WHAT IS BIG DATA?

David Bechtold

Agenda

1. Introduction
2. What is Big Data?
3. Big Data – a perspective
4. Characteristic of Big Data – Three Vs
5. A Fourth V..?
6. Examples…
7. How did we get here?... A historical look back
8. Interesting FACTs about Big Data… what, where, how, why?
9. Big Data and the Analysis of that data (Analytics) in Industry
10. Big Data risks… It is not necessarily all good news…
11. Big Data and YOU… It is not hard to create, but it is hard to extinguish
12. Big Data benefits – spawning new careers
13. Analytics, Cognitive Computing… Applications of Big Data
14. Video examples
15. Do you feel the effects of Information Overload…?

David Bechtold

Tenure with IBM:	**1989-1997;2008-Present**
Tenure with StorageTek:	**1998-2000**
Tenure with Veritas Software:	**2000-2006**
Tenure with Hewlett-Packard:	**2007-2008**

David holds a BSEE degree from RIT (Rochester NY) and an MSCS degree from Marist College (Poughkeepsie NY).

Background in Electrical Engineering and Software Development then to Client Technical Architect (Solution Architecture). Been in the Enterprise Data Management Field for the past 20 years.

Currently an IBM Client Technical Architect covering New England based and large International Accounts. Experienced with Cloud, Big Data, Flash, Storage Virtualization and Performance Optimization, Elastic Storage as well as High Availability and Disaster Recovery Solutions.

What is BIG DATA?

- “Big Data is similar to small data, but bigger in size.”
- “Data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges.”
- “An all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using on-hand data management tools or traditional data processing applications.”
- “Datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.”
- “The belief that the more data you have the more insights and answers will rise automatically from the pool of ones and zeros.”
- “A new attitude by businesses, non-profits, government agencies, and individuals that combining data from multiple sources could lead to better decisions.”

What is BIG DATA?

- Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone.

- This data comes from everywhere:
  - sensors used to gather climate information,
  - posts to social media sites,
  - digital pictures and videos,
  - purchase transaction records,
  - cell phone GPS signals to name a few.

This data is "**big data**."

Let’s look at

Big Data

in a different way… a perspective…

Byte : one grain of rice

Byte

Byte : one grain of rice

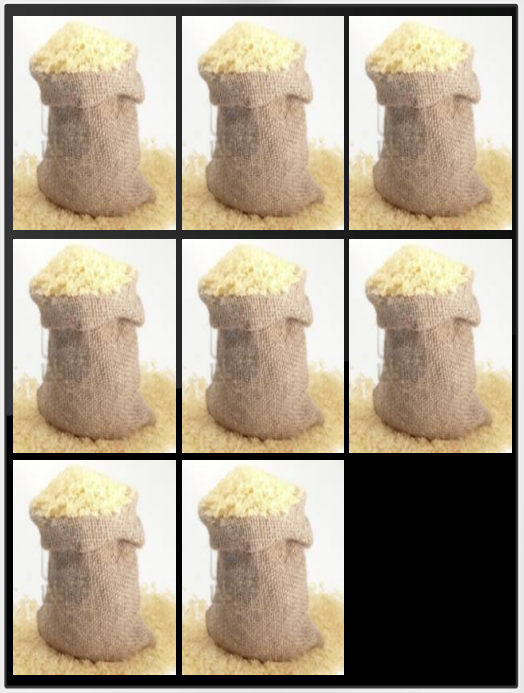
Kilobyte : cup of rice

Kilobyte

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice



Megabyte

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 Semi trucks

Gigabyte

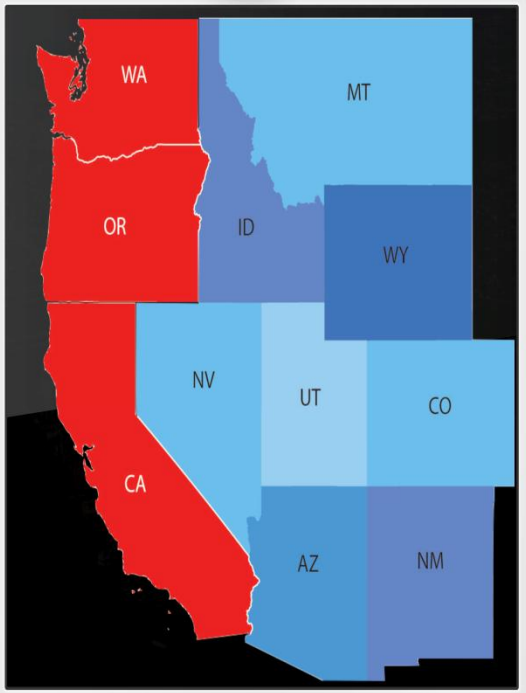
Byte : one grain of rice
Kilobyte : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte : 2 Container Ships

Terabyte

Byte : one grain of rice
Kilobyte : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte : 2 Container Ships
Petabyte : Blankets Manhattan

Petabyte

Byte : one grain of rice
Kilobyte : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte : 2 Container Ships
Petabyte : Blankets Manhattan
Exabyte : Blankets west coast states



Exabyte

Byte : one grain of rice
Kilobyte : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte : 2 Container Ships
Petabyte : Blankets Manhattan
Exabyte : Blankets west coast states
Zettabyte : Fills the Pacific Ocean

Zettabyte

Byte : one grain of rice
Kilobyte : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte : 2 Container Ships
Petabyte : Blankets Manhattan
Exabyte : Blankets west coast states
Zettabyte : Fills the Pacific Ocean
Yottabyte : A EARTH SIZE RICE BALL!

Yottabyte

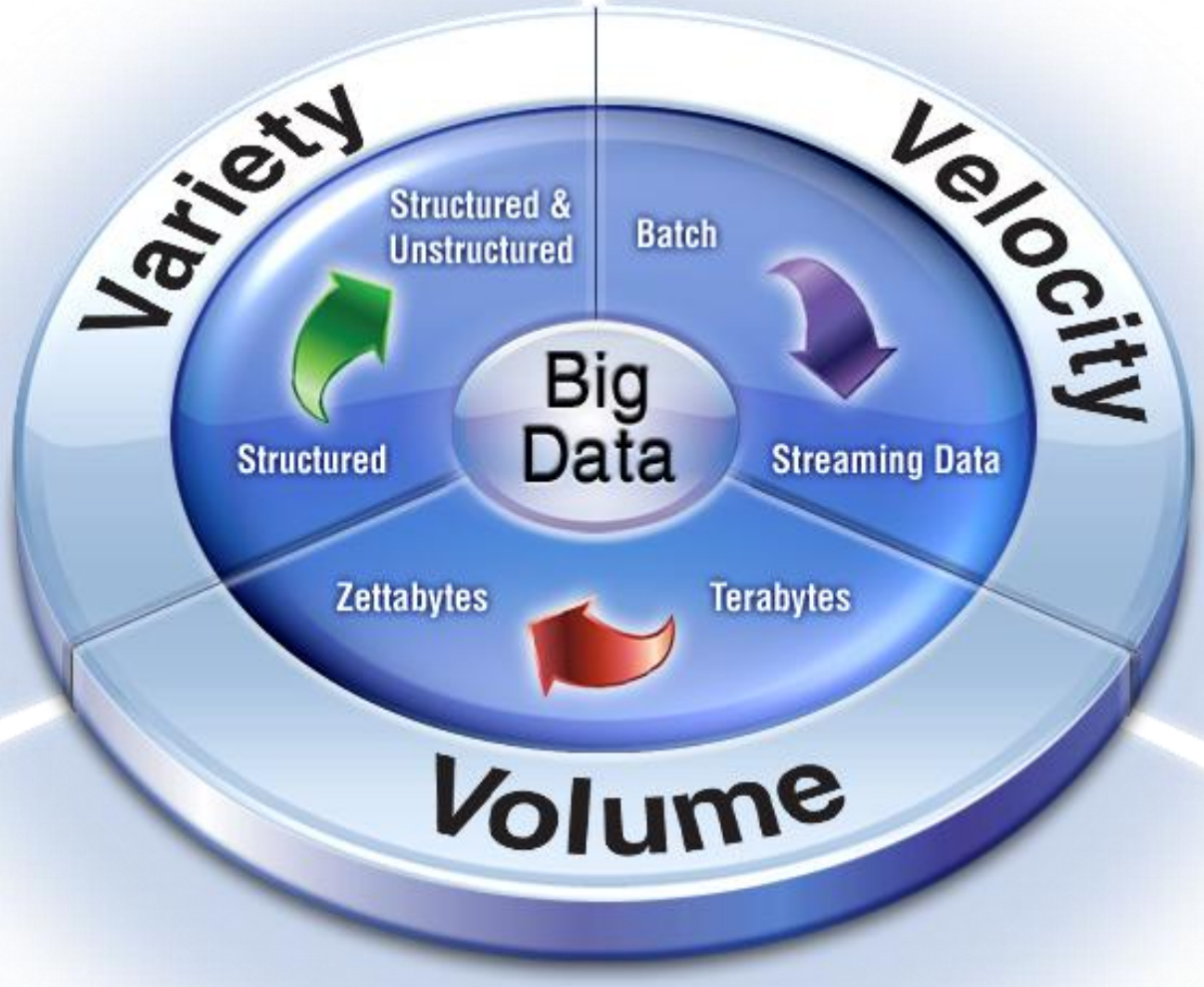
Byte	: one grain of rice		Hobbyist
Kilobyte	: cup of rice		
Megabyte	: 8 bags of rice		Desktop
Gigabyte	: 3 Semi trucks		
Terabyte	: 2 Container Ships		Internet
Petabyte	: Blankets Manhattan		
Exabyte	: Blankets west coast states		Big Data
Zettabyte	: Fills the Pacific Ocean		
Yottabyte	: A EARTH SIZE RICE BALL!		

