

# BIG DATA TRENDS

11/15/2012

David Feinleib

THE BIG DATA GROUP

[THEBIGDATAGROUP.COM](http://THEBIGDATAGROUP.COM)

# WHAT IS BIG DATA?







Elevation

Speed



Total Time	h:m:s	Total Distance	mi	Elevation	ft	Speed	mph
00:01:32	07:13:43	0.00	108.14	-18	3,801	2.8	30.9



Price: \$399.99

+

GARMIN<sup>®</sup> Connect™

Home

Features

Get Started

[Sign in](#) | [Help](#) | [English](#)

## Total Miles Logged by Connect Users

2,509,889,029

- 1) **Landscape** - Number of Big Data Companies and Categories Is Growing
- 2) **Basic Stats** - Consumer Scale Driving Big Data Transformation
- 3) **Transformation** - Many Industries
- 4) **Top 8 Laws Of Big Data**
- 5) **Big Data Whitespace**

# BASIC STATS - BIG DATA MARKET ROBUST, CONSUMER SCALE FUELING TECH ADOPTION



# BIG DATA LANDSCAPE

## Apps

### Vertical

MYRRIX Placedio bloomreach  
ellucian. Atgeo PREDICTIVE POLICING

### Operational Intelligence

splunk sumologic  
loggly New Relic

### Ad/Media

METAMARKETS collective  
rocketfuel Recorded Future  
bluefin Quantivo  
collective the audience engine  
Media Science TURN DataXu  
LUCKYSORT

### Business Intelligence

ORACLE Hyperion  
SAP Business Objects RJMetrics  
Microsoft Business Intelligence  
IBM COGNOS birst  
pentaho JASPERSOFT  
MicroStrategy  
Autonomy bime  
Chart.io  
WATTIVO GoodData

### Analytics and Visualization

Palantir  
tableau metaLayer dataspora  
OPERA centrifuge  
TERADATA ASTER  
sas TIBCO KARMASPHERE  
panopticon  
Datameer 1010data  
platform ClearStory CIRRO  
alteryx visual.ly  
QlikView AYATA

### Data As A Service

factual FICO kaggle knoema  
GNIP DATASIFT Windows Azure Marketplace INRIX LexisNexis  
LOGATE

## Infrastructure

### Analytics

Hortonworks VERTICA MAPR  
cloudera INFOBRIGHT  
EMC PARACCEL  
NETEZZA GREENPLUM  
DATASTAX EXASOL koginitio

### Operational

Couchbase 10gen  
HADAPT  
TERRACOTTA VoltDB  
MarkLogic INFORMATICA

### As A Service

Qubole amazon web services  
Windows Azure MORTAR  
infochimps  
Google BigQuery

### Structured DB

ORACLE MySQL  
Microsoft SQL Server PostgreSQL  
IBM DB2 SYBASE  
memsql TERADATA



