A Big Data Arsenal for the 21st Century

Matt Asay

VP, Marketing & Business Development, MongoDB

mongoDB

7 million
downloads

150,000
online
education
registrations

1,000 active
subscribers

20,000
MongoDB Days
attendees

30,000 user
group members

World's fastest-
growing database

What We Don't Do

"The relational database market is a $9 billion a year market. I want to shrink it to $3 billion and take a third of the market."

- Marten Mickos

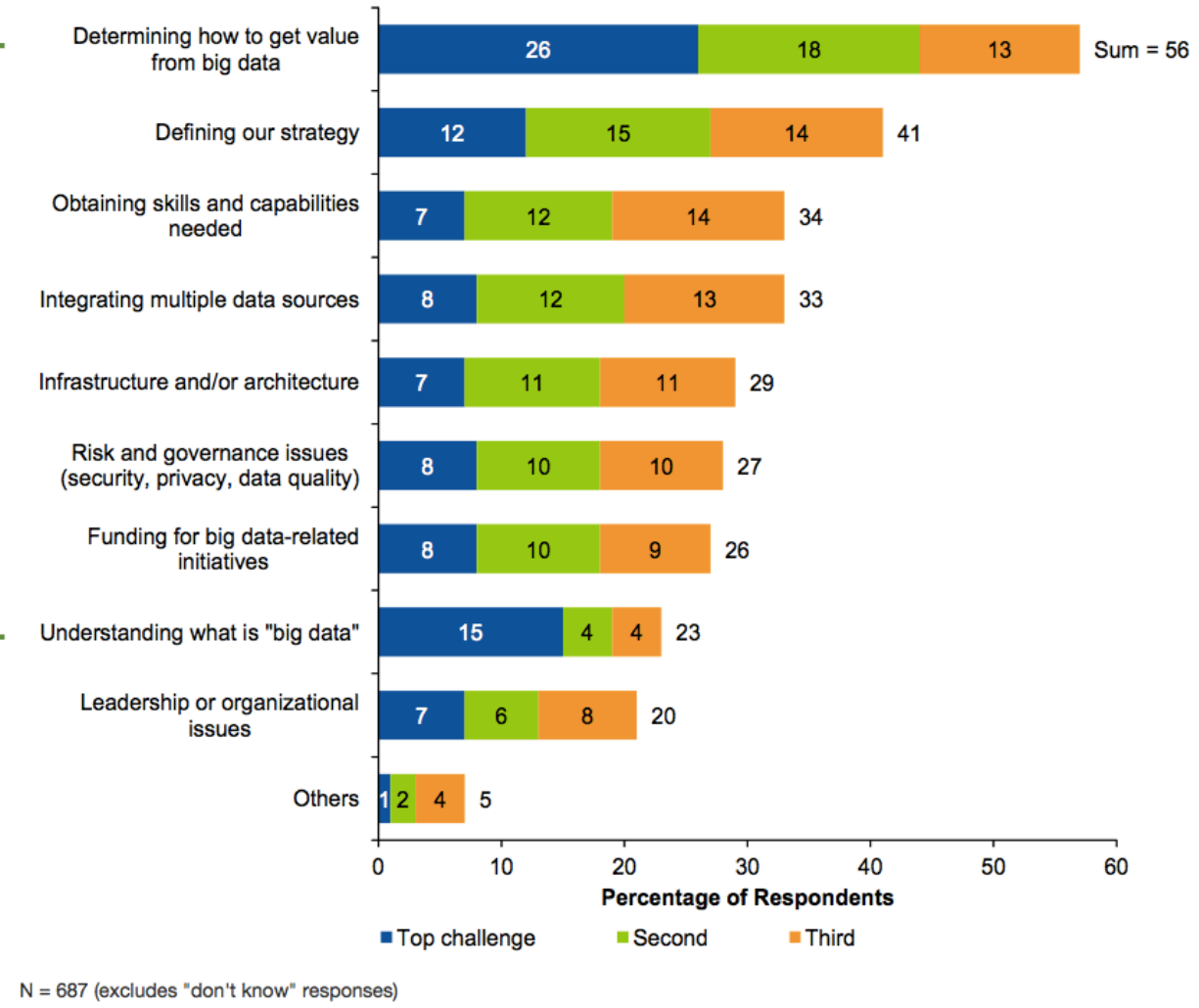
What We Do

Enable a Generation of Innovative, Modern Applications Previously Impossible Or Too Difficult to Achieve


The Big Data Unknown

Top Big Data Challenges?

Translation? Most struggle to know what Big Data is, how to manage it and who can manage it