Byte	: one grain of rice
Kilobyte	: cup of rice
Megabyte	: 8 bags of rice
Gigabyte	: 3 Semi trucks
Terabyte	: 2 Container Ships
Petabyte	: Blankets Manhattan
Exabyte	: Blankets west coast states
Zettabyte	: Fills the Pacific Ocean
Yottabyte	: A EARTH SIZE RICE BALL!

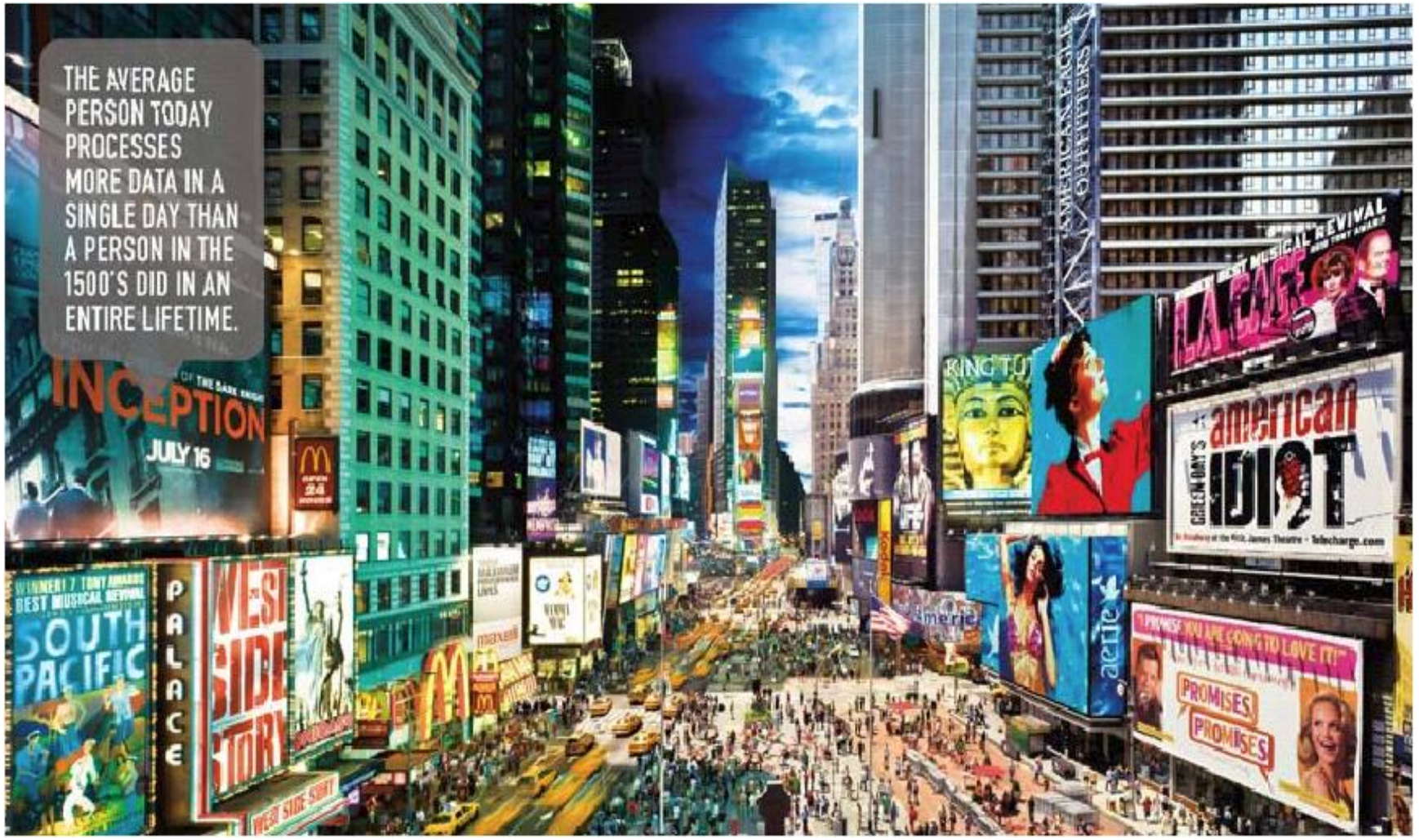
Byte	: one grain of rice		Hobbyist
Kilobyte	: cup of rice		
Megabyte	: 8 bags of rice		Desktop
Gigabyte	: 3 Semi trucks		
Terabyte	: 2 Container Ships		Internet
Petabyte	: Blankets Manhattan		
Exabyte	: Blankets west coast states		Big Data
Zettabyte	: Fills the Pacific Ocean		
Yottabyte	: A EARTH SIZE RICE BALL!	*The Future?*	

BIG DATA spans three dimensions: Volume, Velocity and Variety

- **Volume(size):** Enterprises are awash with ever-growing data of all types, easily amassing terabytes—even petabytes—of information.
  - Turn 12 terabytes of Tweets created each day into improved product sentiment analysis
  - Convert 350 billion annual meter readings to better predict power consumption
- **Velocity (speed):** Sometimes 2 minutes is too late. For time-sensitive processes such as catching fraud, big data must be used as it streams into your enterprise in order to maximize its value.
  - Scrutinize 5 million trade events created each day to identify potential fraud
  - Analyze 500 million daily call detail records in real-time to predict customer churn faster
- **Variety (sources):** Big data is any type of data - structured and unstructured data such as text, sensor data, audio, video, click streams, log files and more. New insights are found when analyzing these data types together.
  - Monitor 100's of live video feeds from surveillance cameras to target points of interest
  - Exploit the 80% data growth in images, video and documents to improve customer satisfaction


Variety
Velocity
Volume
Structured &
Unstructured
Batch
Structured
Big
Data
Streaming Data
Zettabytes
Terabytes

Data Volume…


THE AVERAGE PERSON TODAY PROCESSES MORE DATA IN A SINGLE DAY THAN A PERSON IN THE 1500'S DID IN AN ENTIRE LIFETIME.
INCEPTION
JULY 16
WINNER! 7 TONY AWARDS BEST MUSICAL REVIVAL
SOUTH PACIFIC
PALACE
WEST SIDE STORY
AMERICA EAGLE OUTFITTERS
LA CAGE
BEST MUSICAL REVIVAL
KING TUT
GREEN DAY'S
american IDIOT
aerie
PROMISES PROMISES