Technologies



APACHE HBASE

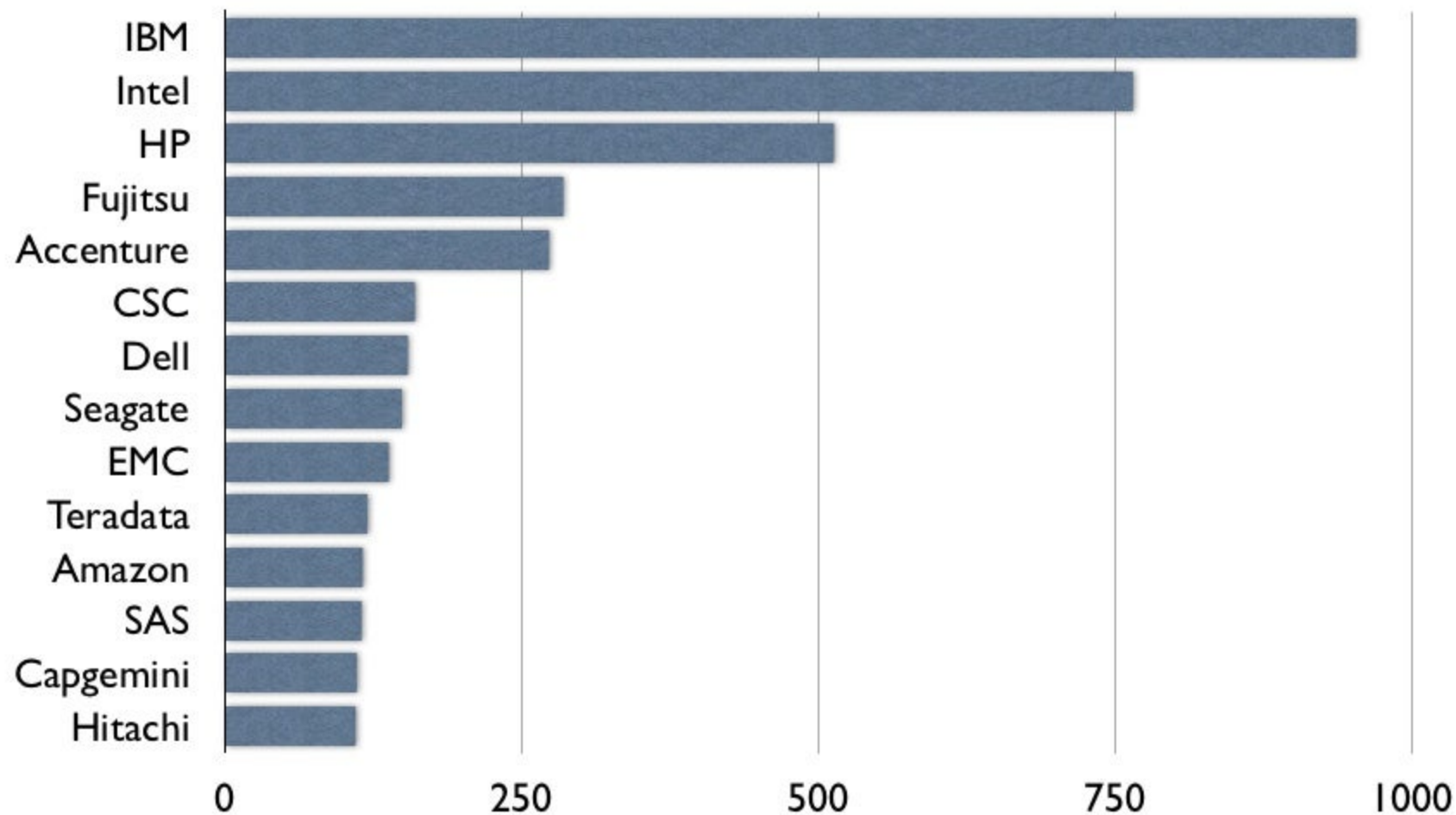
Cassandra



# MARKET GROWTH

# Big Data Overall Revenue \$5.1B in '11

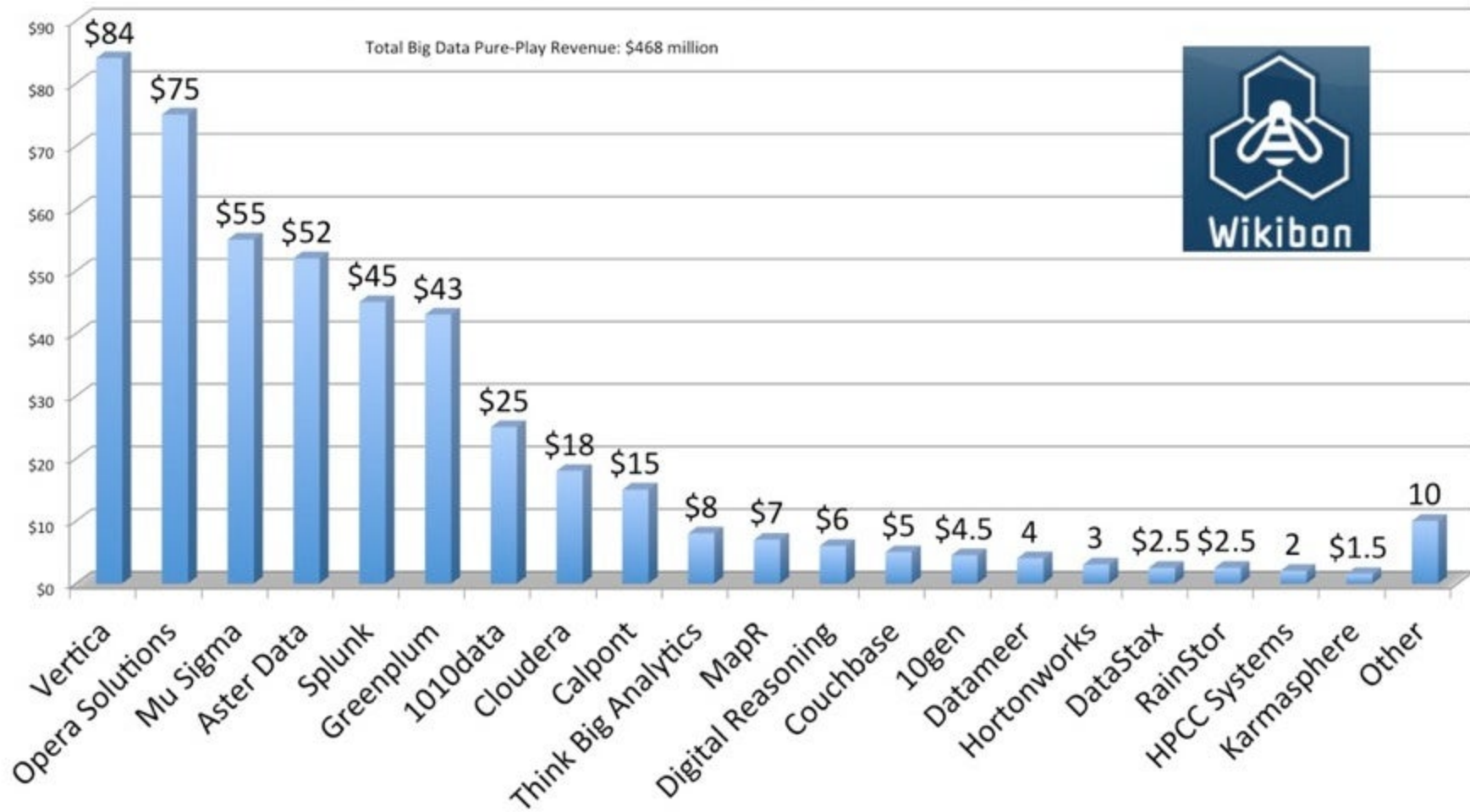
Vendors With Big Data Revenues Over \$100M



Source: Wikibon

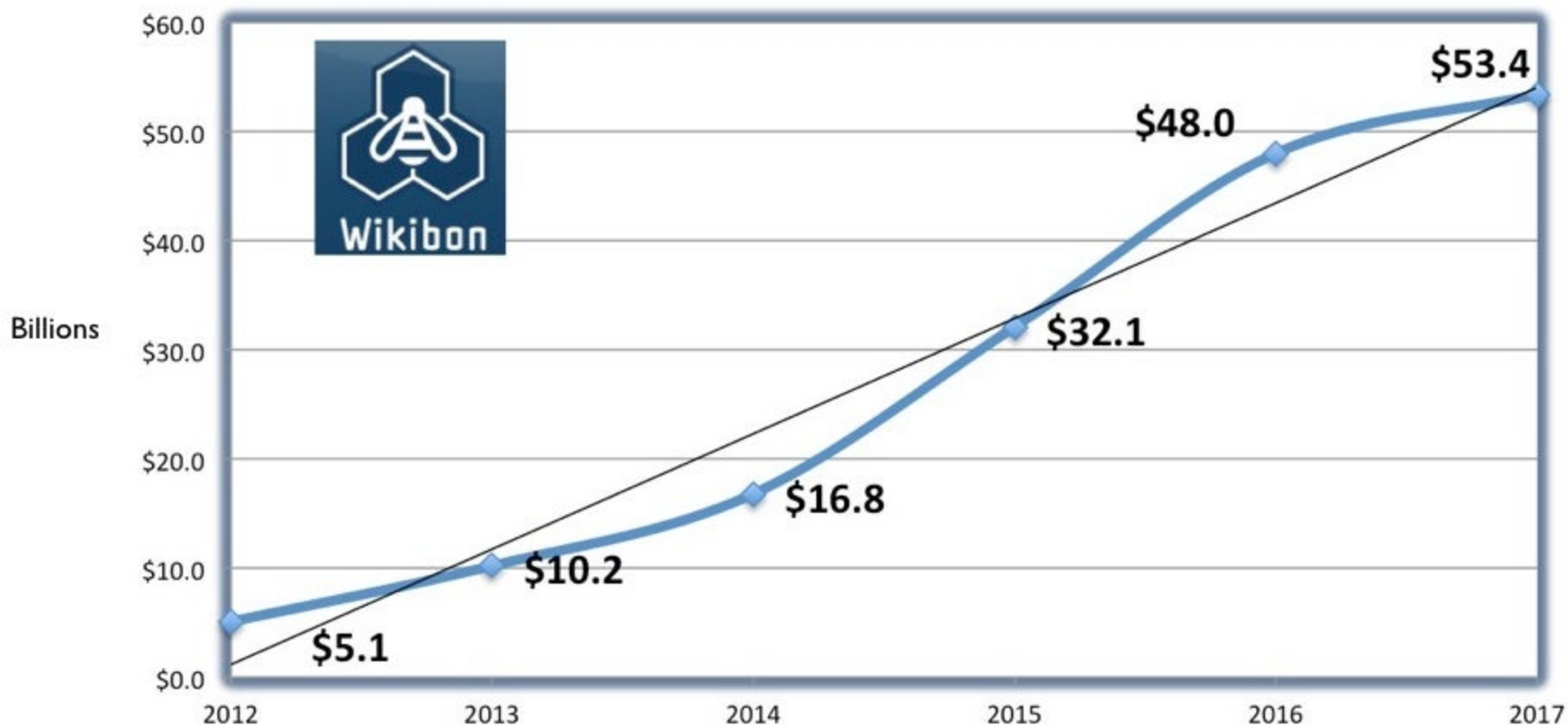
# Big Data Pure-Play Revenue \$468M in '11

