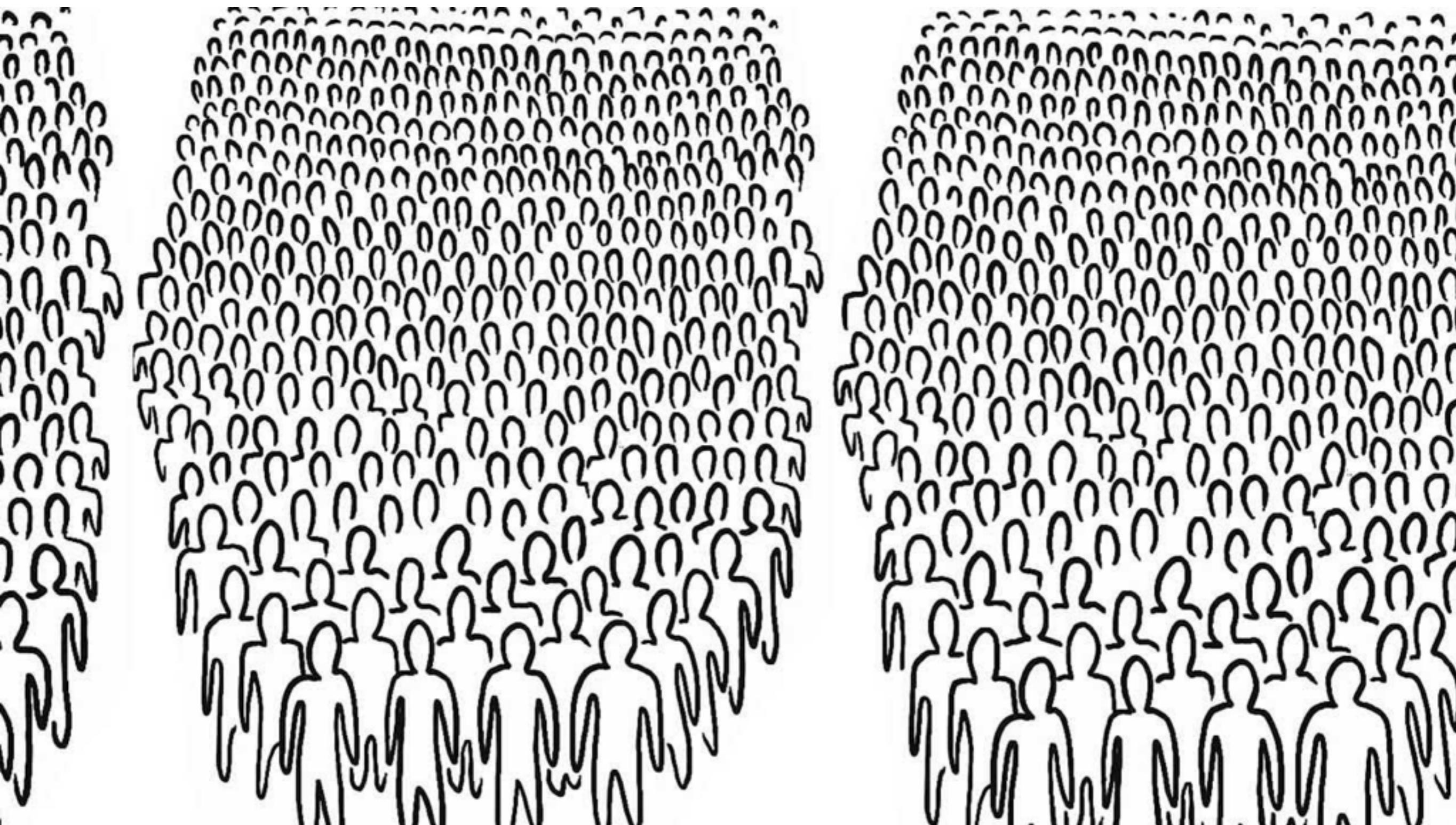
The world then



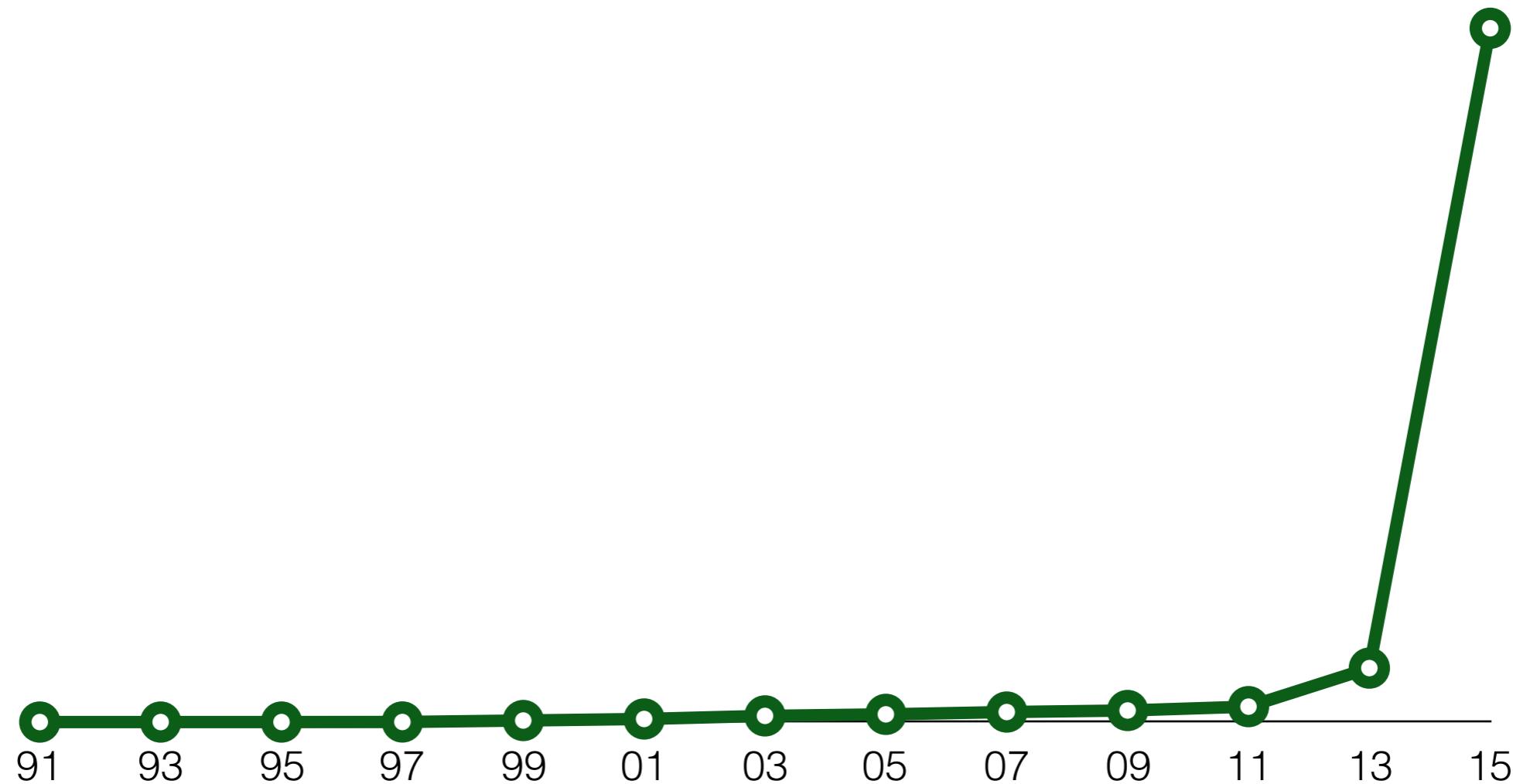
The world now



but really...



A full **90 percent** of all the data in the world
has been generated over the last two years



Vertical Scalability

Name	Favorite Software	Name	Platforms	Release Date
Goldschmidt	Notepad	Notepad	Windows	1983
Moorthy	Observatory	Dr. Memory	Cross-platform	2011
Magdon-Ismail	LaTeX	Observatory	Web	2010
Cutler	Dr. Memory	LaTeX	Cross-platform	1985



Horizontal Scalability

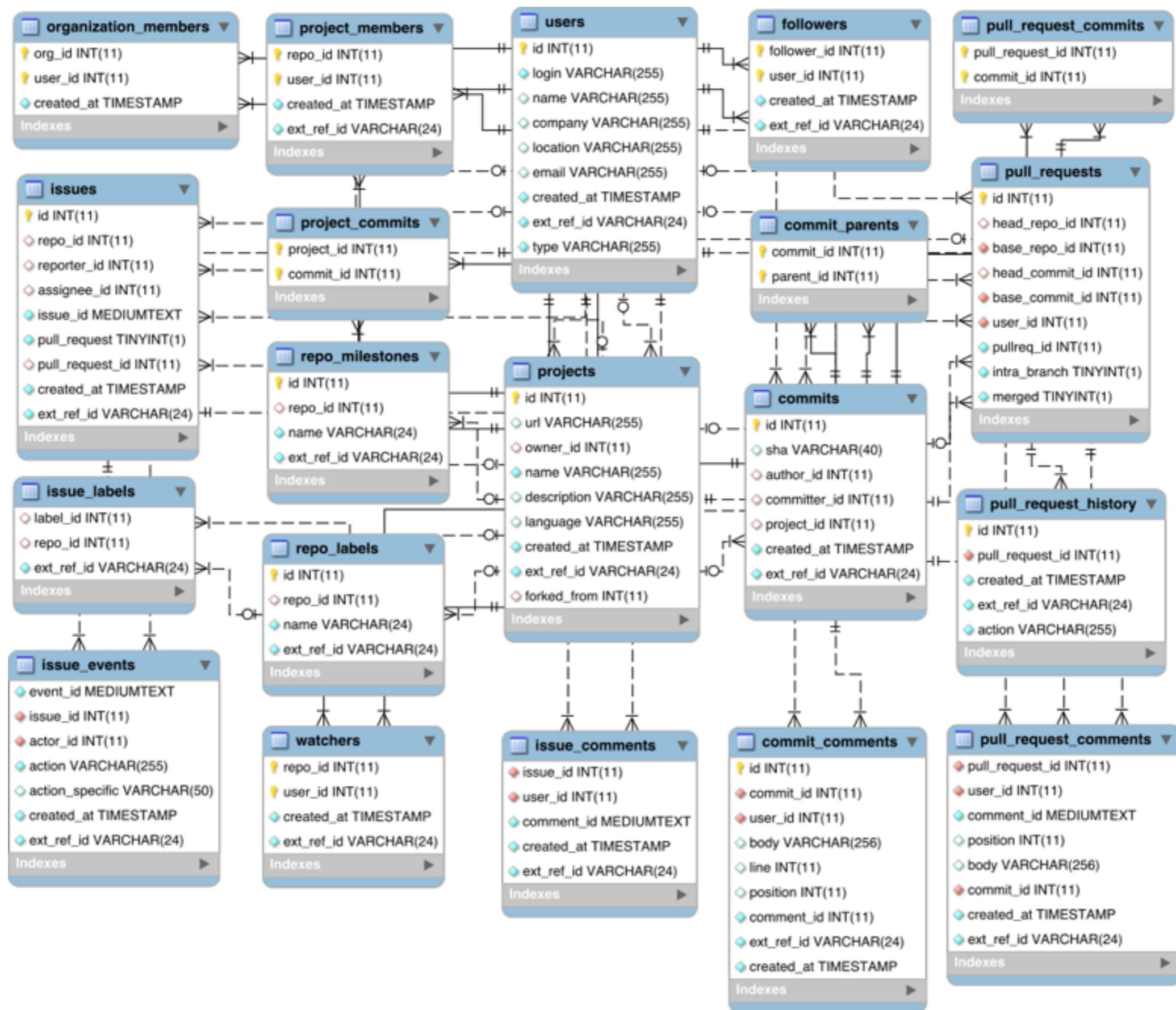
Find me what professors like software that runs on Windows?

Name	Favorite Software
Goldschmidt	Notepad
Moorthy	Observatory
Magdon-Ismail	LaTeX
Cutler	Dr. Memory

Name	Platforms	Release Date
Notepad	Windows	1983
Dr. Memory	Cross-platform	2011
Observatory	Web	2010
LaTeX	Cross-platform	1985

Expensive Join





Impedance Mismatch

```
1 class AudioDevice:  
2     def __init__(self, manufacturer, type):  
3         self.manufacturer = manufacturer  
4         self.type = type  
5  
6 class MP3Player(AudioDevice):  
7     def __init__(self, manufacturer, type, file_format):  
8         AudioDevice.__init__(self, manufacturer, type)  
9         self.file_format = file_format  
10  
11 class RecordPlayer(AudioDevice):  
12     def __init__(self, manufacturer, type, speed):  
13         AudioDevice.__init__(self, manufacturer, type)  
14         self.speed = speed
```

Name	Manufacturer	File Format	Speed
iPod	Apple	.m4a	>:(
Turntable	Pro-Ject	>:(33 rpm
Zune	Microsoft	.wma	>:(