Data Volume…

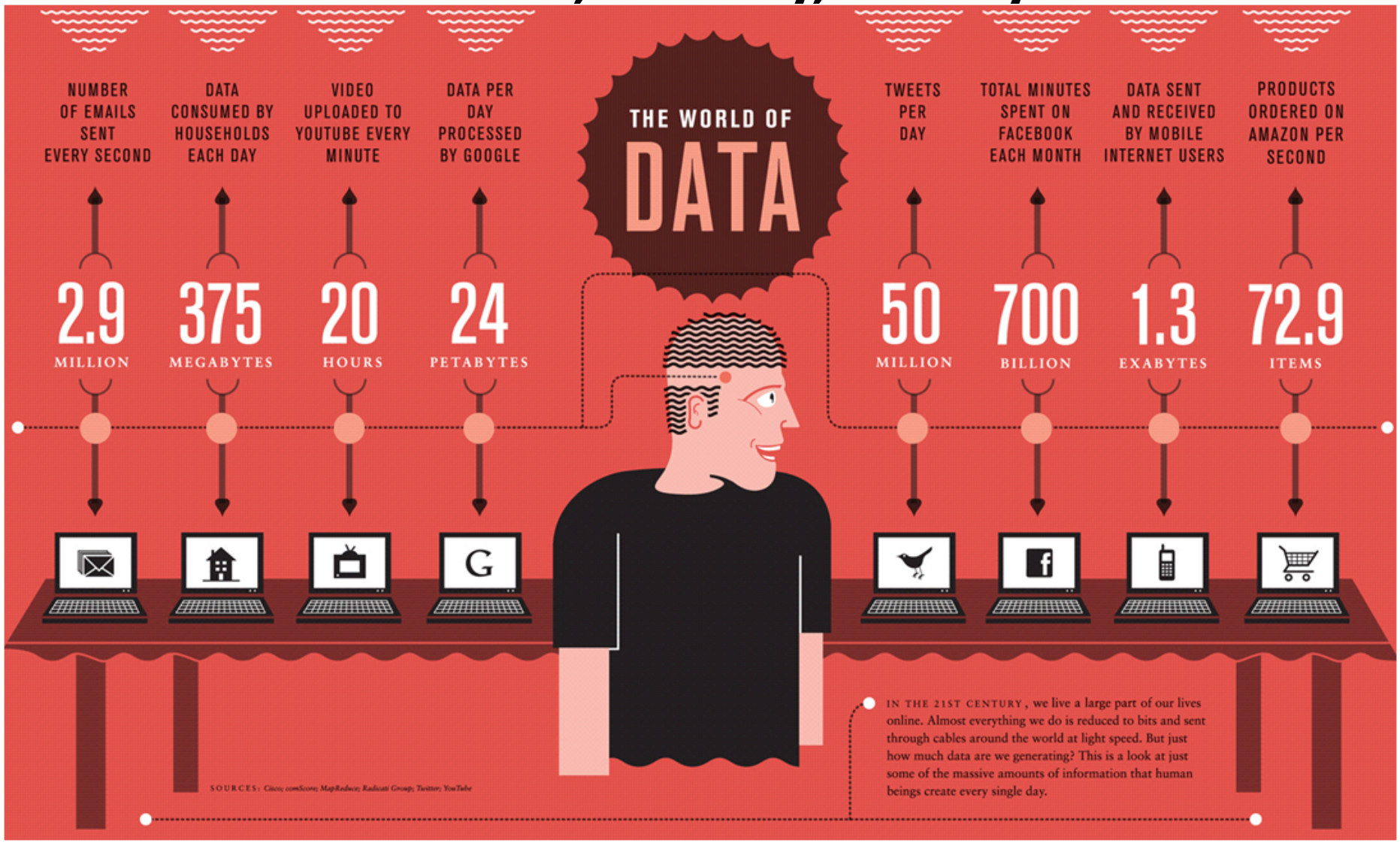

DURING THE FIRST DAY OF A BABY'S LIFE, THE AMOUNT OF DATA GENERATED BY HUMANITY IS EQUIVALENT TO 70 TIMES THE INFORMATION CONTAINED IN THE LIBRARY OF CONGRESS.