2011 Big Data Pure-Play Vendors, Yearly Big Data Revenue (in \$US millions)





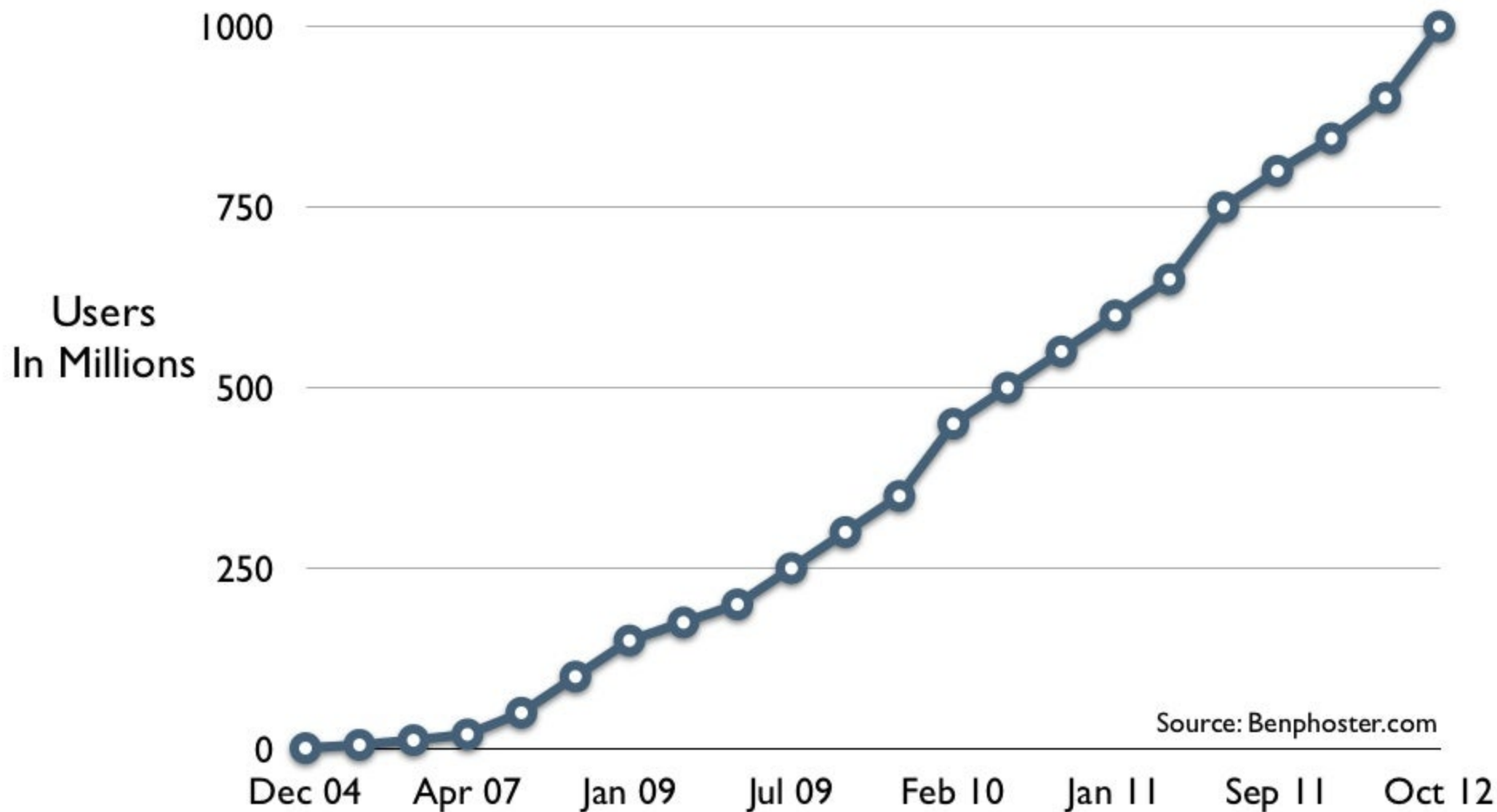
# Big Data Overall Revenue \$53.4B By 2017