Name	Manufacturer	File Format
iPod	Apple	.m4a
Zune	Microsoft	.wma

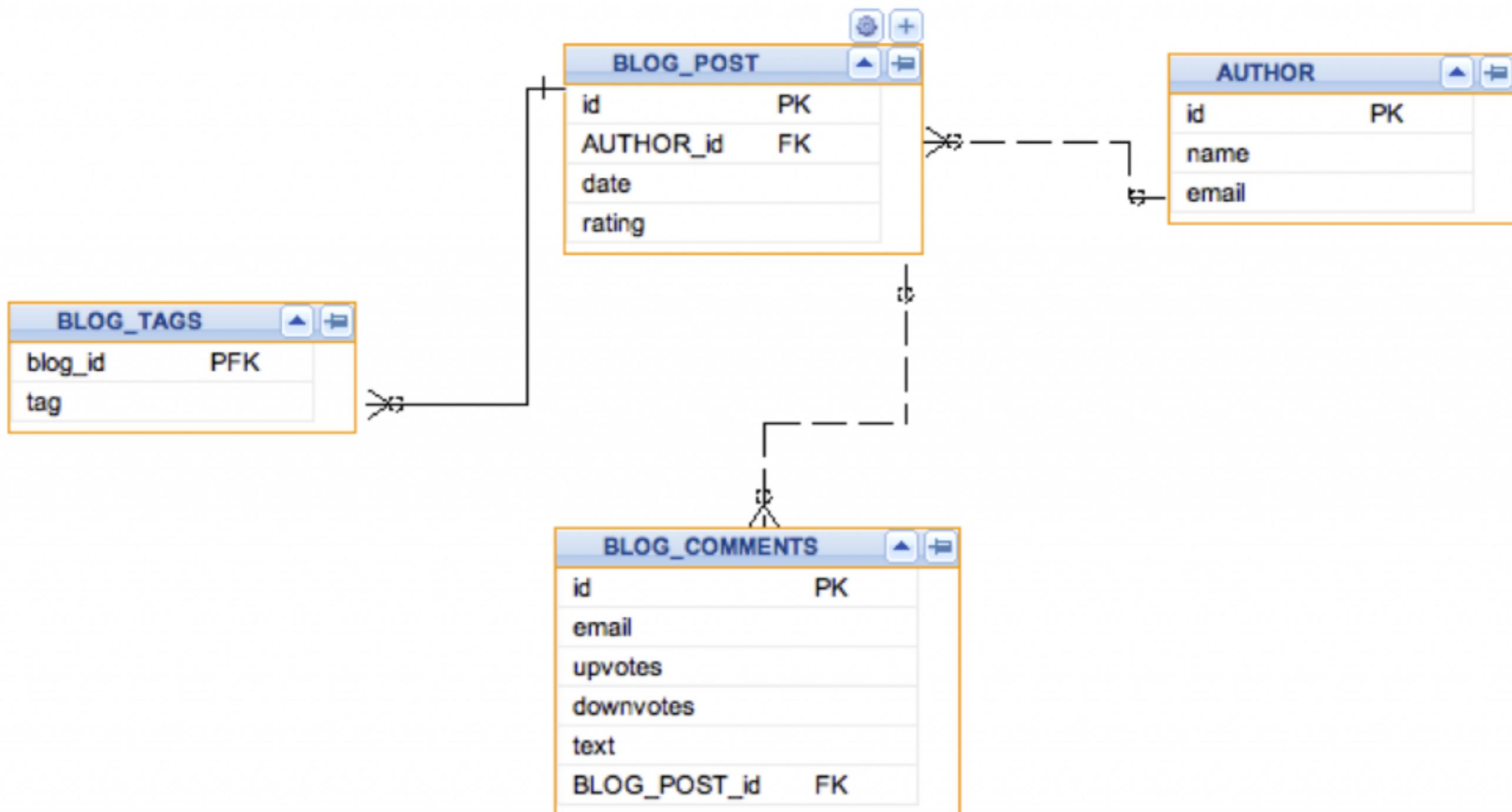
Name	Manufacturer	Speed
Turntable	Pro-Ject	33 rpm

Name	File Format
iPod	.m4a
Zune	.wma

Name	Speed
Turntable	33 rpm

Name	Manufacturer
Turntable	Pro-Ject
iPod	Apple
Zune	Microsoft

So... What is MongoDB?



A document database

```
{  
  _id: 1234,  
  author: { name: "Bob Davis", email : "bob@bob.com" },  
  post: "In these troubled times I like to ...",  
  date: { $date: "2010-07-12 13:23UTC" },  
  location: [ -121.2322, 42.1223222 ],  
  rating: 2.2,  
  comments: [  
    { user: "jgs32@hotmail.com",  
      upVotes: 22,  
      downVotes: 14,  
      text: "Great point! I agree" },  
    { user: "holly.davidson@gmail.com",  
      upVotes: 421,  
      downVotes: 22,  
      text: "You are a moron" }  
  ],  
  tags: [ "Politics", "Virginia" ]  
}
```