Data Velocity – 60 Seconds…

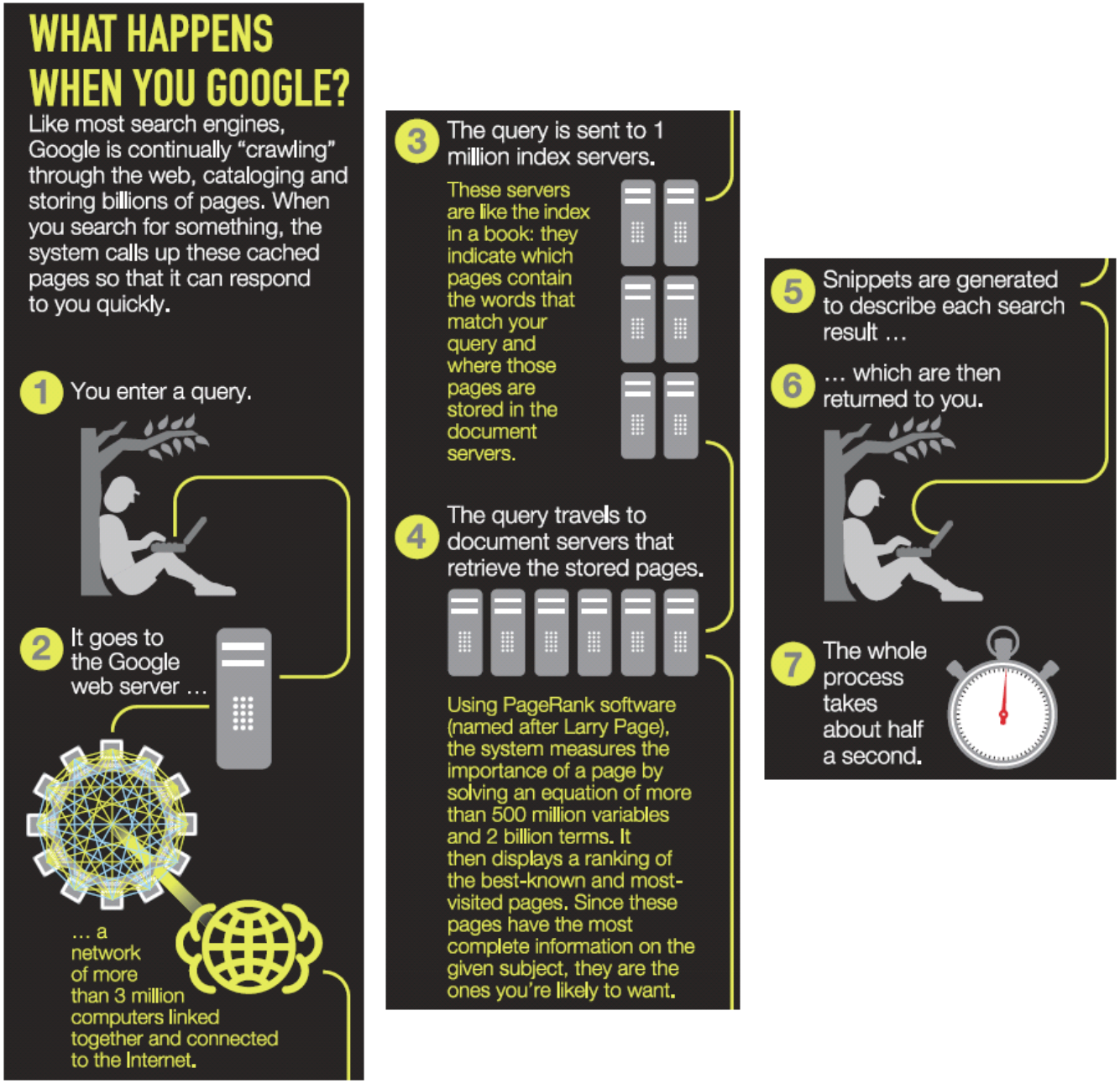
Data Velocity & Transparency…

Data Variety…

A look at Data (sec, min, day… month)
Volume, Velocity, Variety

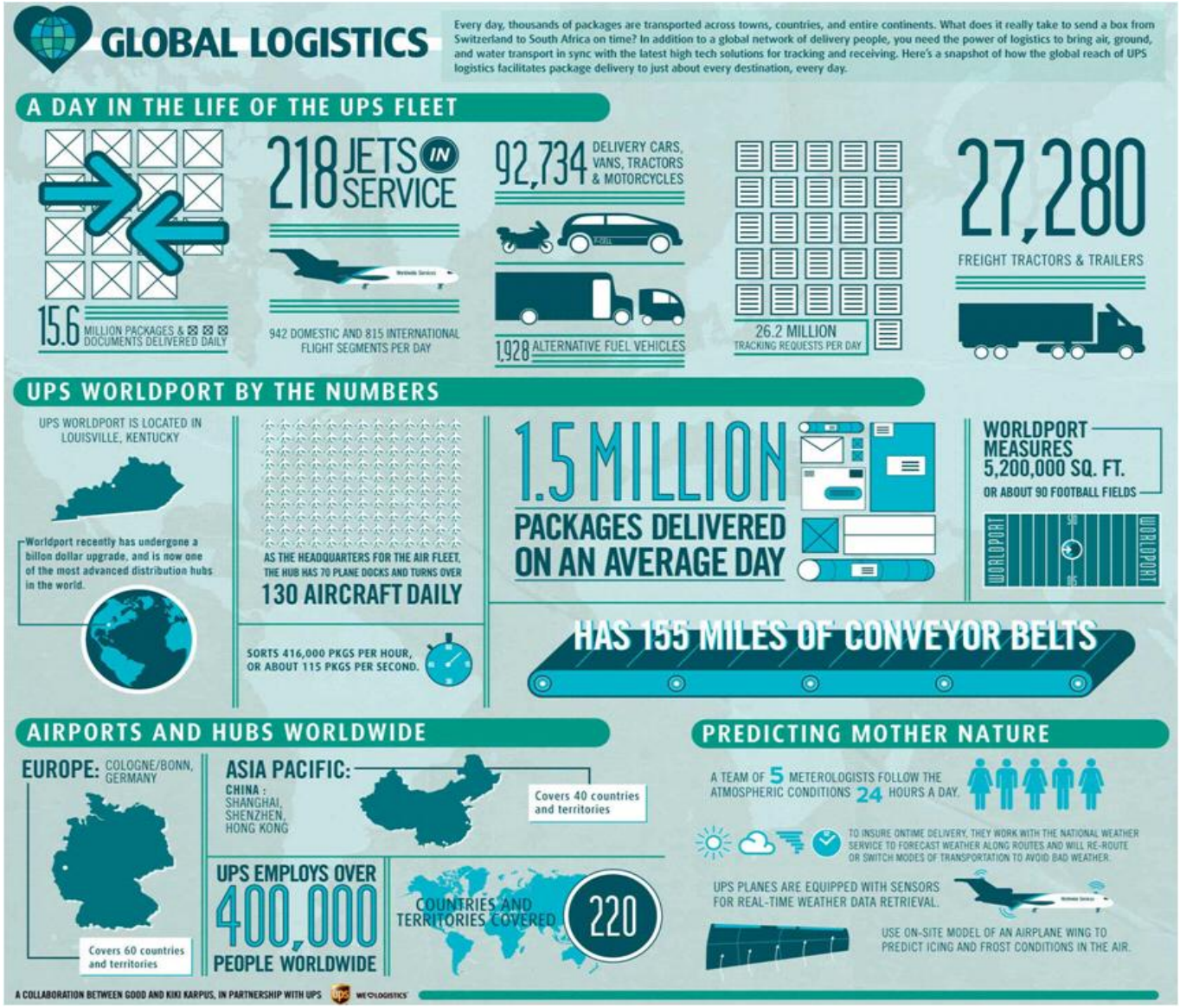


What happens when you search Google?

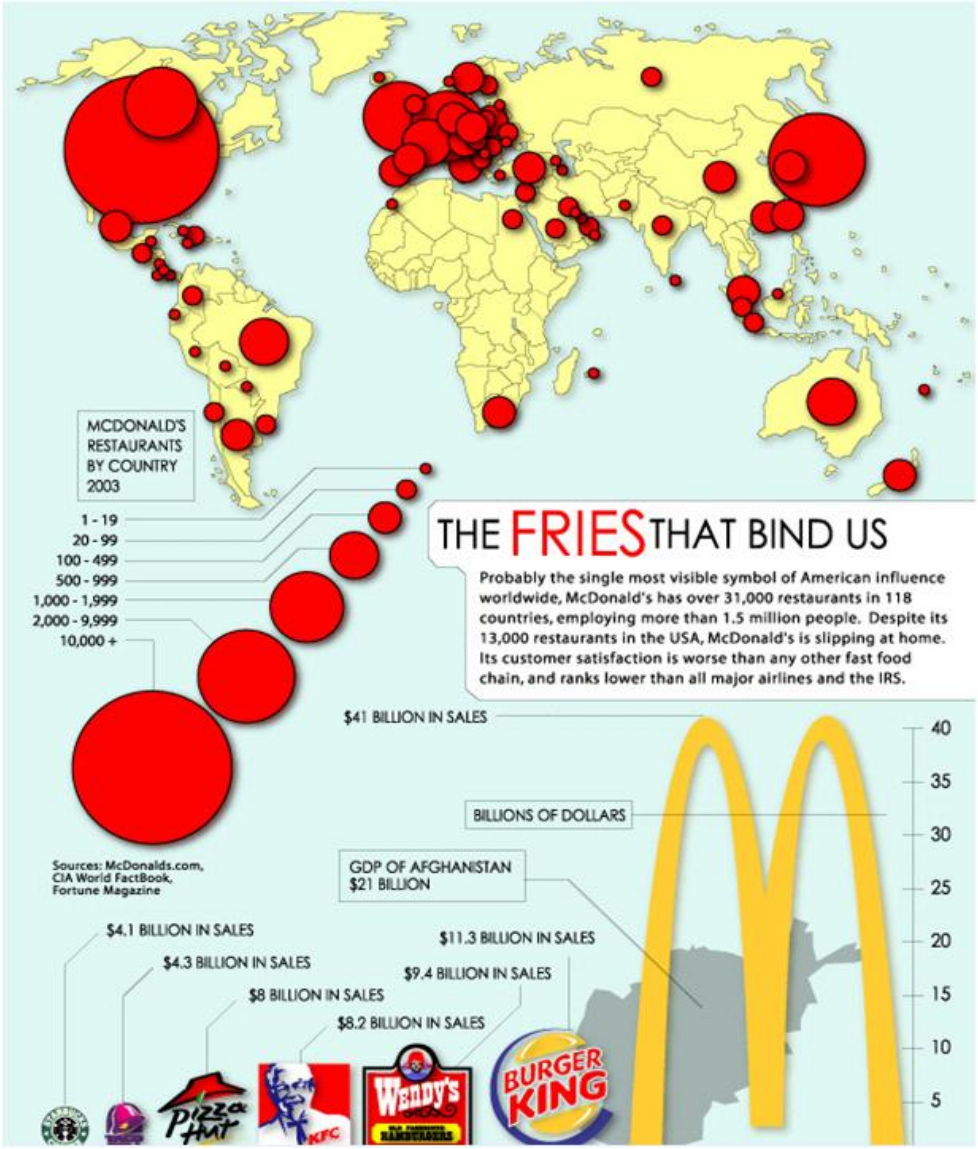


Your Turn: This is an example of Data V….?

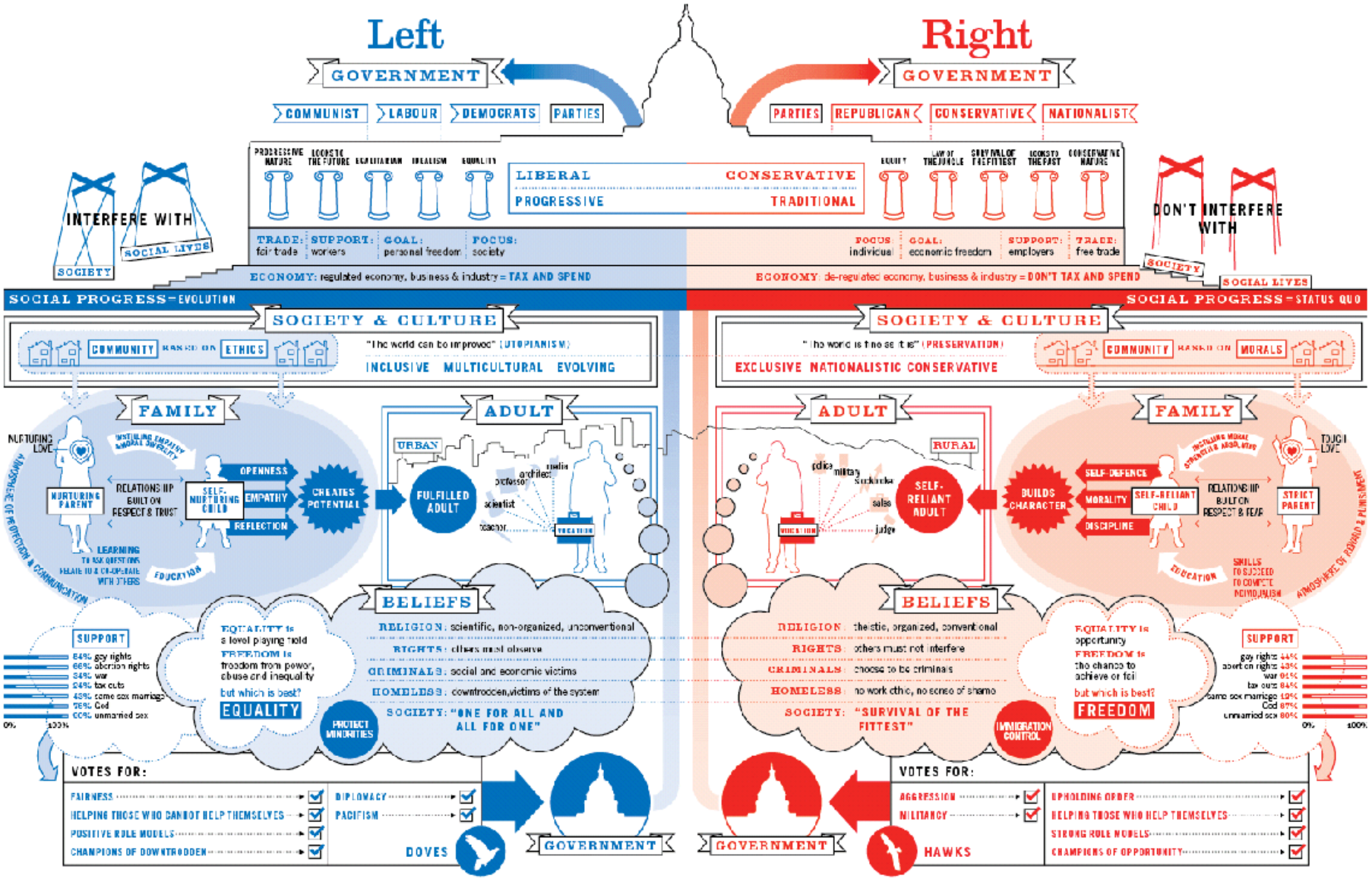
What to make of all that data…?