Source: Wikibon

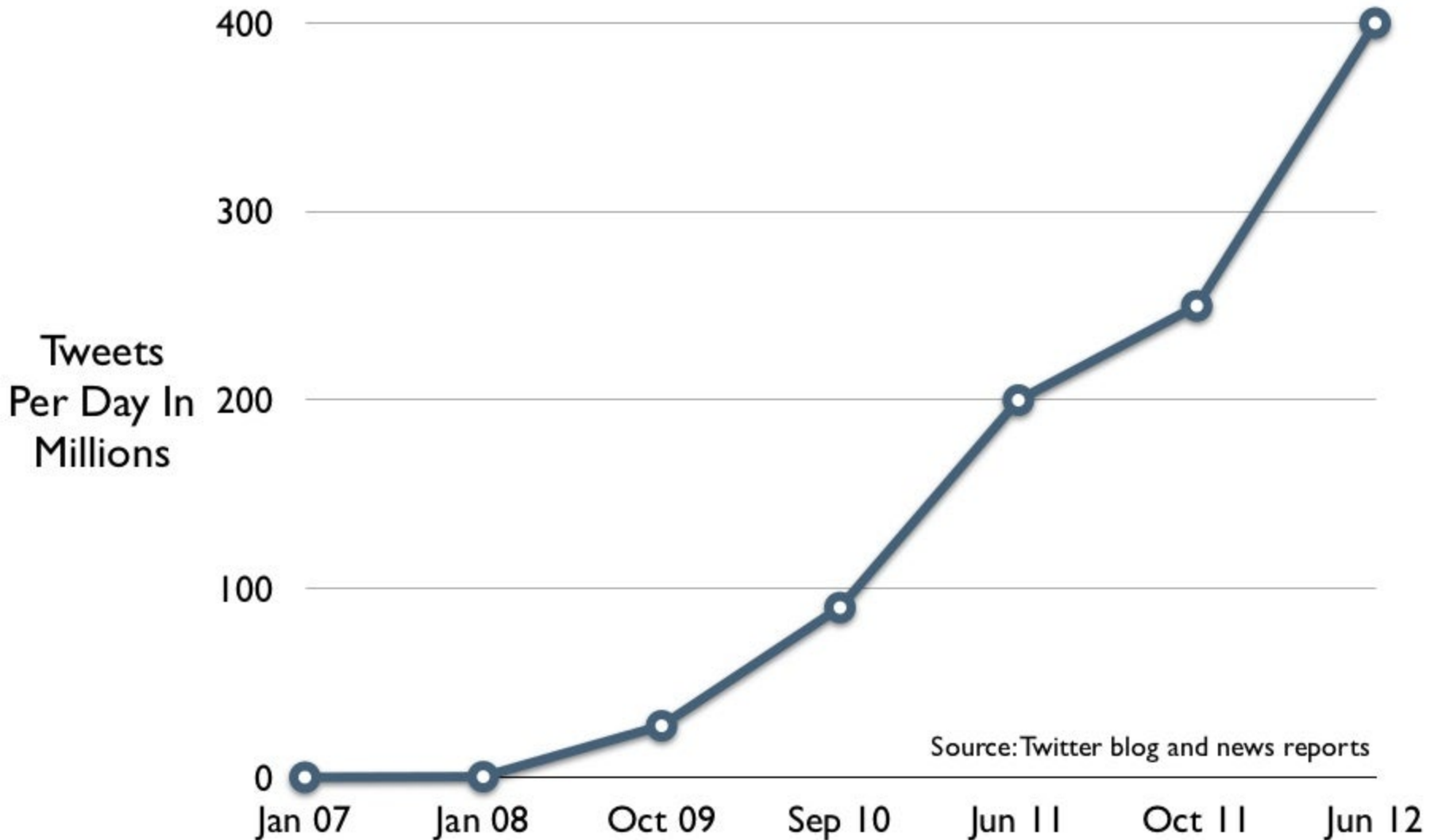
# DATA GROWTH

# Facebook at 1B Users in Oct '12

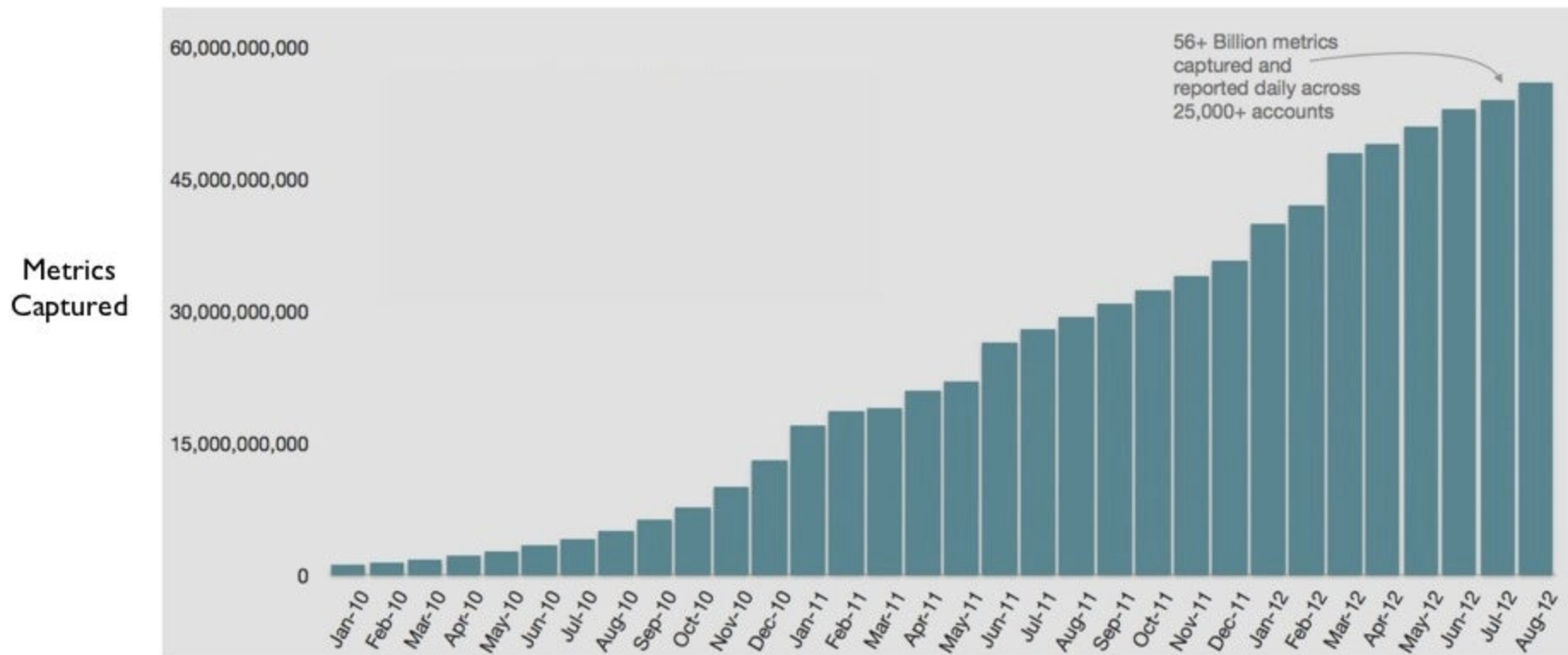




# Twitter at 400M Tweets Per Day in Jun '12



# New Relic at 56B Metrics Captured in Aug '12



# 94% Corporate Data Growth Y/Y

<b>Area</b>	<b>Growth Rate</b>
Database systems	97%
Overall corporate data	94%
Data of average organization	50%