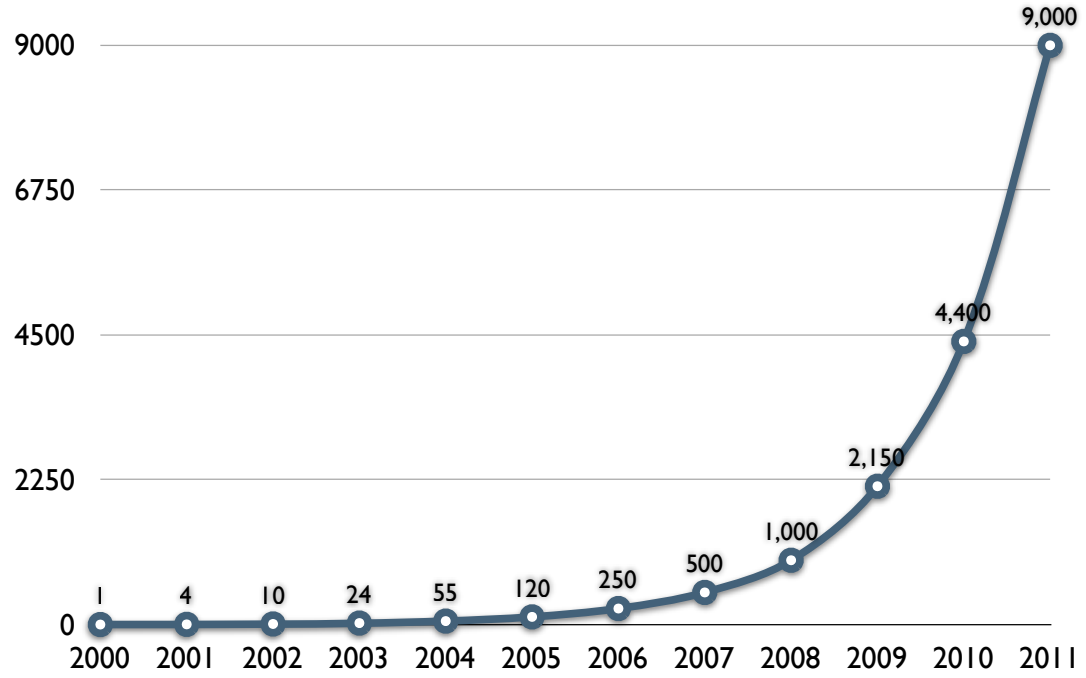


Source: Gartner

mongoDB

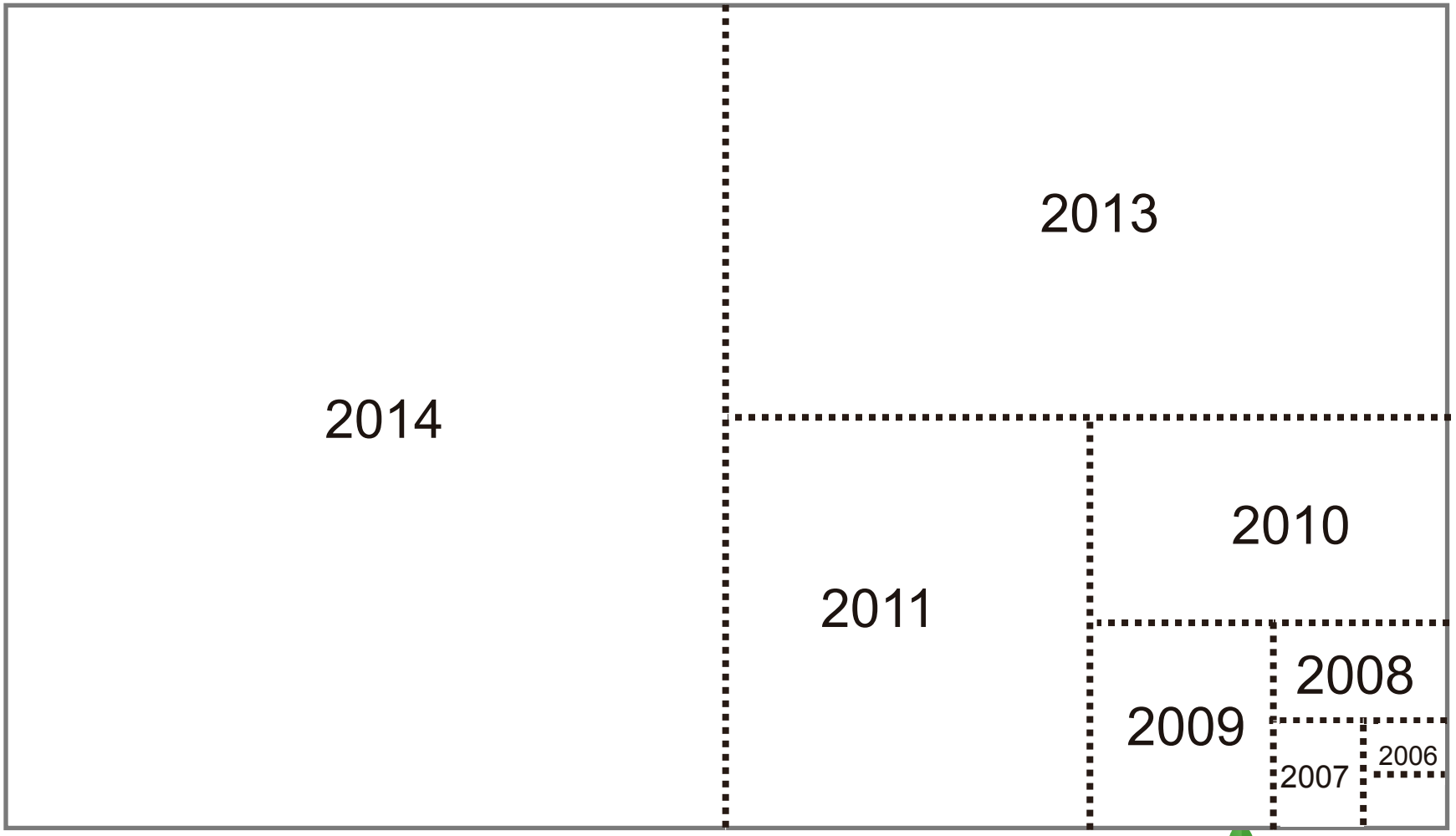
Big Data Is Sort of a Matter of Volume

- More than 90% of today's data was created in the last 2 years
- Moore's Law for data: Doubles at regular intervals