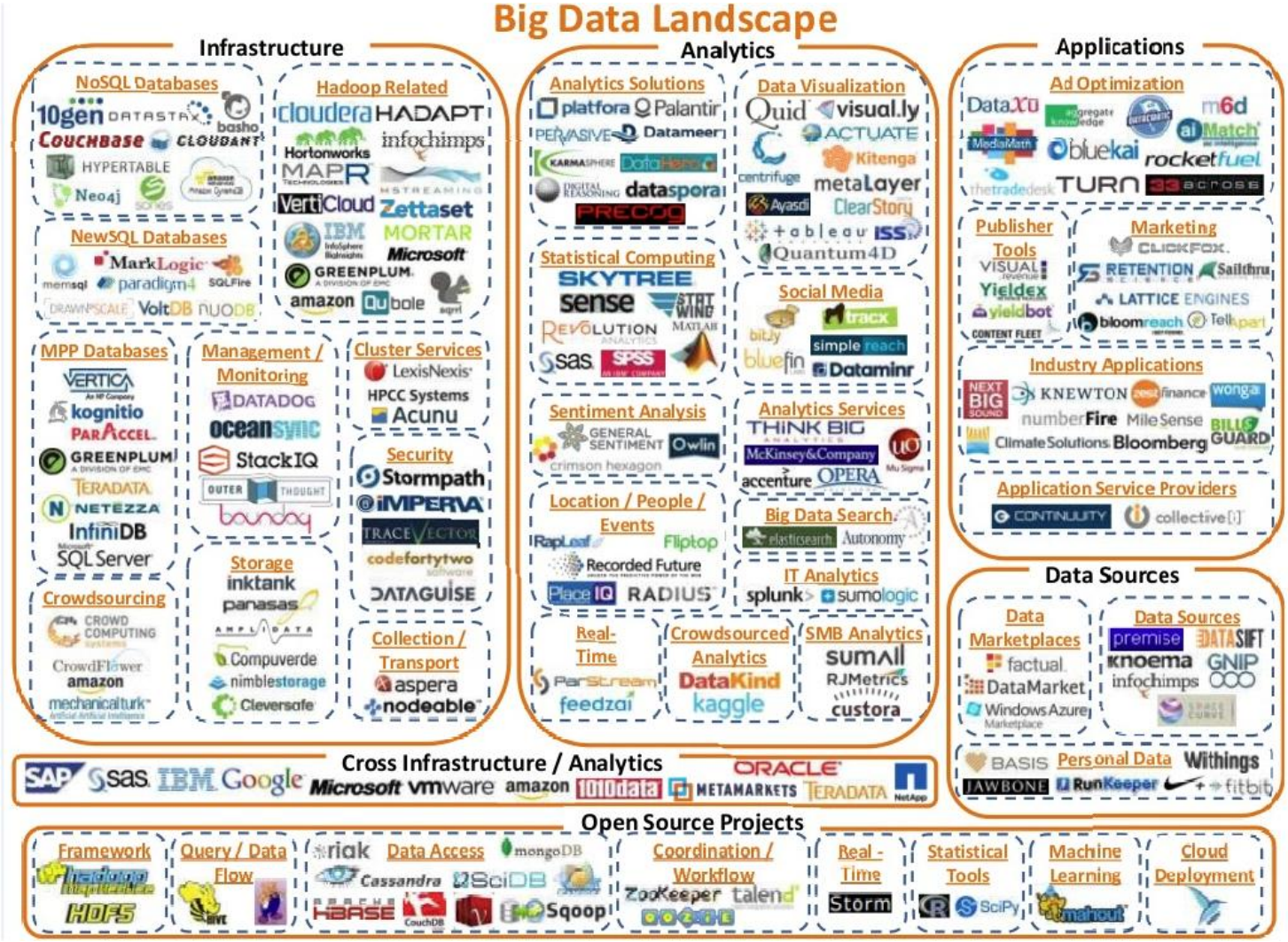
What to make of all that data…?



What to make of all that data…?



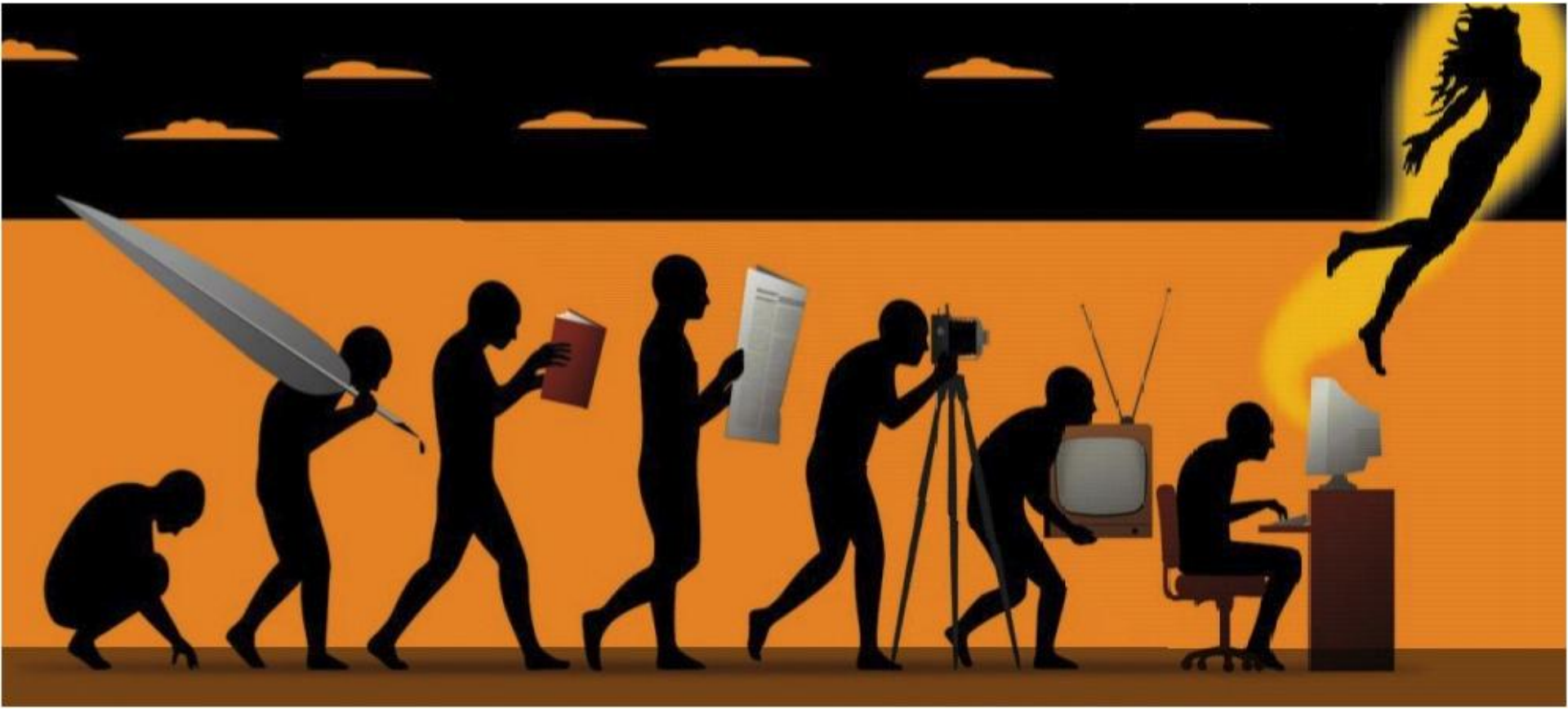
The BIG DATA landscape



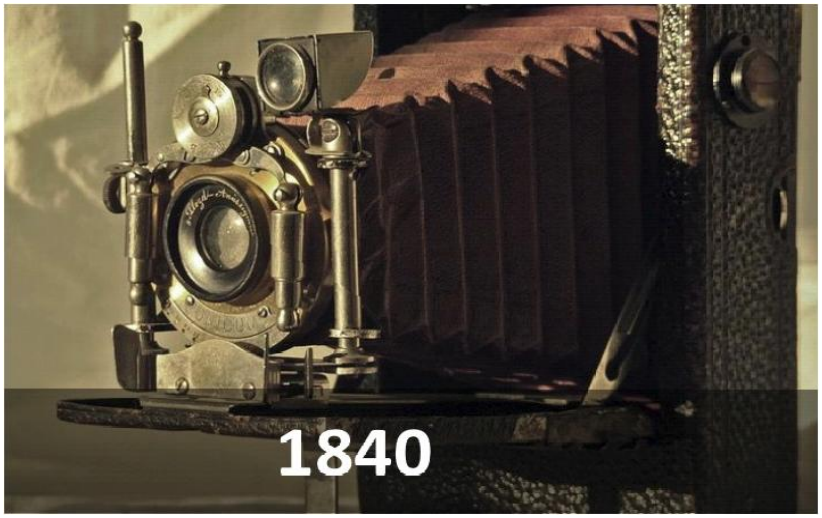
How did we get here…?