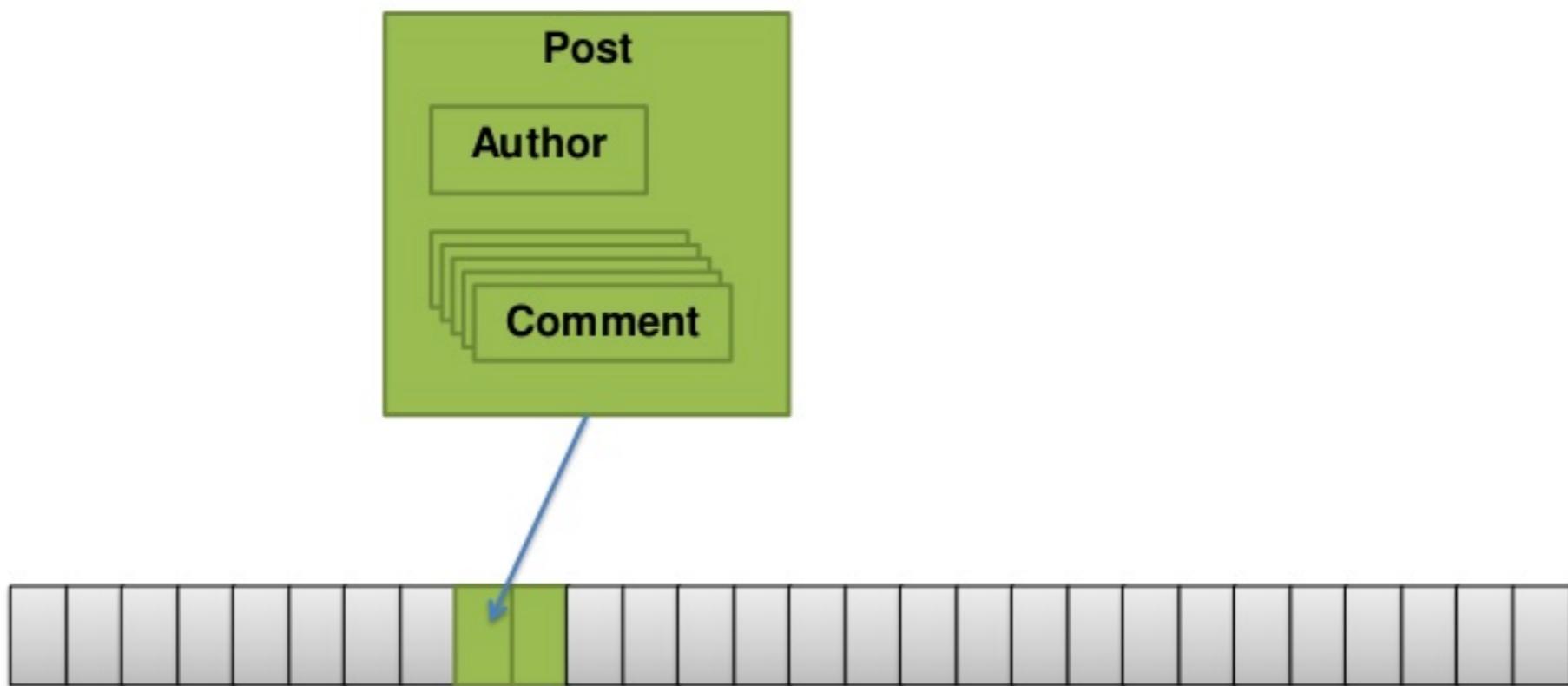
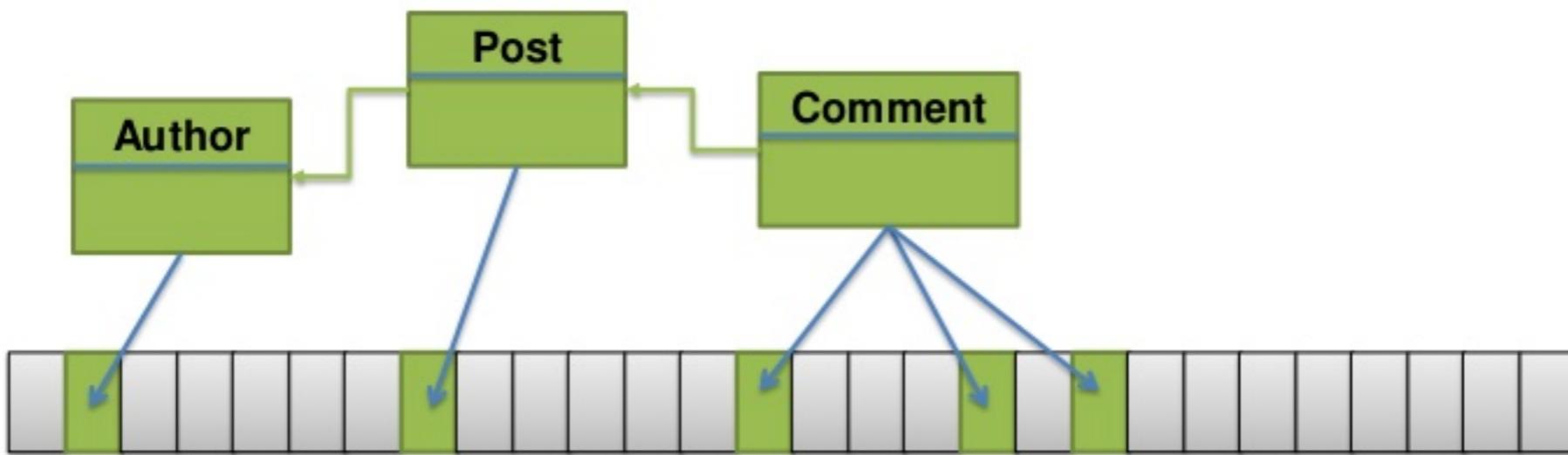
Actually, BSON (Binary JSON)
<http://bsonspec.org/>

A document database

```
{  
  _id: 1234,  
  author: { name: "Bob Davis", email : "bob@bob.com" },  
  post: "In these troubled times I like to ...",  
  date: { $date: "2010-07-12 13:23UTC" },  
  location: [ -121.2322, 42.1223222 ],  
  rating: 2.2,  
  comments: [  
    { user: "jgs32@hotmail.com",  
      upVotes: 22,  
      downVotes: 14,  
      text: "Great point! I agree" },  
    { user: "holly.davidson@gmail.com",  
      upVotes: 421,  
      downVotes: 22,  
      text: "You are a moron" }  
  ],  
  tags: [ "Politics", "Virginia" ]  
}
```