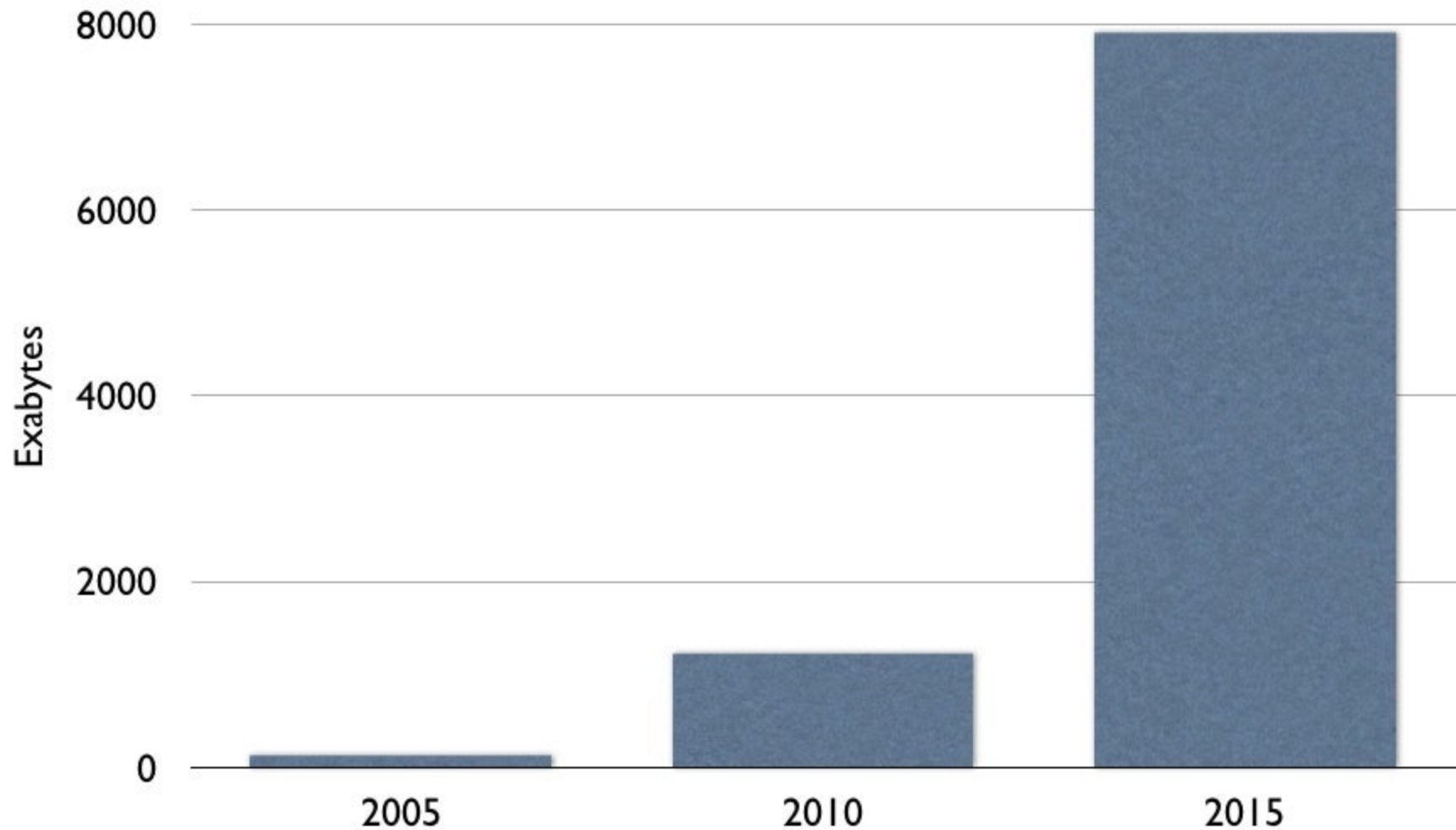
Source: Forrester



# Big Data: By The Numbers

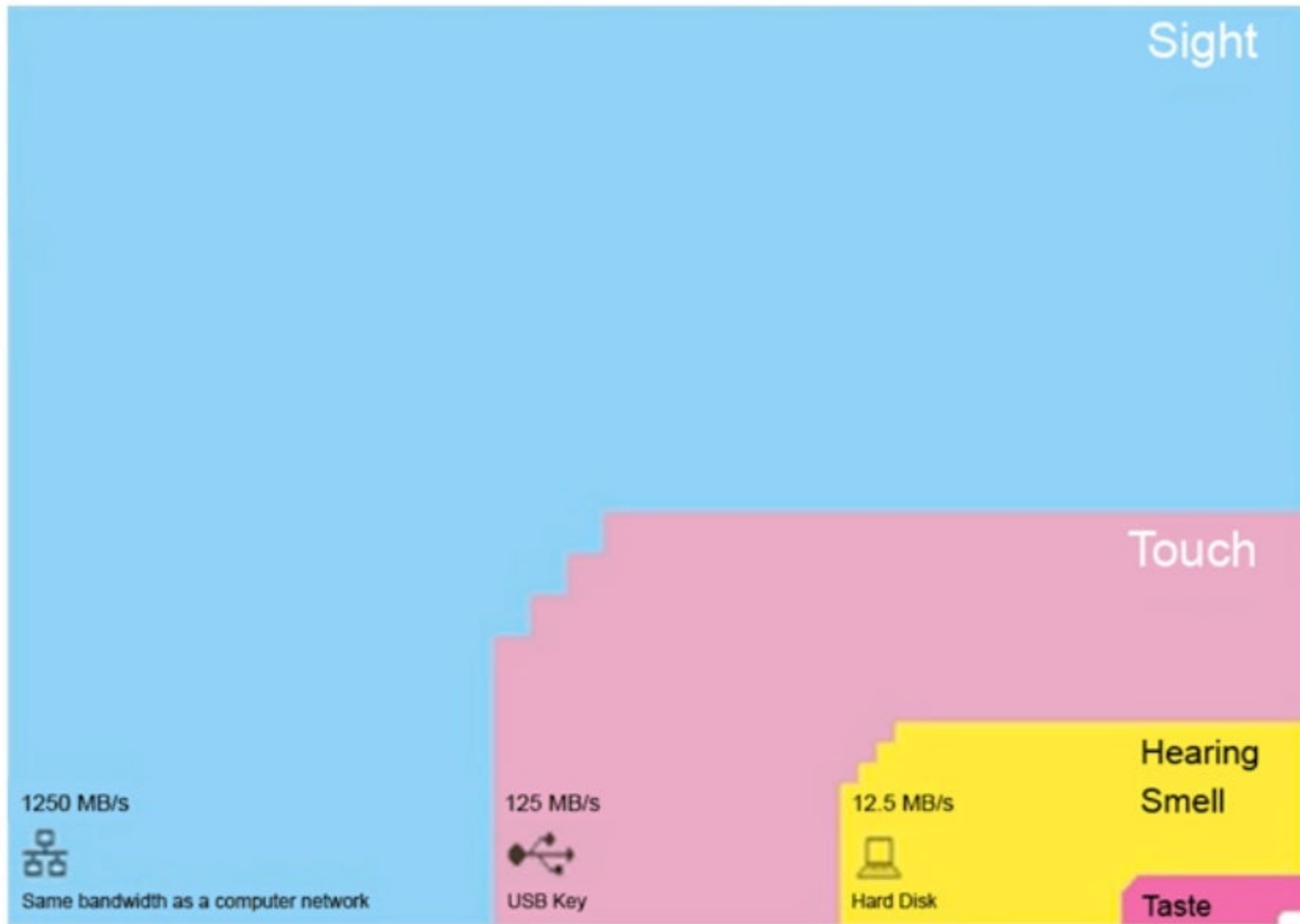
- Walmart handles 1M transactions per hour
- Google processes 24PB of data per day
- AT&T transfers 30PB of data per day
- 90 trillion emails are sent per year
- World of Warcraft uses 1.3PB of storage

# Worldwide Data Growth at 7.9EB / Yr in '15



Source: IDC, EMC. 1EB = 1 Billion GB.