Volume
Velocity
Variety

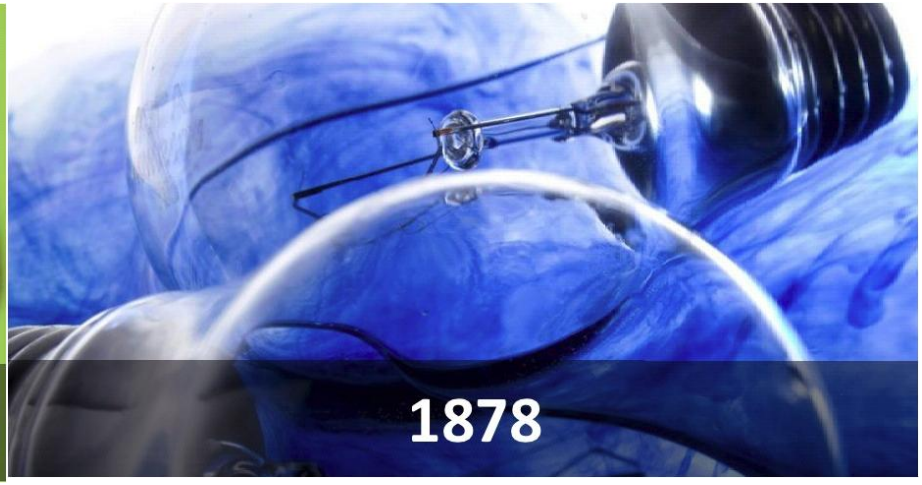
The rise of communication over time

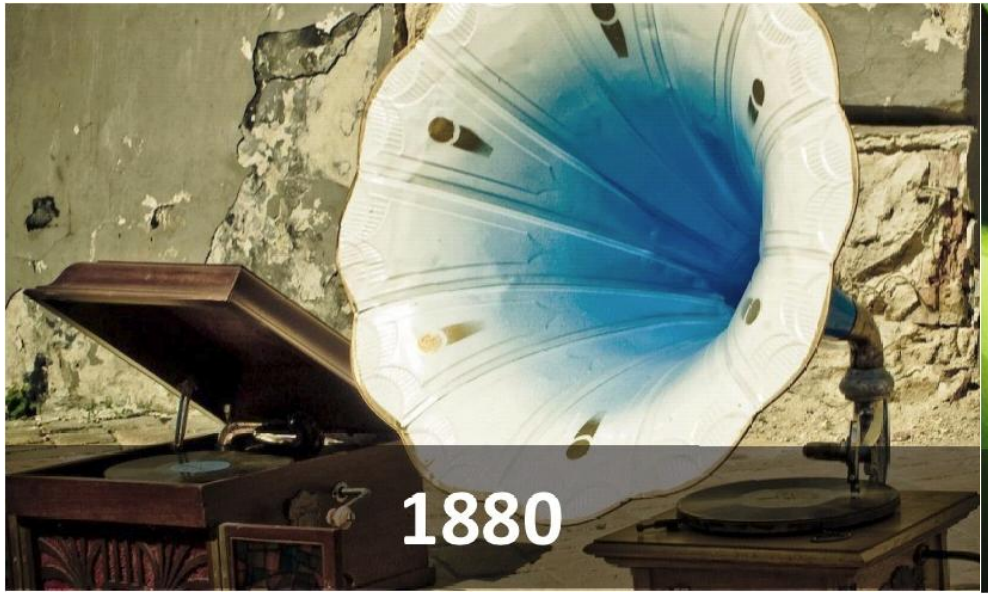


Data generation & communication over time

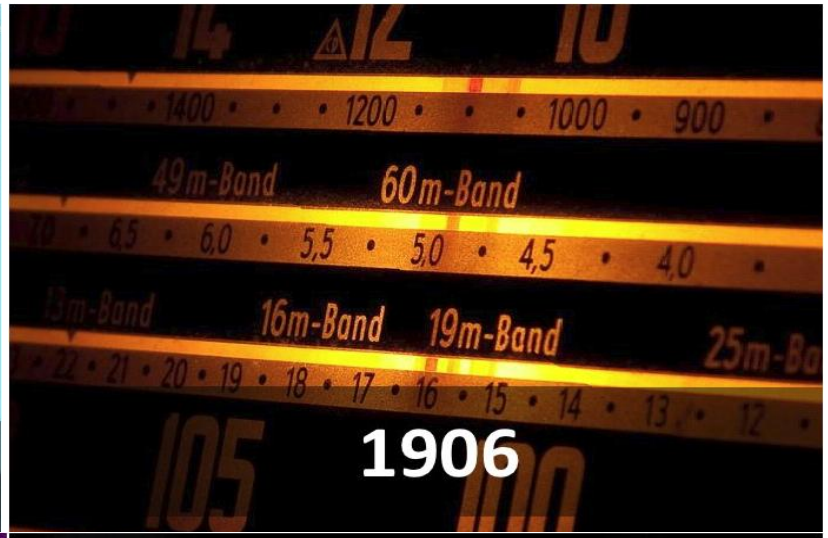


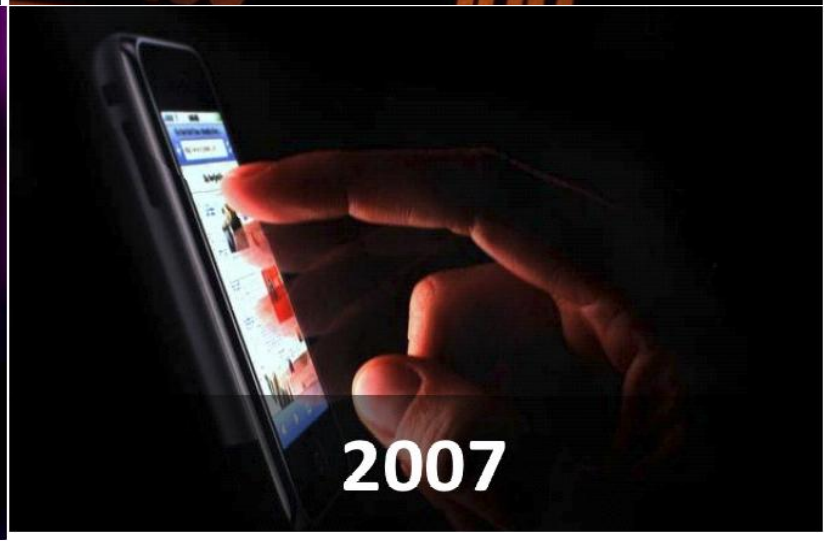
Data generation & communication over time