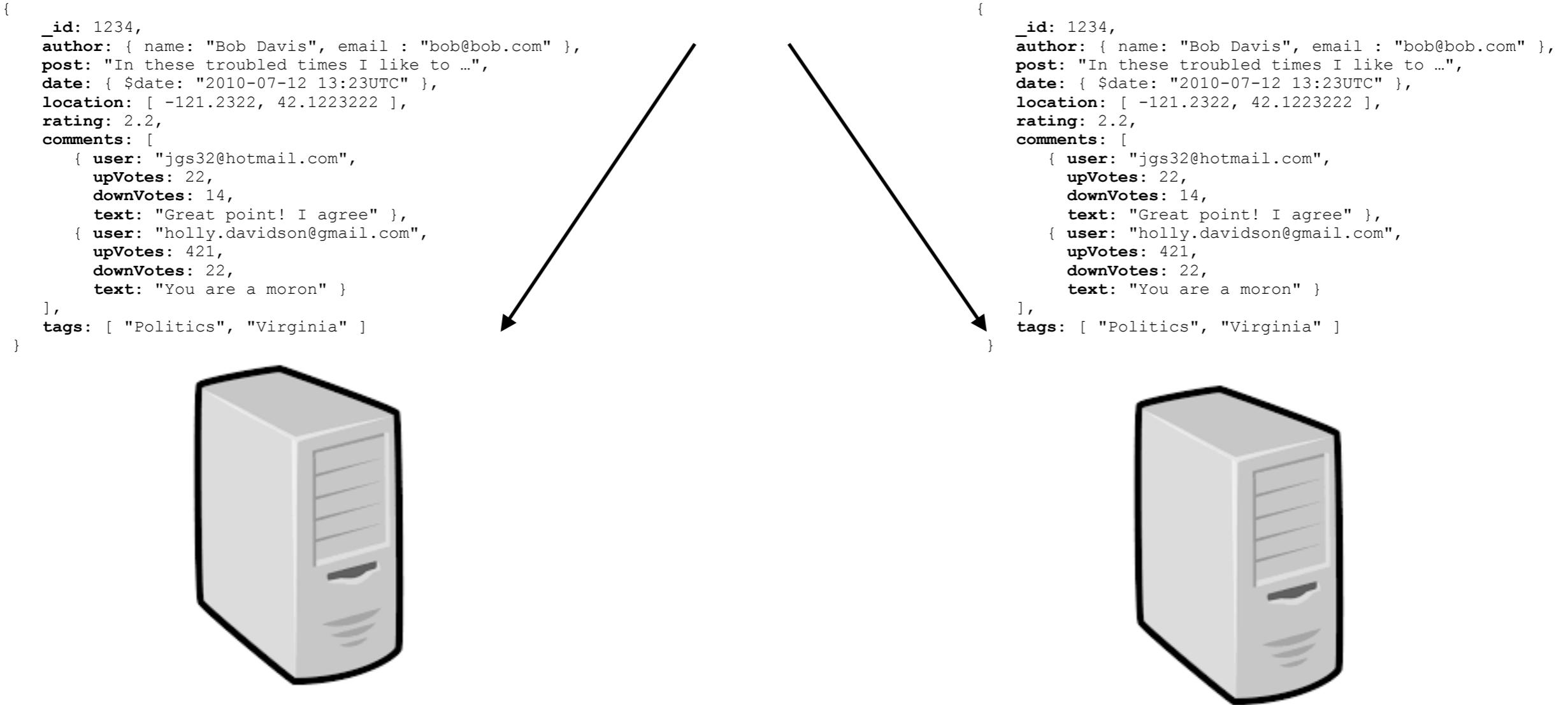
Normalization

What is it and why?