# Bandwidth of Our Senses



# DATA SENSED PER YEAR

0.04 EB

7,910 EB



200,000X





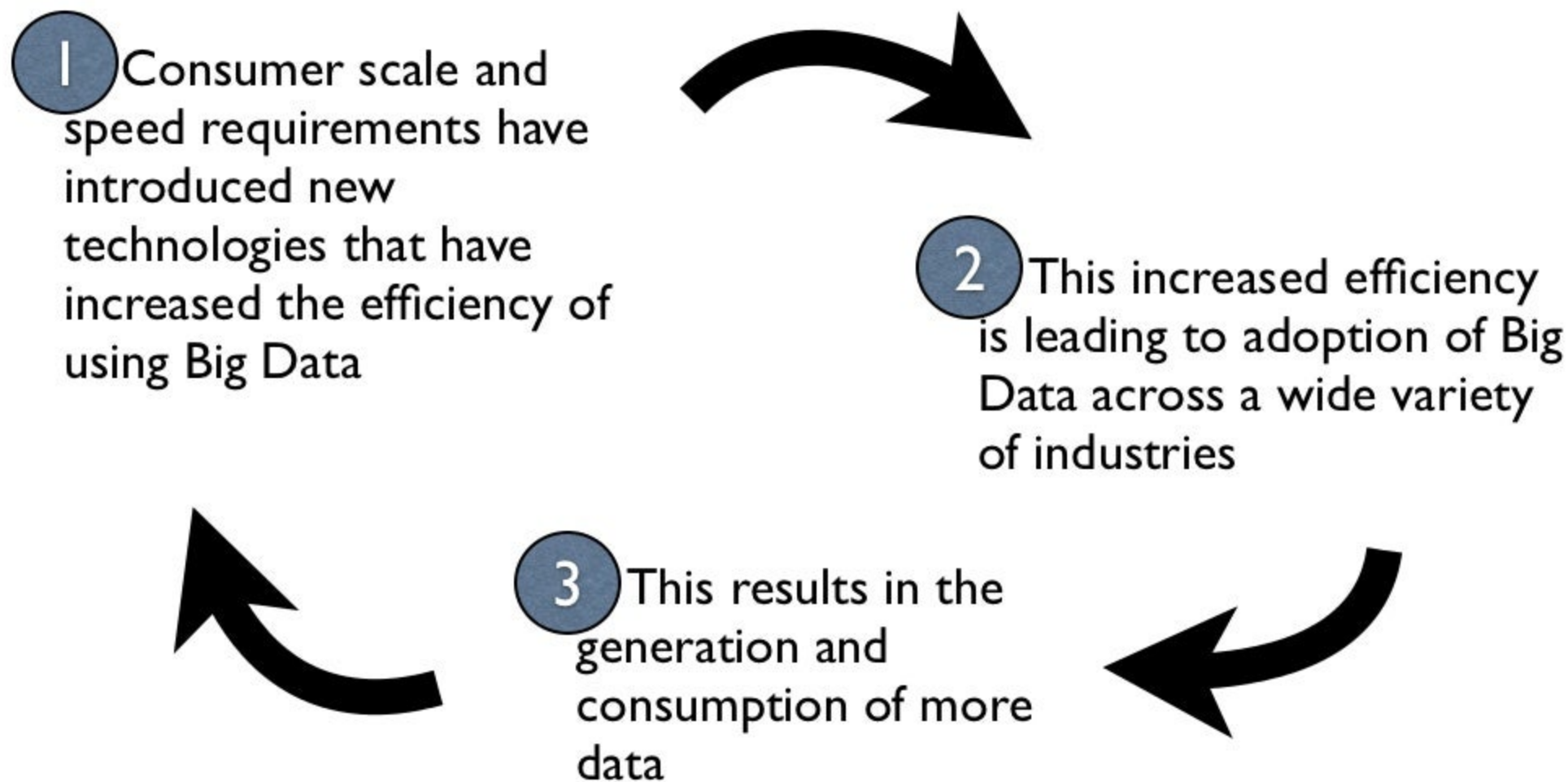
# THE BIG DATA CYCLE

# As Efficiency Increases, So Does Consumption



Source: Cloudyn

# The Big Data Cycle



# 6 Insights From Facebook's Former Head Of Big Data

**Analytics on 900M users**  
**25PB of compressed data - 125PB uncompressed**

- 1) “What data to store” => “What can we do with more data”
- 2) Simplify data analytics for end users
- 3) More users means analytics systems have to be more robust
- 4) Social networking works for Big Data
- 5) No single infrastructure can solve all Big Data problems
- 6) Building software is hard; running a service is even harder

# BIG DATA TRANSFORMING BUSINESS



# Transformation of Retail

THEN...

Sales



NOW...

Data driven pricing and recommendations



# Transformation of Online Marketing

THEN...

Leads

Company	First	Last	Oppty	Created
Acme	Fred	Langan	\$250K	6/08/12
BigCo	Tom	Jones	\$100K	6/17/12
DealCo	Jan	Sedor	\$50K	7/01/12
Stor Works	Liza	Grear	\$750K	7/14/12
RF Group	Carl	Tomer	\$47K	7/18/12

NOW...

Marketing and Sales  
Recommendations



# Transformation of IT

THEN...

Log files

```
[Sun Dec 21 09:17:09 2008] [error]
[Sun Dec 21 10:04:53 2008] [error]
[Sun Dec 21 10:45:50 2008] [error]
[Sun Dec 21 11:14:09 2008] [error]
[Sun Dec 21 12:26:04 2008] [error]
[Sun Dec 21 13:41:05 2008] [error]
[Sun Dec 21 14:37:16 2008] [error]
[Sun Dec 21 15:19:39 2008] [error]
[Sun Dec 21 15:26:05 2008] [error]
[Sun Dec 21 15:51:07 2008] [error]
```

NOW...