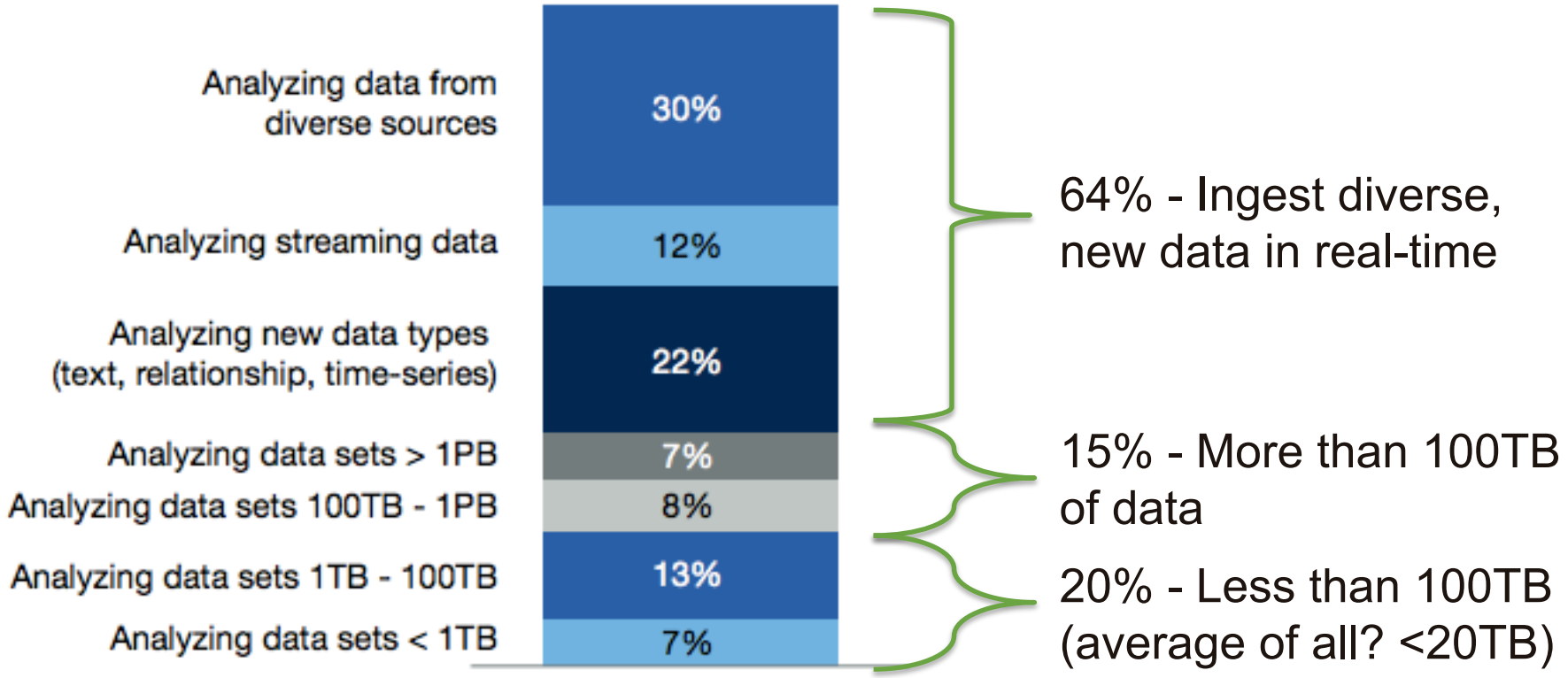
mongoDB

Big(ger) Is the New Normal



mongoDB

Understanding Big Data – It's Not Very "Big"



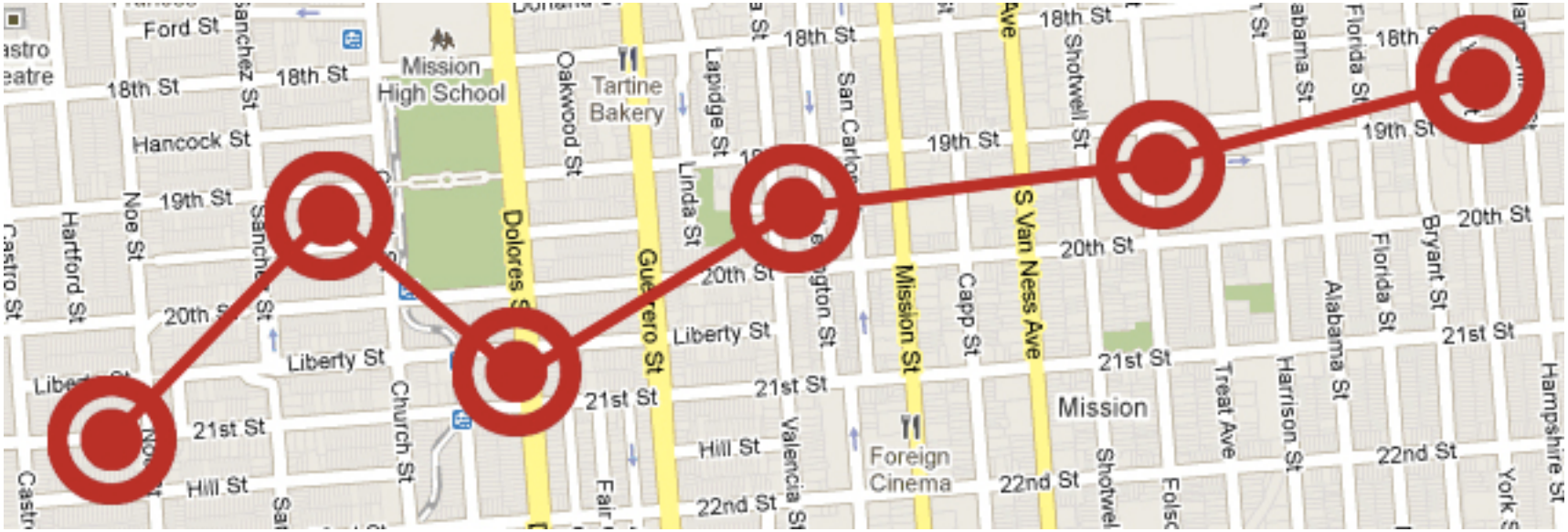
from Big Data Executive Summary – 50+ top executives from Government and F500 firms