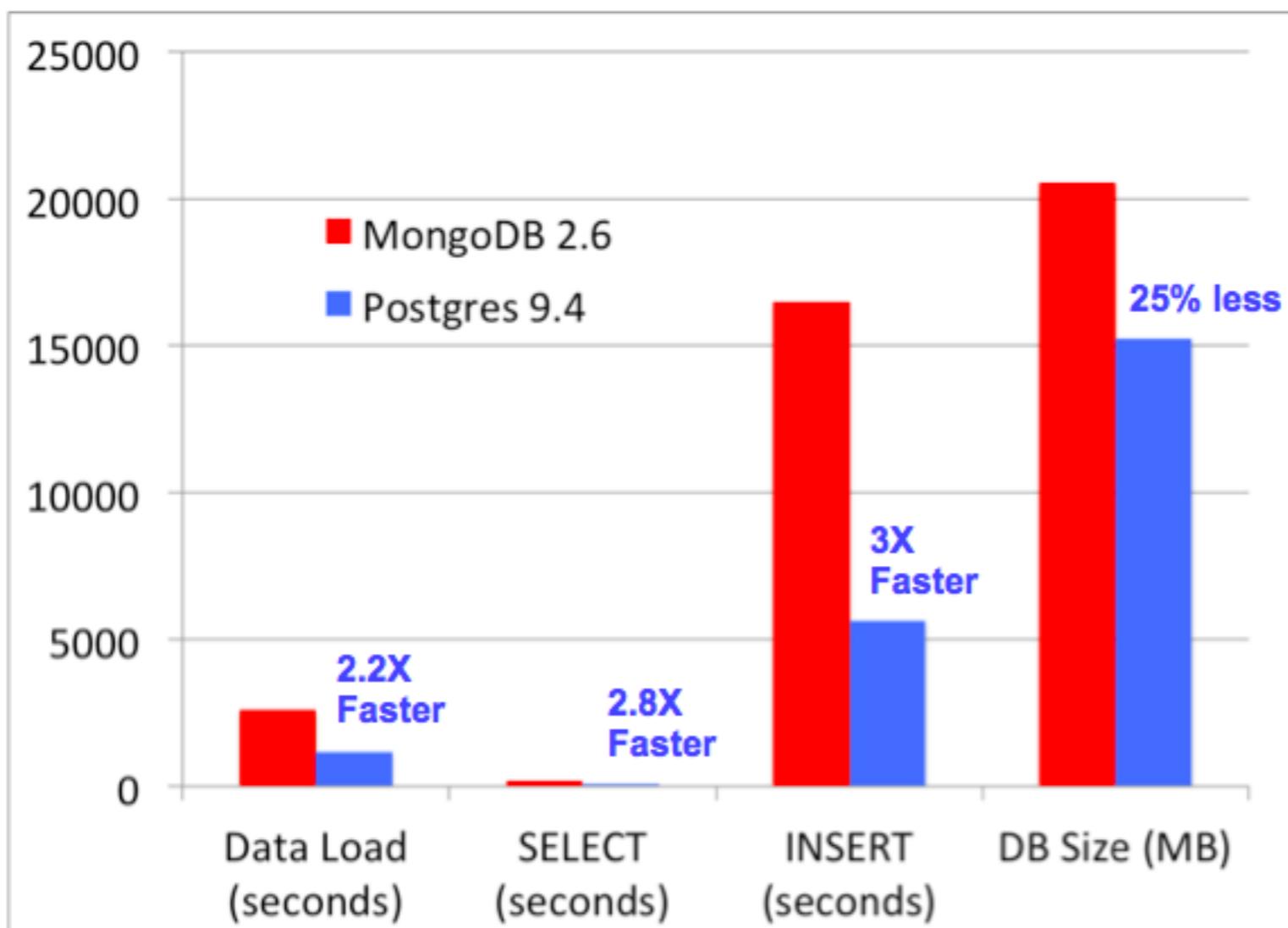


Horizontal Scaling

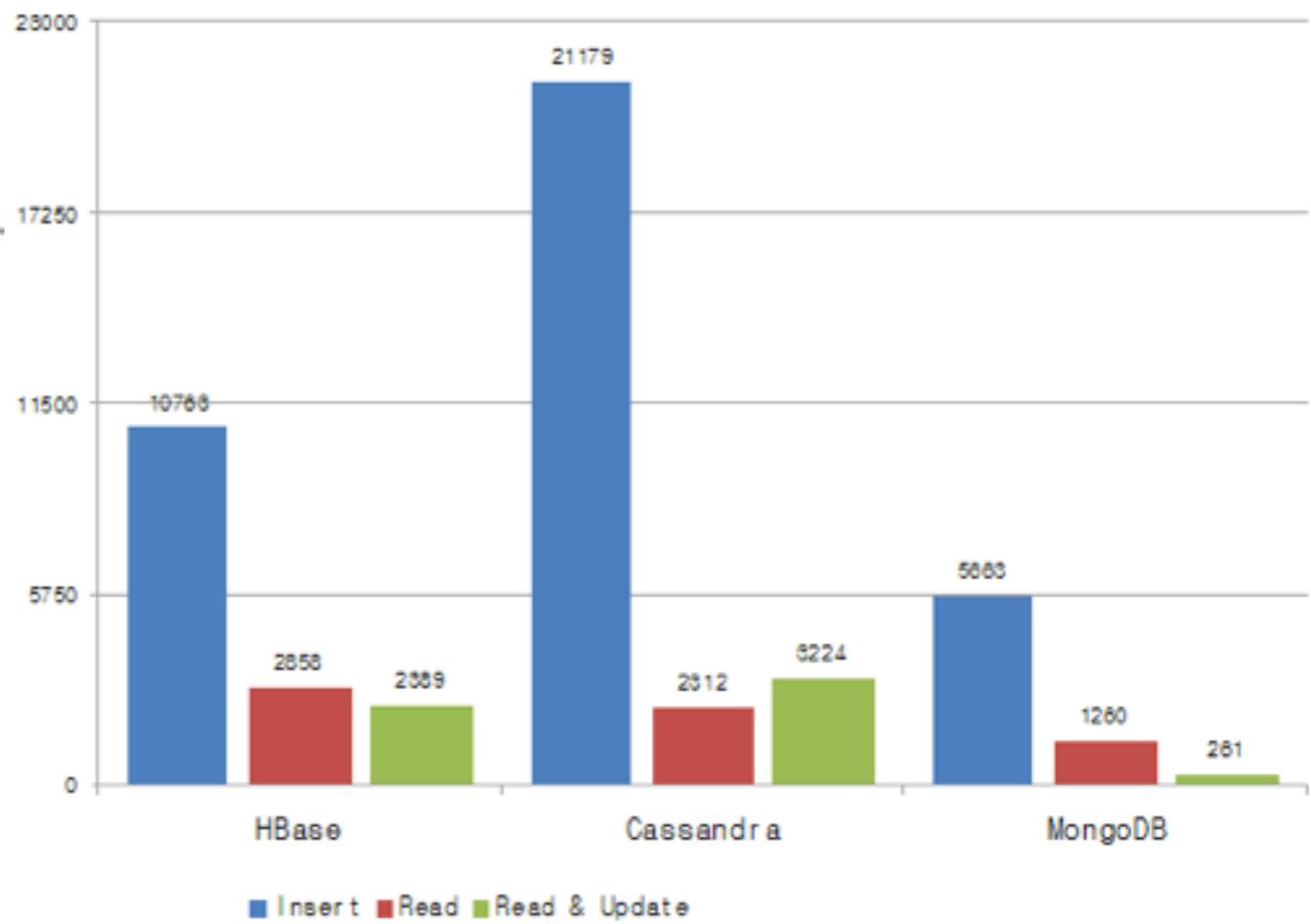
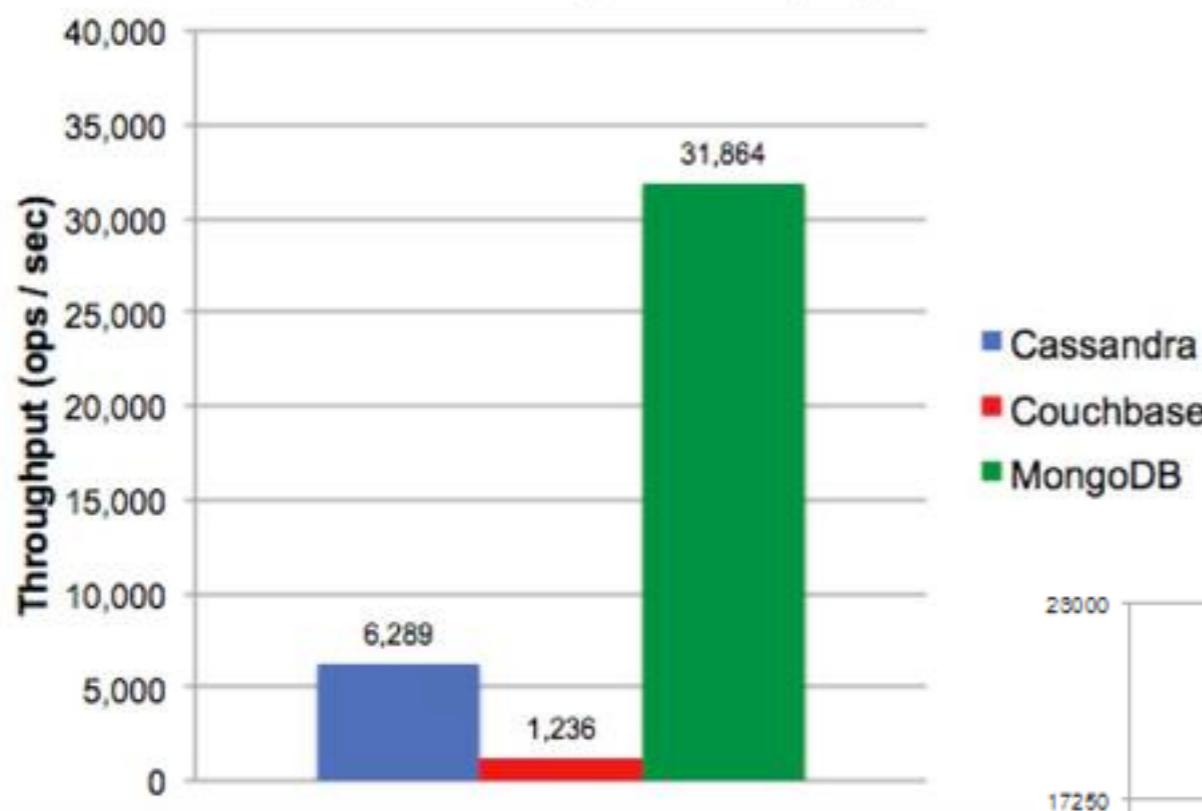
Router



MongoDB 2.6 vs PostgreSQL 9.4 Performance



YCSB – Workload A, Durability Optimized



Flexible Schema

Relational DB

1. Set up schema
2. Insert data
3. Change schema
4. Insert data with new structure
5. How do I change the schema?
Am I breaking something?
6. Insert data with new structure

MongoDB

1. Insert data
2. Insert data with new structure
3. Insert data with new structure

Problems with Flexible Schema

```
{  
  _id: 1,  
  author: { name: "Bob Davis", email : "bob@bob.com" },  
  post: "In these troubled times I like to ...",  
  date: { $date: "2010-07-12 13:23UTC" },  
}  
  
{  
  _id: 1928571982758,  
  author: { name: "Peter Brown", email : "brownp@rpi.edu" },  
  post: "First blog post ever",  
  date: { $date: "2014-11-12 13:23UTC" },  
  tags: [ "Food", "DIY" ]  
}
```