Operational intelligence



# Transformation of Customer Service

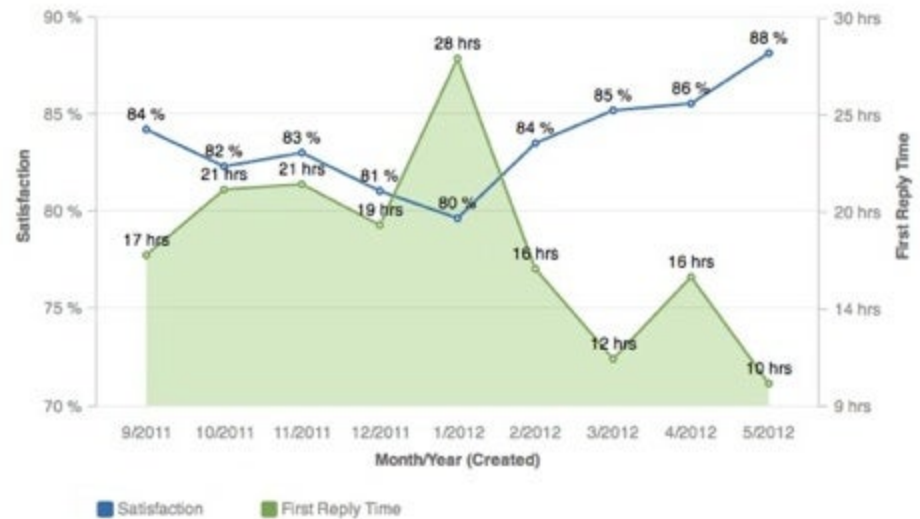
THEN...

Unhappy  
customers



NOW...

Customer insight



# Transformation of Billing

THEN...

Manual coding



NOW...

Intelligent coding





# Transformation Of Fraud Management

THEN...

Credit databases



NOW...

Social profiles



# Transformation Of Operations Management

THEN...

Operations?



NOW...

Automated  
operations



# Transformation of Law Enforcement

THEN...

Gut instinct



NOW...

Crime hotspot  
prediction



# Transformation of Medical Research

## THEN...

Keyword  
searches

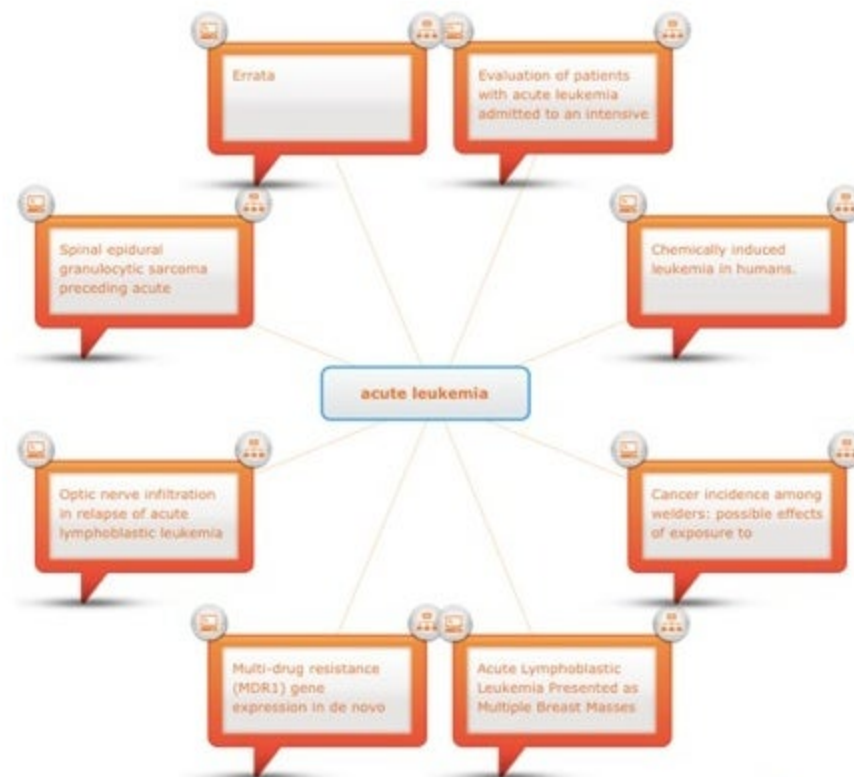
Results: 1 to 20 of 245953

<< First < Prev Page 1 of 12298 Next > Last >>

- ☐ [Abrogation of Src Homology Region 2 Domain-Containing Phosphatase 1 in Tumor-Specific T Cells Improves Efficacy of Adoptive Immunotherapy by Enhancing the Effector Function and Accumulation of Short-Lived Effector T Cells In Vivo.](#)  
Stromnes IM, Fowler C, Casamina CC, Georgopoulos CM, McAfee MS, Schmitt TM, Tan X, Kim TD, Choi I, Blattman JN, Greenberg PD.  
J Immunol. 2012 Jul 13. [Epub ahead of print]  
PMID: 22798667 [PubMed - as supplied by publisher]  
[Related citations](#)
- ☐ [MLL5 Maintains Genomic Integrity by Regulating the Stability of the Chromosomal Passenger Complex via a Functional Interaction with Borealin.](#)  
Liu J, Cheng F, Deng LW.  
J Cell Sci. 2012 Jul 13. [Epub ahead of print]  
PMID: 22797924 [PubMed - as supplied by publisher]  
[Related citations](#)
- ☐ [The Btk pathway functions as a tumor suppressor in chronic myeloid leukemia stem cells.](#)  
Zhang H, Peng C, Hu Y, Li H, Sheng Z, Chen Y, Sullivan C, Cerny J, Hutchinson L, Higgins A, Miron P, Zhang X, Brehm MA, Li D, Green MR, Li S.  
Nat Genet. 2012 Jul 15. doi: 10.1038/ng.2350. [Epub ahead of print]  
PMID: 22797726 [PubMed - as supplied by publisher]  
[Related citations](#)

## NOW...

Relevance

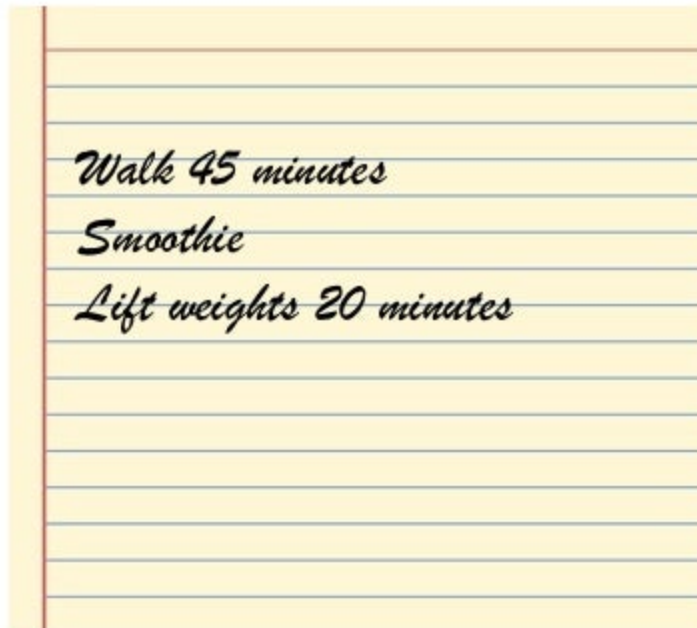




# Transformation of Fitness

## THEN...

Manual  
tracking



## NOW...

Goals +  
measurement





# TOP 8 LAWS OF BIG DATA

**The faster you analyze your data,  
the greater its predictive value.**

Companies are moving away from batch processing to real-time to gain competitive advantage.