mongoDB

Modern, Big Data Is Messy

mongoDB

Data Now Looks Like This



mongoDB

And This

And This

mongoDB

Doesn't Fit Neatly into a "Spreadsheet"

- 90% of the world's data was created in the last two years
- 80% of enterprise data is unstructured
- Unstructured data growing 2X faster than structured

Back in 1970...Cars Were Great!

The New American Car.

This is the American Motors Gremlin.
It is the kind of car this country has needed for a long, long time.
It is designed to give the American motorist a car that is easy to buy, easy to handle, easy to take care of, and, at the same time, fun to drive.
The Gremlin is the smallest production car made in America.
It is 161 inches long, just 2½ inches longer than the Volkswagen.
Yet its turning circle, at 32 feet, 8 inches, is about 3 feet less than VW's.
Which makes the Gremlin about the easiest car in the world to park and handle.
The Gremlin gets the best gas mileage of any car made in America. It goes about 500 miles without stopping for gas.
This is great gas mileage, when you consider that the Gremlin has a bigger standard engine than any car near its size and price. 128 hp to VW's 57.
This engine gets from 0 to 60 in 15.3 seconds, the pickup you need on expressways.
And nobody's going to push you around in a Gremlin. It is 10 inches wider, 7 inches lower and 765 pounds heavier than a VW.
Which gives you about the smoothest, most stable ride possible in a car this size.
The Gremlin is remarkably easy to service and maintain.
Its normal oil change interval is 6 months or 6,000 miles; lubrication is normally needed only every 24,000 miles.
There are two basic Gremlin models.
A two-passenger, with storage area in the rear.
A four-passenger with fold-down rear seats for extra storage and flip-up rear window for easy access.
Both models cost about what you'd pay for an imported economy car.
The four-passenger lists for $1,959!
The lowest list price of any car made in America.
Except for the two-passenger Gremlin.
It lists for $1,879!
Which is quite a bargain, when you consider what you get for your money.
The new American car.

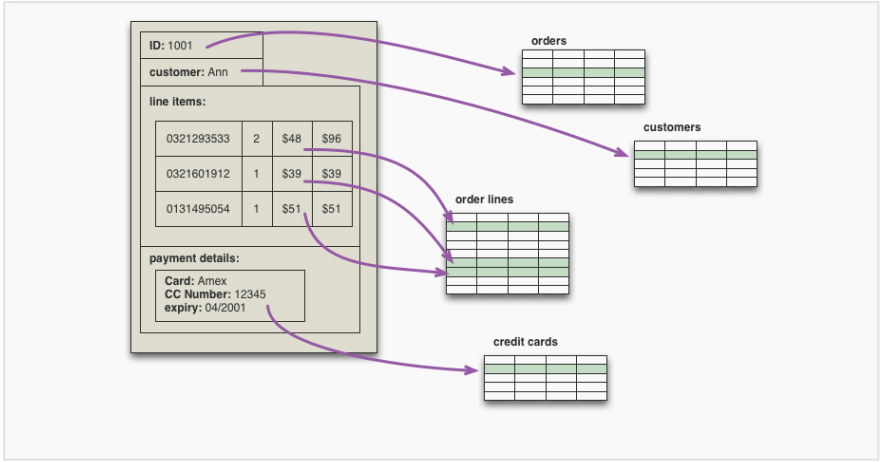
American Motors Gremlin

$1,879* 2-Passenger **$1,959*** 4-Passenger

So Were Computers!