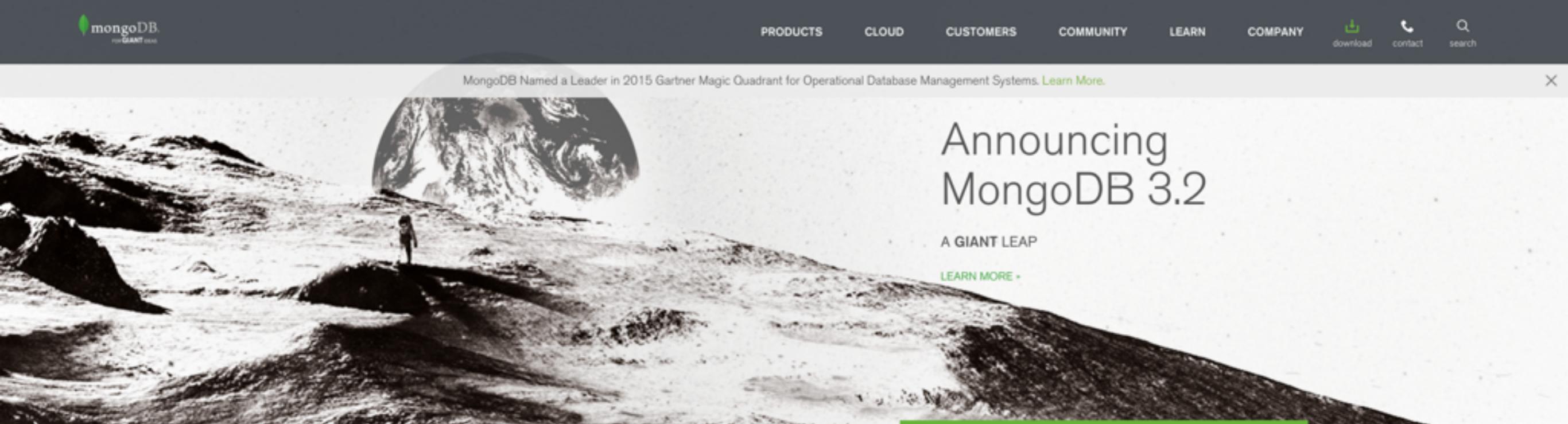
Why is MongoDB open source?

- Community
- Documentation
- Ease of adoption
- Trust in open source

“Open Source Business Model”



mongodb.com



MongoDB Named a Leader in 2015 Gartner Magic Quadrant for Operational Database Management Systems. [Learn More.](#)

PRODUCTS CLOUD CUSTOMERS COMMUNITY LEARN COMPANY

download contact search

Announcing MongoDB 3.2

A GIANT LEAP

[LEARN MORE »](#)

Download MongoDB Enterprise | Download Now



MongoDB:
A Gartner Magic Quadrant Leader

A 2015 Magic Quadrant Leader for Operational Database Management Systems

[Read The New Report](#)

Magic Quadrant for Operational Database Management Systems
12 October 2015 ID: G00271405
Analyst(s): Donald Fehlberg, Herv Adrian, Nick Heudecker, Adam M. Ronitoff, Terriy Peacock

Market Definition/Description

The operational DBMS market enjoys a dynamic resurgence, new entrants are challenging established leaders. Information managers will be especially interested in the changes in the Leaders quadrant, the strength of dBMS vendors, and the emerging feature parity of new and old vendors.

STRATEGIC PLANNING ASSUMPTIONS

By 2017, all leading operational DBMS will offer multiple data models, relational and NoSQL, in a single DBMS platform.

Through 2016, a wave of consolidation will affect the operational DBMS market as smaller companies, through mergers, acquisitions, and business become.

By 2017, the "big data" trend will cause to distinguish DBMS, which will result in it falling out of use.

**NOTE 1
DEFINITION OF AN OPERATIONAL DBMS WORKLOAD**

For the purposes of this evaluation, the workload we expect to be managed by an operational DBMS include batch/periodic, real-time or continuous data loading, transactional and analytical processing, and update transactions, operational reporting, and management of extremely distributed processes such as "near-eject" queries. Operational DBMS products must provide the ability to prioritize these multiple workloads to ensure they are met when they occur.

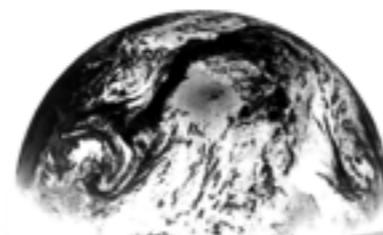
LICENSED FOR DISTRIBUTION
Gartner

Learn how Gartner can help you succeed
Become a Client now!

mongodb.org

[DOCS](#)[UNIVERSITY](#)[COMMUNITY](#)[BLOG](#)[ENTERPRISE](#)

A GIANT LEAP MongoDB 3.2

[Download MongoDB](#)[MongoDB Enterprise](#)[Learn More about MongoDB 3.2 >](#)

MongoDB Enterprise Advanced

The best way to run MongoDB in your data center. Supported. Secured. Certified.

[Learn More](#)

Join the MongoDB Community

Events

Find out when the next MongoDB event is happening near you, or meet the MongoDB team at an industry conference or trade show.

[View Upcoming Events](#)

Webinars

Attend a live broadcast and ask questions with a MongoDB expert. We host weekly sessions on a variety of development and operations topics.

[View Upcoming Webinars](#)

Launch your
GIANT idea

[Download MongoDB](#)[MongoDB Cloud Manager](#)

MongoDB v3.0 is now available



Awesome Documentation & Language Support

Community website <https://www.mongodb.org/>

Online classes, tutorials

Actionscript, C, C#, C++, Clojure, ColdFusion, D, Dart, Delphi, Erlang, Go, Groovy, Haskell, Java, JavaScript, Lisp, Lua, MatLab, Perl, PHP, PowerShell, Prolog, Python, R, Ruby, Rust, Scala, Smalltalk

What's it like working at an
open source company?