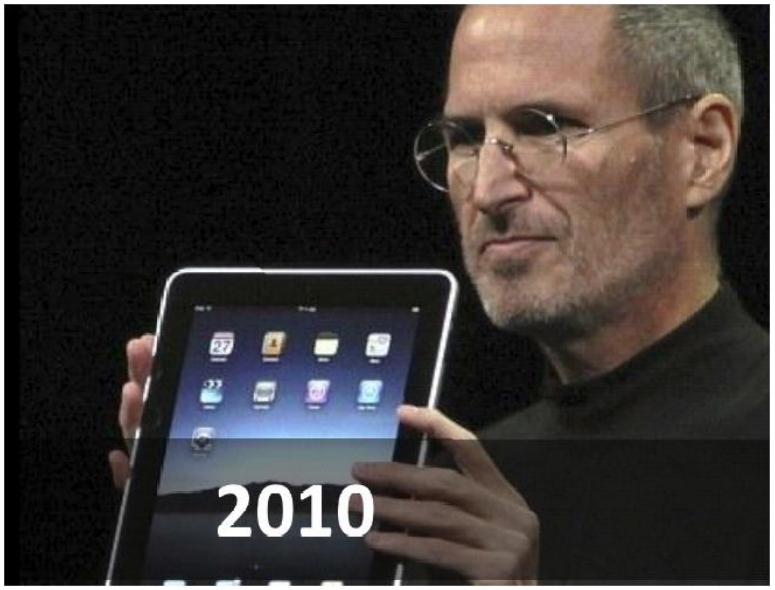


Data generation & communication over time





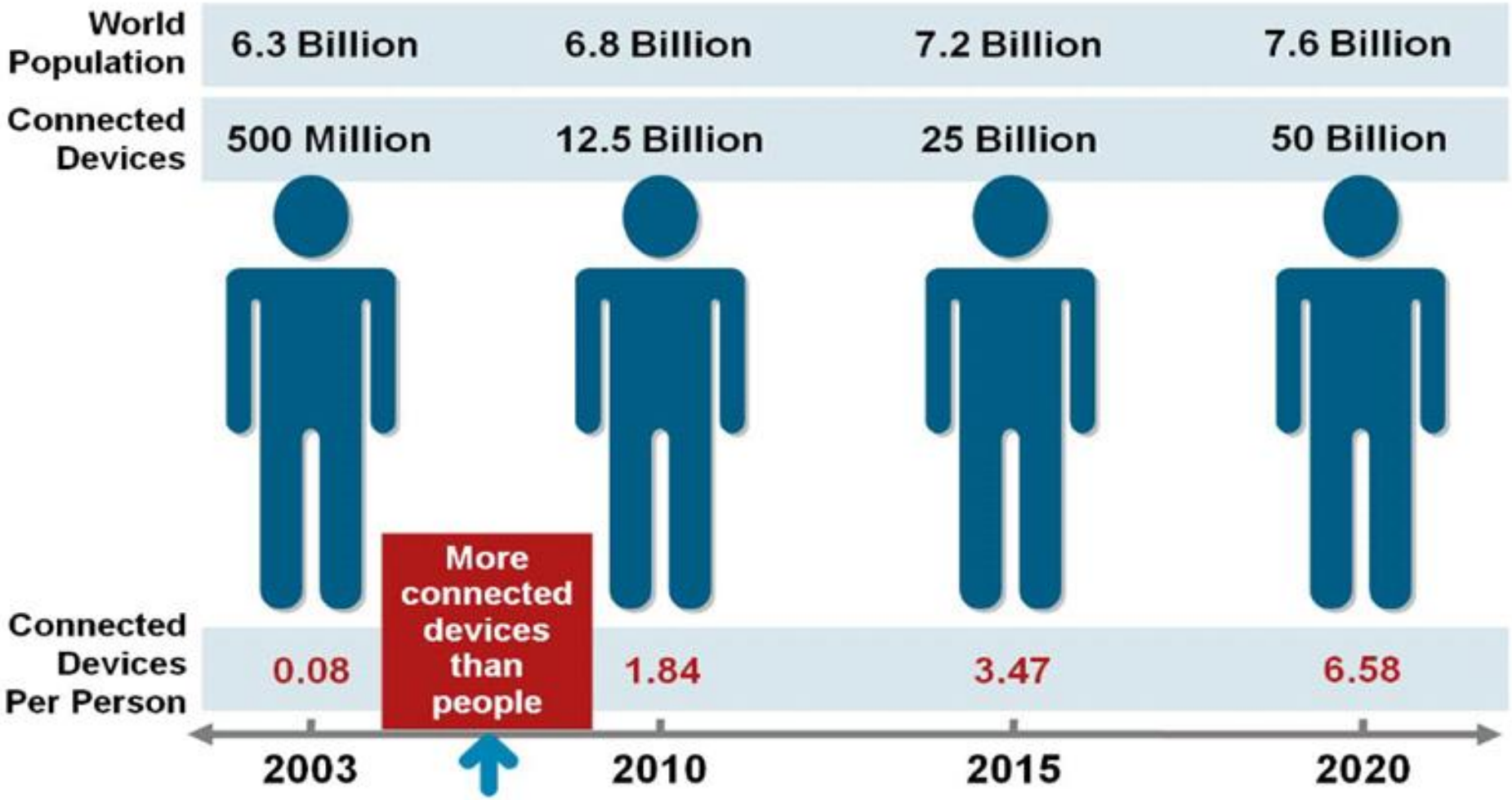
Data generation & communication over time



ENCYCLOPÆDIA Britannica®

About 16 oz on an iPad
About 128 lbs in hard copy text books

Driving Data Generation – Example:



15 Need-to-Know Facts

Fact 1

Every 2 days we create as much information as we did from the beginning of time until 2003

Fact 2

Over 90% of all the data in the world was created in the past 2 years.

Fact 3

The total amount of data being captured and stored by industry doubles every 1.2 years

Fact 4

Every minute we send 204 million emails, generate 1,8 million Facebook likes, send 278 thousand Tweets, and up-load 200,000 photos to Facebook

Fact 5

Google alone processes on average over 40 thousand search queries per second, making it over 3.5 billion in a single day.

Fact 6

Around 100 hours of video are uploaded to YouTube every minute and it would take you around 15 years to watch every video uploaded by users in one day.

Fact 7

If you burned all of the data created in just one day onto DVDs, you could stack them on top of each other and reach the moon – twice.

Fact 8

1.9 million IT jobs will be created in the US by 2015 to carry out big data projects. Each of those will be supported by 3 new jobs created outside of IT – meaning a total of 6 million new jobs thanks to big data.

Fact 9

1.570 new websites spring into existence every minute of every day.

Fact 10

This year, there will be over 1.2 billion smart phones in the world (which are stuffed full of sensors and data collection features), and the growth is predicted to continue.

Fact 11

12 million RFID tags – used to capture data and track movement of objects in the physical world – had been sold in by 2011. By 2021, it is estimated that number will have risen to 209 billion.

Fact 12

Big data has been used to predict crimes before they happen – a "predictive policing" trial in California was able to identify areas where crime will occur three times more accurately than existing methods of forecasting.

Fact 13

By better integrating big data analytics into healthcare, the industry could save $300bn a year – that's the equivalent of reducing the healthcare costs of every man, woman and child by $1,000 a year.

Fact 14

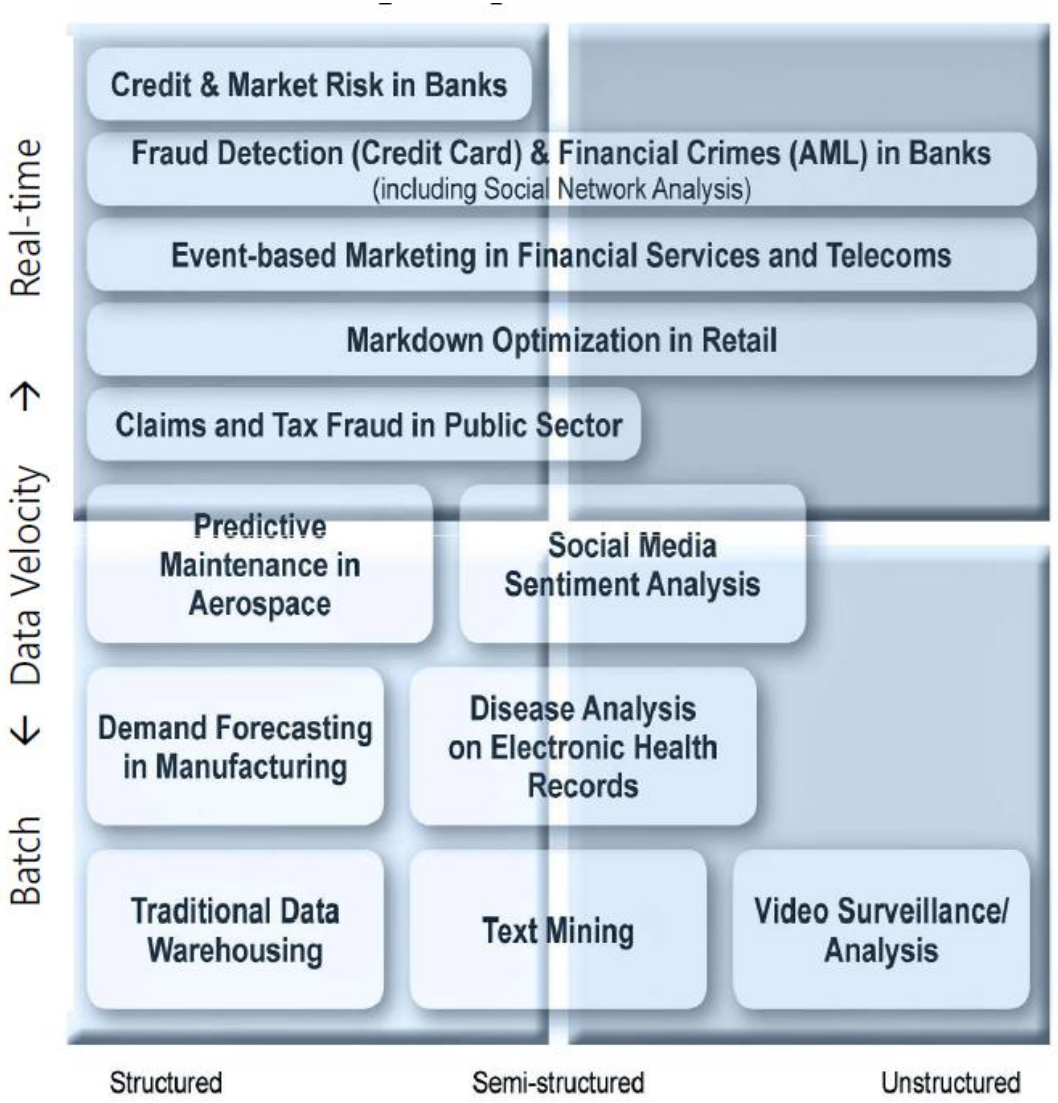
Retailers could increase their profit margins by more than 60% through the full exploitation of big data analytics.