mongoDB

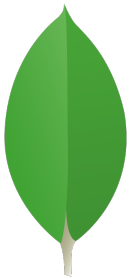
Lots of Great Innovations Since 1970



mongoDB

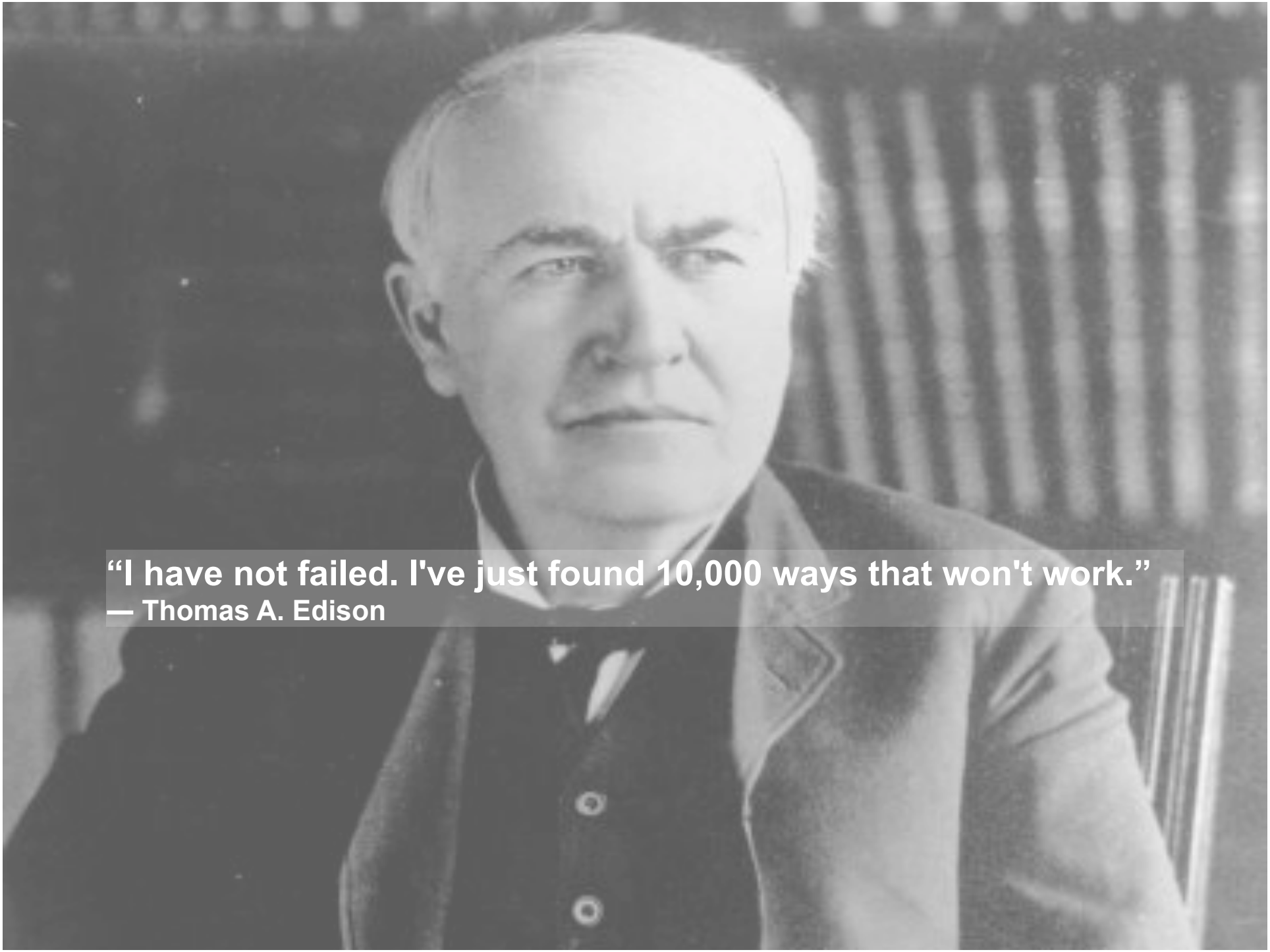
New Tools for New Data

elasticsearch.



mongoDB

Innovation As Iteration


“I have not failed. I've just found 10,000 ways that won't work.”
— Thomas A. Edison

Must Be Open Source

Modernizing IT

ERIC KNORR

SEPTEMBER 18, 2012

Open source in 2012: Bigger and better than ever

This year's Best of Open Source Software awards includes a whopping 125 products in 7 categories. The real story is the technology leadership so many of these products display