Wide variety of people to interact with

Limewire, Turntable.fm

Ambitious in its mission

Tough Problems with a **LOT** of public attention

- Scalability
- Fault tolerance/Availability
- Consistency
- Speed
- Putting up with Hacker News

The screenshot shows the GitHub profile for the MongoDB repository. At the top right, the 'Star' button is circled in red. Below it, the star count is displayed as 8,309. Other metrics shown include 32,366 commits, 15 branches, 294 releases, and 234 contributors. The bottom navigation bar includes links for 'New pull request', 'Find file', 'HTTPS', and download options.

mongodb / mongo

Code Pull requests 10 Pulse Graphs

32,366 commits 15 branches 294 releases 234 contributors

Branch: master New pull request Find file HTTPS https://github.com/mongo/ Download ZIP



Raymond Jacobson

raymondjacobson

 Rensselaer Polytechnic Institute

Contributions Repositories Public activity Edit profile

Popular repositories

-  [monsoon](#) 8 ★
Unlimited* cloud storage
-  [cranium.css](#) 5 ★
ML approach to tailored web UI, Hack Upstate...
-  [sprawl](#) 3 ★
Hack Upstate IV Syracuse, N.Y. October 2014
-  [bubblio](#) 3 ★
Hack Upstate I Syracuse, N.Y. April 2013
-  [matlab](#) 3 ★
matlab for python

Repositories contributed to

-  [whereuat/whereuat-ios](#) 1 ★
The iOS app for whereu@.
-  [mongodb/bsonspec.org](#) 57 ★
site for bsonspec.org
-  [adamchel/mongo-tcp-replay](#) 0 ★
A configurable tool used to replay a distributed...
-  [mongodb/mongo](#) 9,055 ★
The Mongo Database
-  [adamchel/mongo-replay-tools](#) 0 ★
Tools to capture MongoDB state and network ...



When everyone's watching, you've got to
stay hip with hottest new technology stack

Databases are an UNSOLVED problem

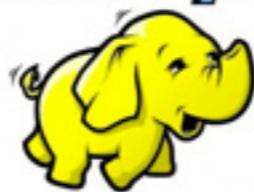


redis



Neo4j

hadoop



RethinkDB.



HYPERTABLE^{INC}

SO MANY OF THEM ARE OPEN SOURCE!!!

Pluggable Storage Engines

(The future of database design)

Storage Engine API

- MMAPv1 (Default in < v3.0)
- WiredTiger (Default in v3.2) - BerkleyDB
- 3rd Party Storage Engines



<http://www.slideshare.net/mongodb/storage-engine-wars-at-parse>

Lots of potential storage engines

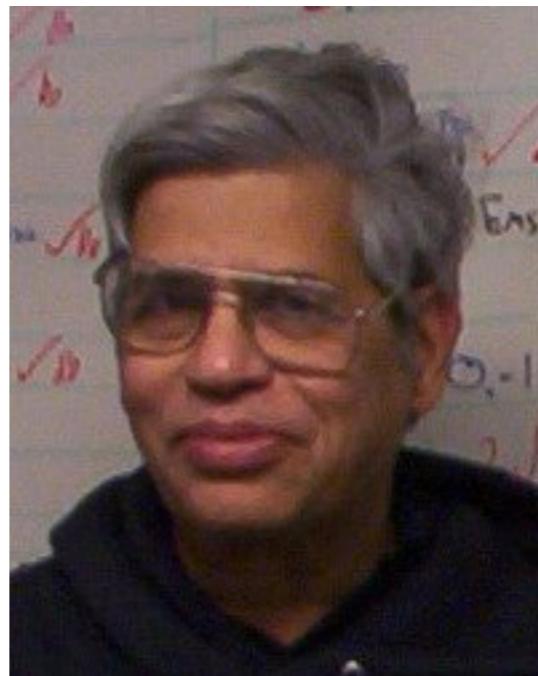
- Oracle
- Facebook
- Solaris
- Machine Learning!?

pluggable storage engines are
possible because of

OPEN SOURCE SOFTWARE

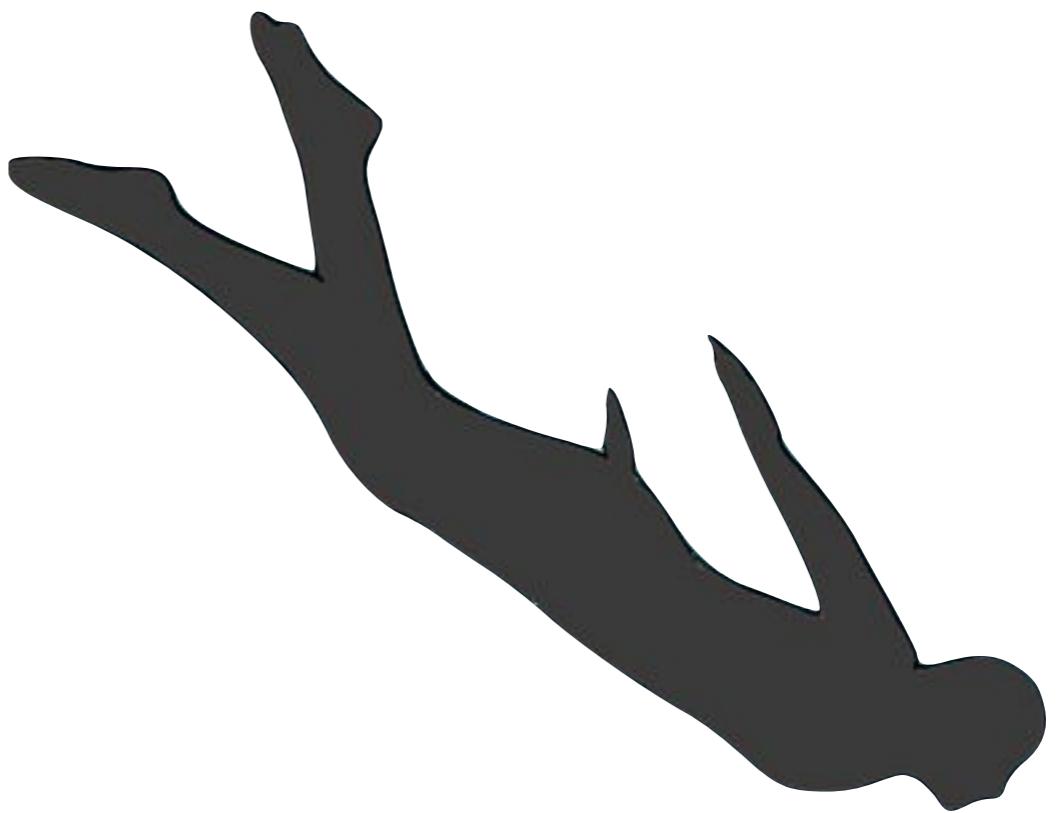
OPEN SOURCE

SOFTWARE



Tech dive

- Run a server
- Connect to a server
- Create a document
- Find a document
- Update a document
- Pluggable storage engines (Mongo 3.0)



Thanks!

Understood: MongoDB doesn't solve all the world's problems. It's a different paradigm solution. SQL has its benefits too.

But, a plea: Please don't buy into everything you read on Hacker News. People are misinformed & **insane**.

(but also funny)

<https://www.youtube.com/watch?v=b2F-DItXtZs>

www.mongodb.com/careers/college-students

me@raymondjacobson.com