**Maintain one copy of your data, not dozens.** The more you copy and move your data, the less reliable it becomes (example: banking crisis).

**Use more diverse data, not just more data.** More diverse data leads to greater insights. Combining multiple data sources can lead to the most interesting insights of all.

**Data has value far beyond what you originally anticipate.** Don't throw it away.

**Plan for exponential growth.** The number of photos, emails, and IMs while large, is limited by the number of people. Networked “sensor” data from mobile phones, GPS, and other devices is much larger.



**Solve a real pain point.** Don't think of Big Data as a stand-alone new, shiny, technology. Think about your core business problems and how to solve them by analyzing Big Data.

**Put data and humans together to get the most insight.** More data alone isn't sufficient. Look for ways to broaden the use of data across your organization.

**Big Data is transforming business the same way IT did.** Those that fail to leverage the numerous internal and external data sources available will be leapfrogged by new entrants.

## D2: DATA + DESIGN

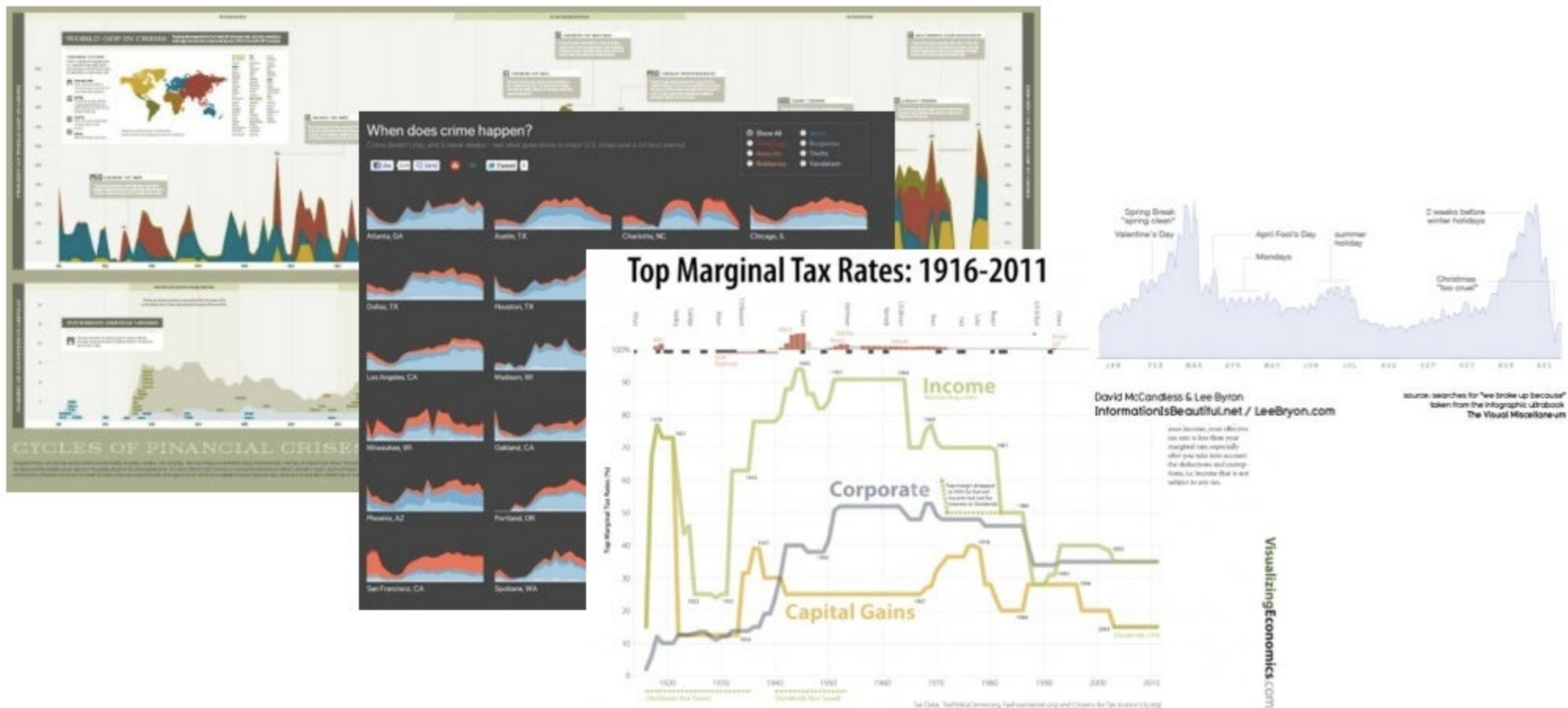
# Data Informs Design



**“The racing technology on the yachts competing for the America’s Cup will be the most advanced ever”**

*- The Wall Street Journal MarketWatch*

# Design Informs Data



**“Visualization is a form of knowledge compression”**  
- David McCandless



# Big Data Versus Romance

## Peak Break-Up Times

According to Facebook status updates



David McCandless & Lee Byron  
InformationIsBeautiful.net / LeeByron.com

source: searches for "we broke up because"  
taken from the infographic ultrabook  
The Visual Miscellane um

# SUMMARY

# Big Data White Space

- 1) Visualization - cloud, mobile, collaboration
- 2) Big Data Apps - verticals
- 3) Trend analysis across multiple data sources
- 4) Consumer behavior
- 5) Public data for scoring
- 6) New information / data service businesses

# Summary

- 1) **Consumer company** speed and **scale** requirements driving efficiencies in Big Data storage and analytics
- 2) Quantity of **machine data** vastly increasing (examples: networked sensor data from mobile phones and GPS devices)
- 3) Move away from **batch to real-time**
- 4) **Cloud** services opening Big Data to all
- 5) New and broader number of **data** sources being **meshed** together
- 6) **Big Data Apps** (BDAs) means using Big Data is faster and easier

David Feinleib

[dave@thebigdatagroup.com](mailto:dave@thebigdatagroup.com)