Must Not Require Big Upfront Payment

mongoDB

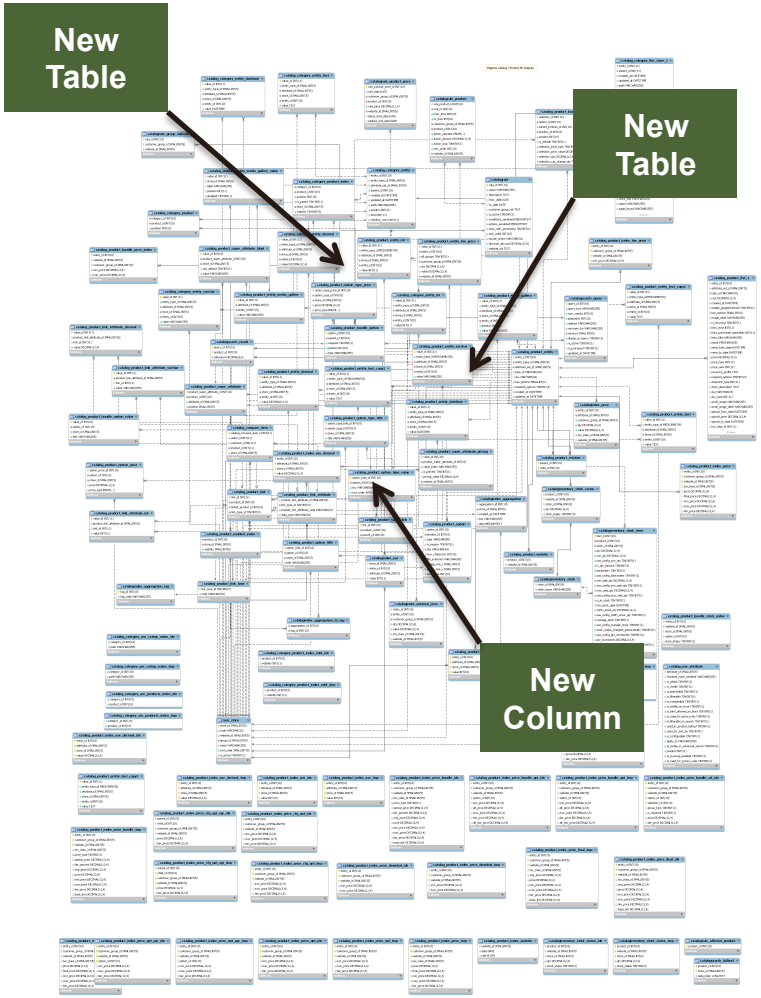
Must Not Penalize Success

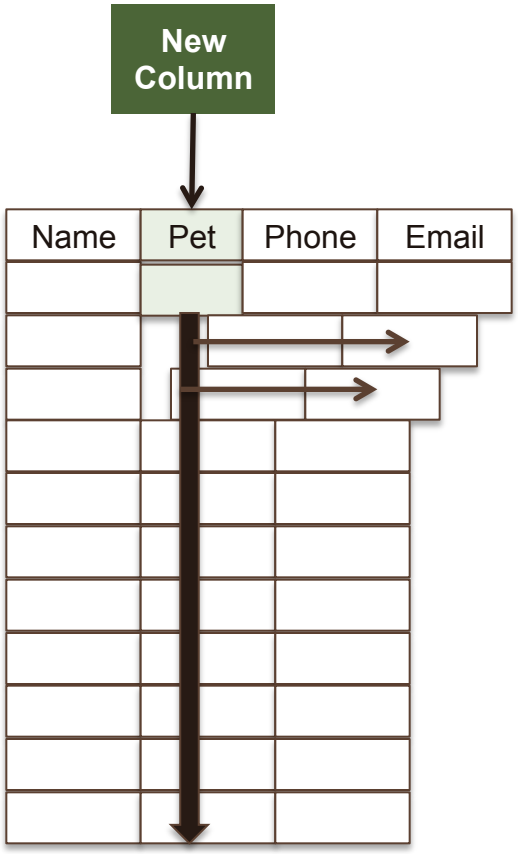
“Clients can also opt to run zEC12 without a raised datacenter floor -- a first for high-end IBM mainframes.”

IBM Press Release 28 Aug, 2012

mongoDB

Must Not Impede Iteration





3 months later…

mongoDB

Must Be Based on Industry Standards

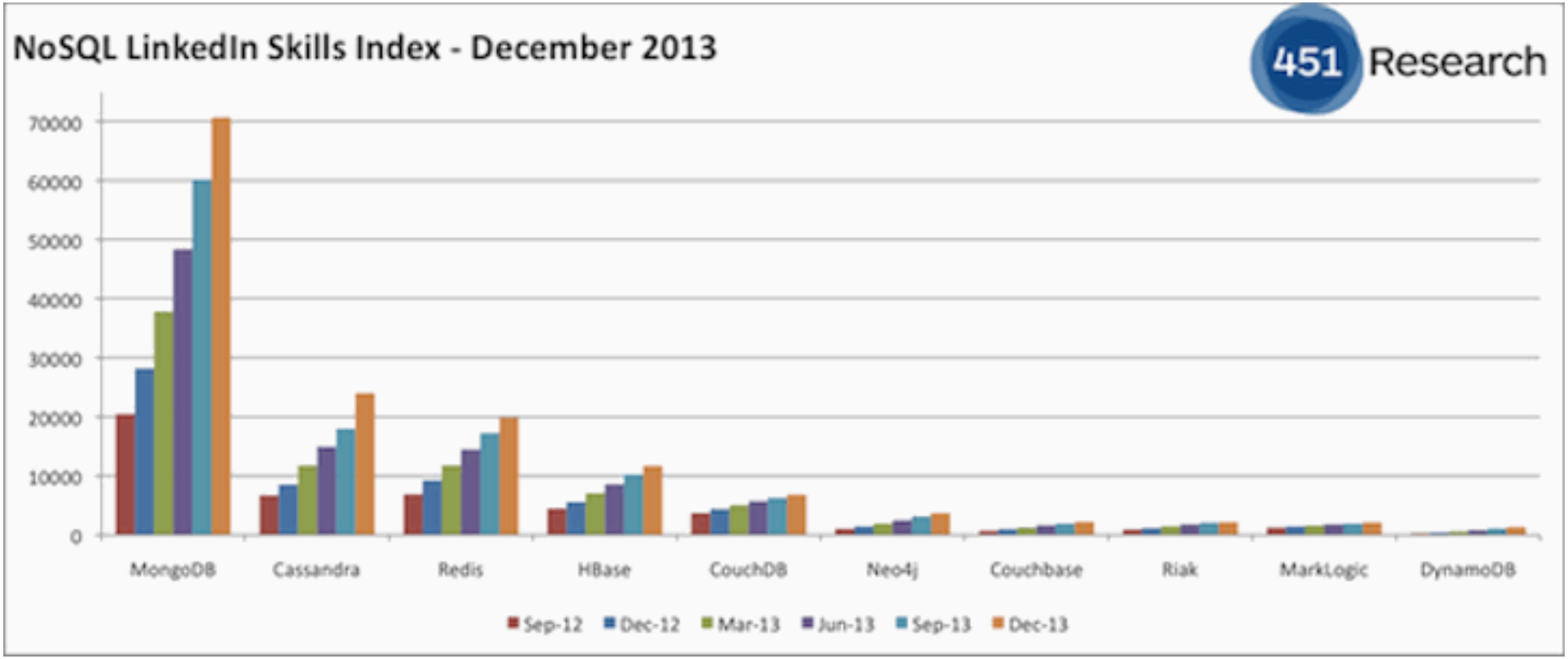
databaseEZ

DB-Engines.com Database Ranking

Ranking	Database	Type	Score	Changes
1	Oracle	Relational	1491.8	-8.43
2	MySQL	Relational	1290.21	1.83
3	Microsoft SQL	Relational	1205.28	-8.99
4	PostgreSQL	Relational	235.06	4.61
5	**MongoDB**	**Document**	**199.99**	**4.81**
6	DB2	Relational	187.32	-1.14
7	Microsoft Access	Relational	146.48	-6.4
8	SQLite	Relational	92.98	-0.03
9	Sybase	Relational	81.55	-6.33
10	Cassandra	Wide Column	78.09	-2.23

mongoDB

Must Be Easy to Find Skills



mongoDB

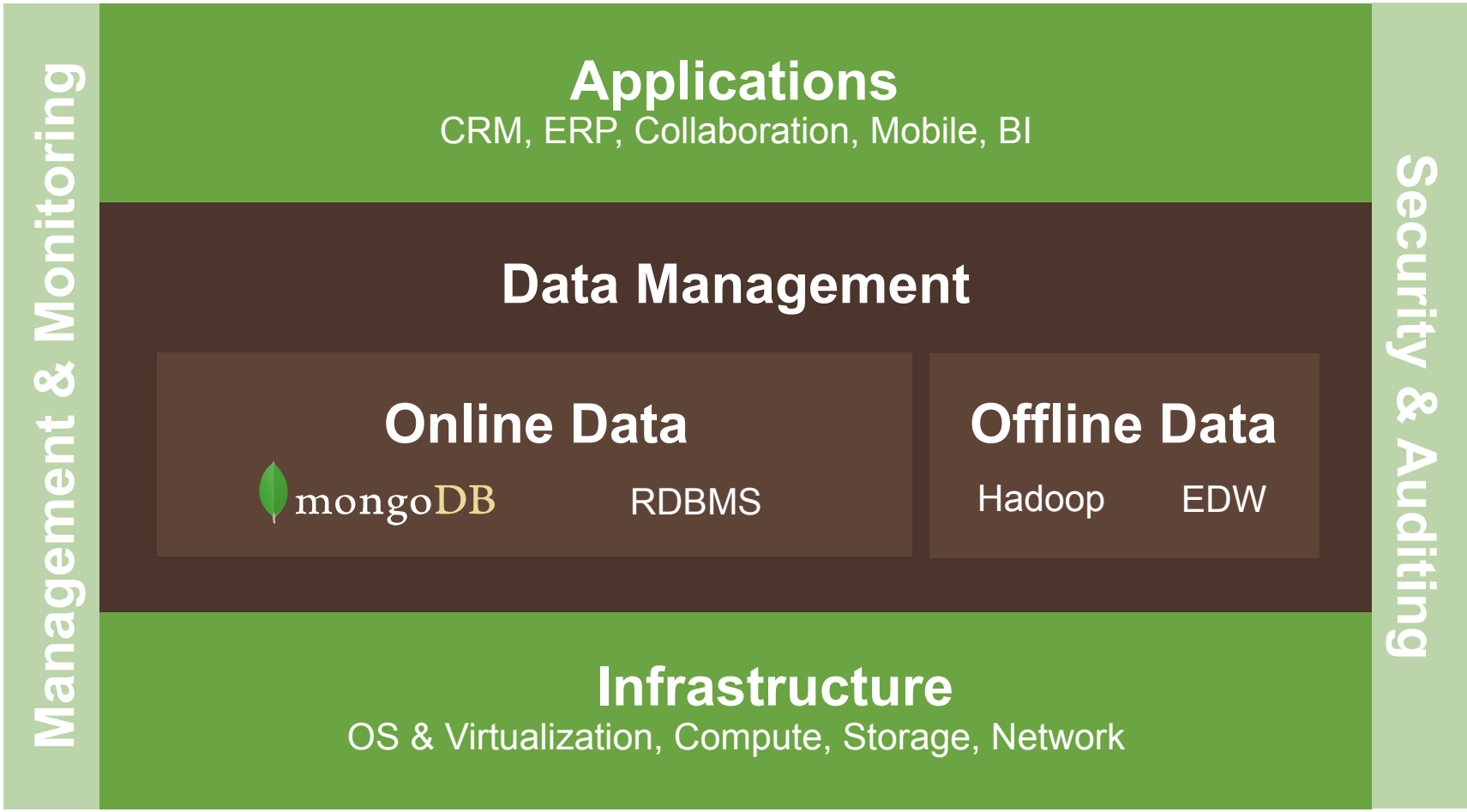
Must Be Easy to Learn/Use

"Organizations already have people who know their own data better than mystical data scientists….Learning Hadoop [or MongoDB] is easier than learning the company's business."

(Gartner, 2012)

When To Use Hadoop, Modern Databases

Enterprise Big Data Stack



mongoDB

Consideration – Online vs. Offline

Online *vs.* Offline

- Real-time
- Low-latency
- High availability

- Long-running
- High-Latency
- Availability is lower priority

Hadoop Is Good for…

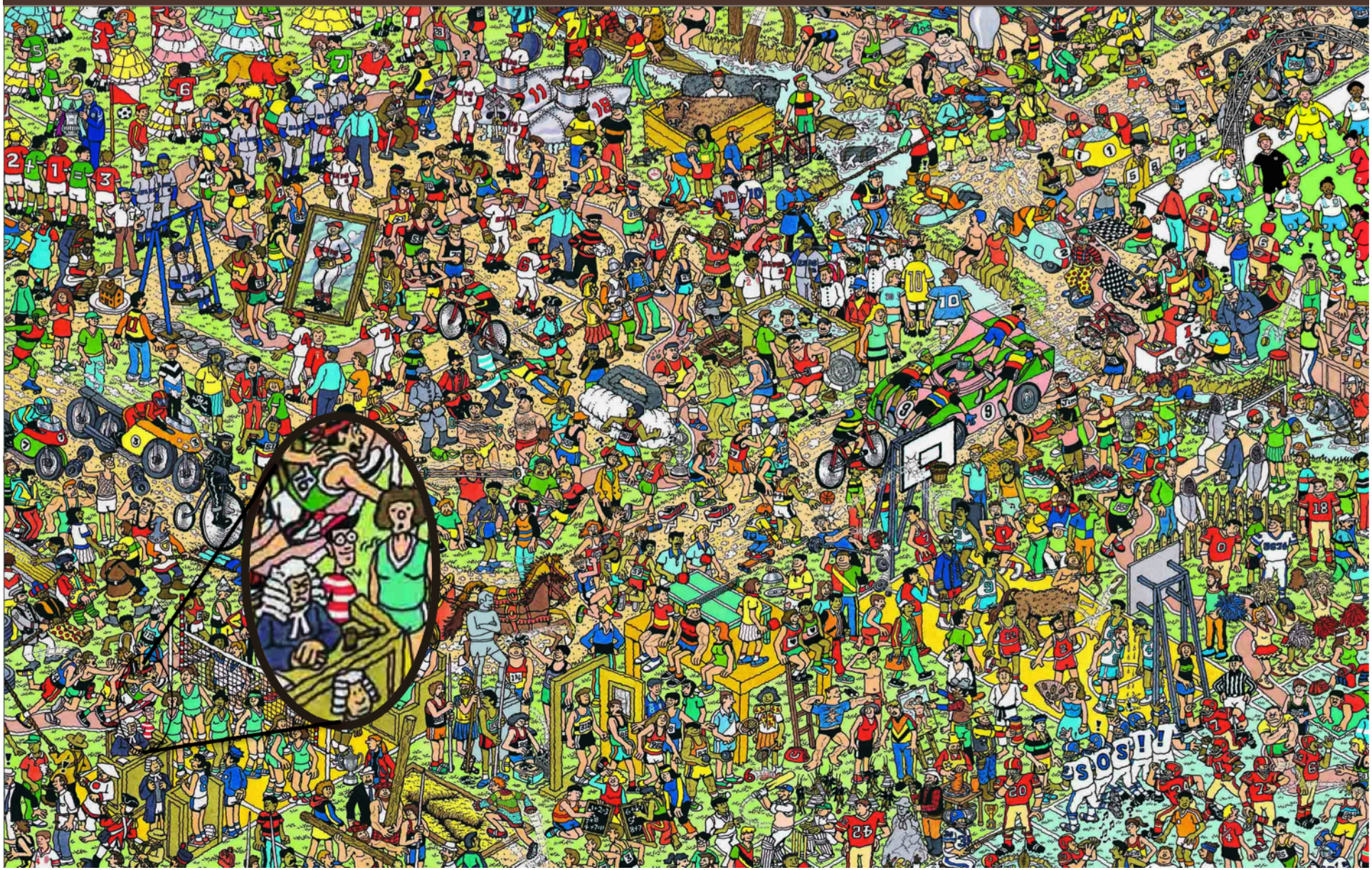
- Risk Modeling
- Churn Analysis
- Recommendation Engine
- Ad Targeting
- Transaction Analysis
- Trade Surveillance
- Network Failure Prediction
- Search Quality
- Data Lake

MongoDB/NoSQL Is Good for…

- 360° View of the Customer
- Mobile & Social Apps
- Fraud Detection
- User Data Management
- Content Management & Delivery
- Reference Data
- Product Catalogs
- Machine to Machine Apps
- Data Hub

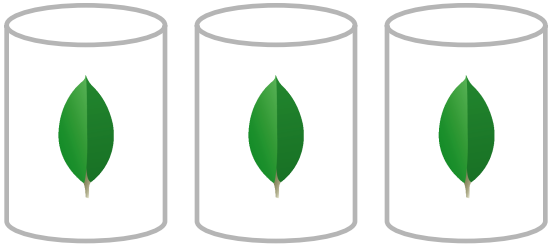
How To Use The Two Together?

Finding Waldo



Predictive Analytics

Government



- Predictive analytics system for crime, health issues
- Diverse, unstructured (incl. geospatial) data from 30+ agencies
- Correlate data in real-time

MongoDB

+ Hadoop

Algorithms

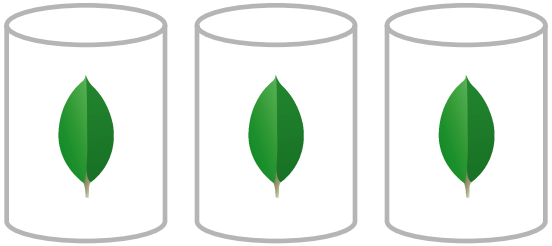
- Long-form trend analysis
- MongoDB data dumped into Hadoop, analyzed, re-inserted into MongoDB for better real-time response

mongoDB

Machine Learning

criteo

Ad-Serving



- Catalogs and products
- User profiles
- Clicks
- Views
- Transactions

MongoDB Connector for Hadoop

Algorithms

- User segmentation
- Recommendation engine
- Prediction engine

mongoDB

Remember…

- Modern data is messy
- Your data infrastructure ***must*** support iteration
- Modern data infrastructure market is crowded
  - But clear winners are distinguishing themselves
  - Bet on general purpose over niche, popular over obscure, open source over proprietary
- Use MongoDB + Hadoop together

@mjasay

Don’t believe me?
MongoDB booth on Floor 3