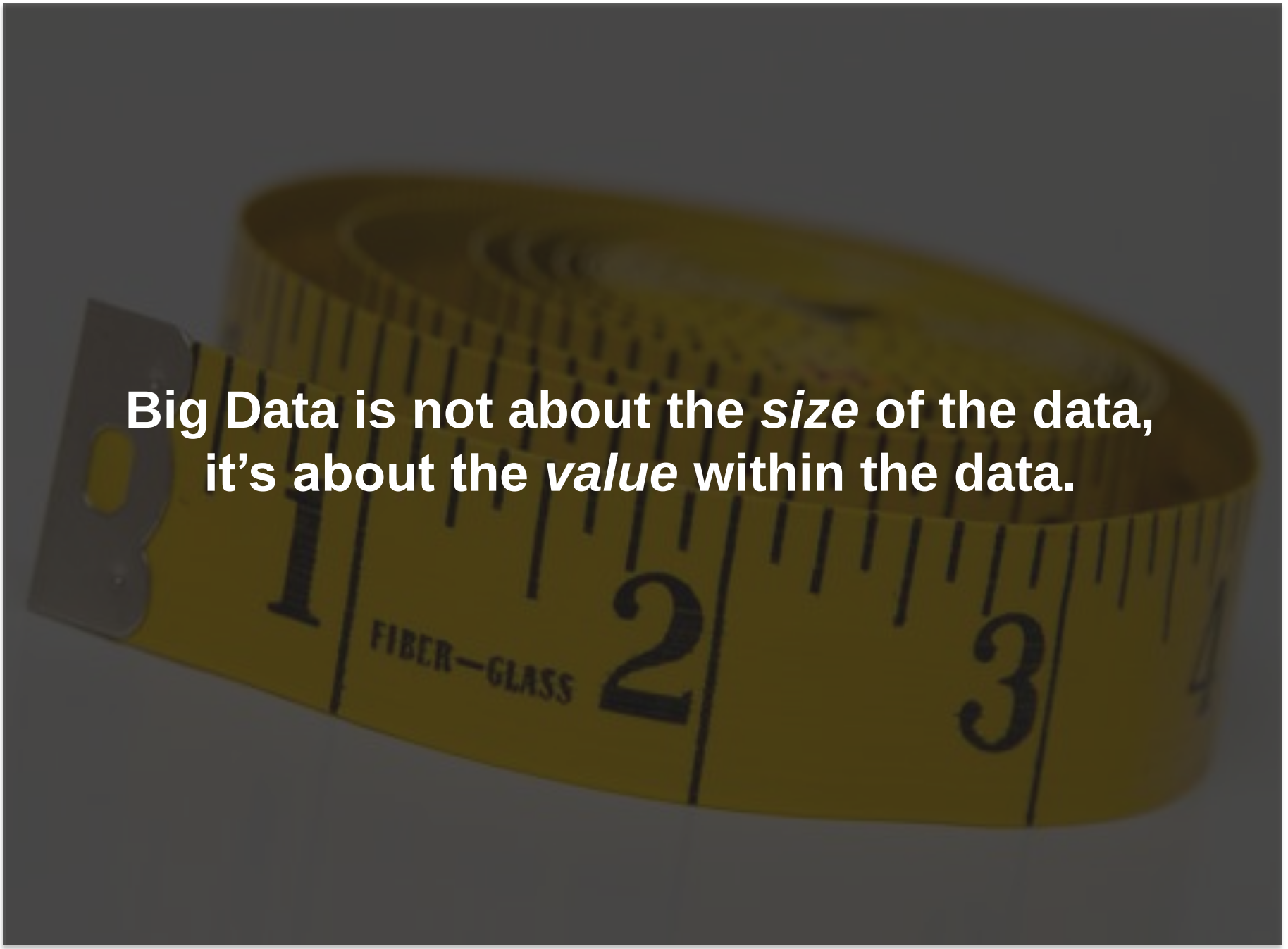
Fact 15

The big data industry is expected to grow from US$10.2 billion in 2013 to about US$54.3 billion by 2017.

How is Big Data different?

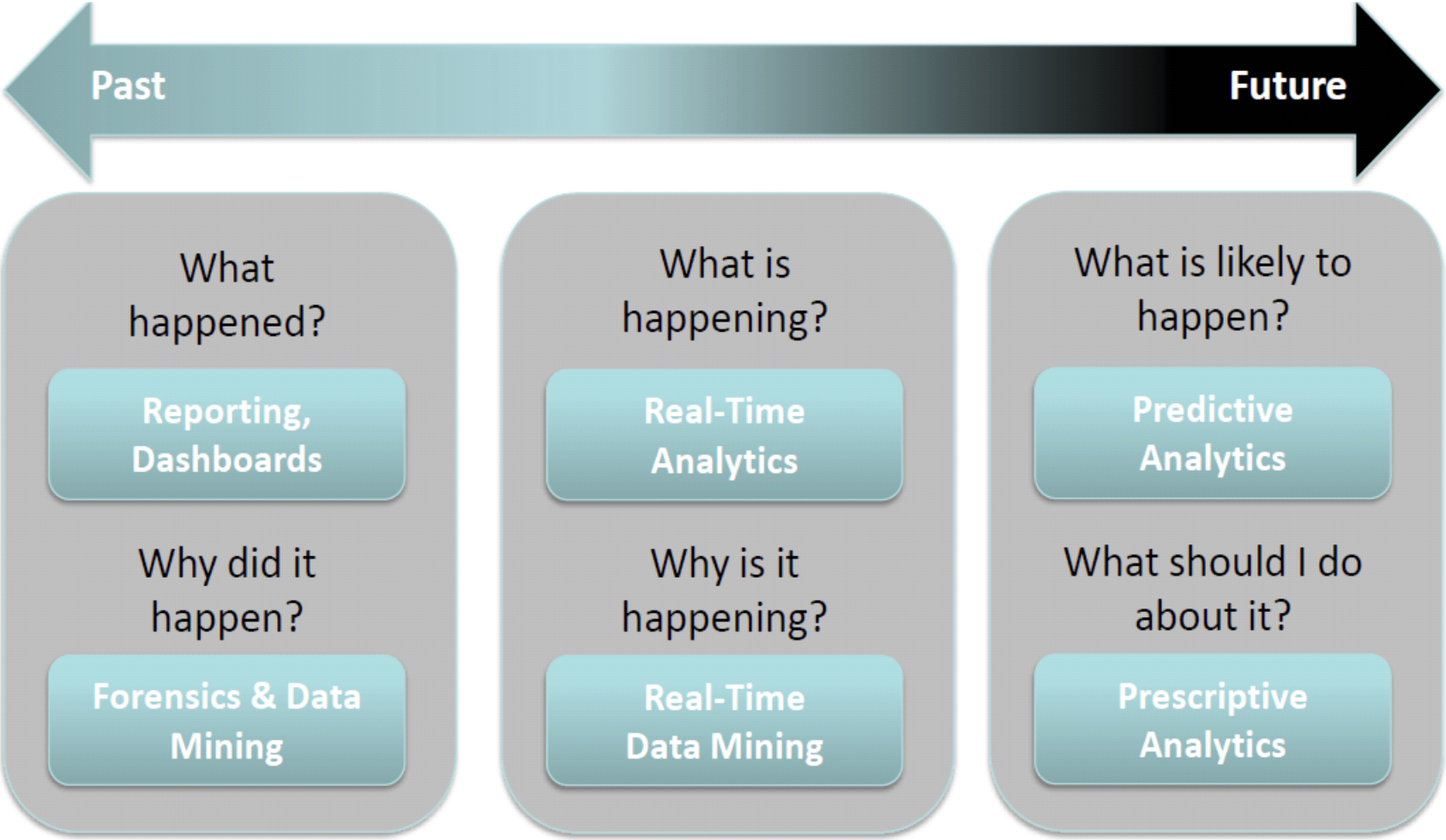
- ❖Structured
  - Most traditional data sources
- ❖Semi-structured
  - Many sources of big data
- ❖Unstructured
  - Video data, audio data



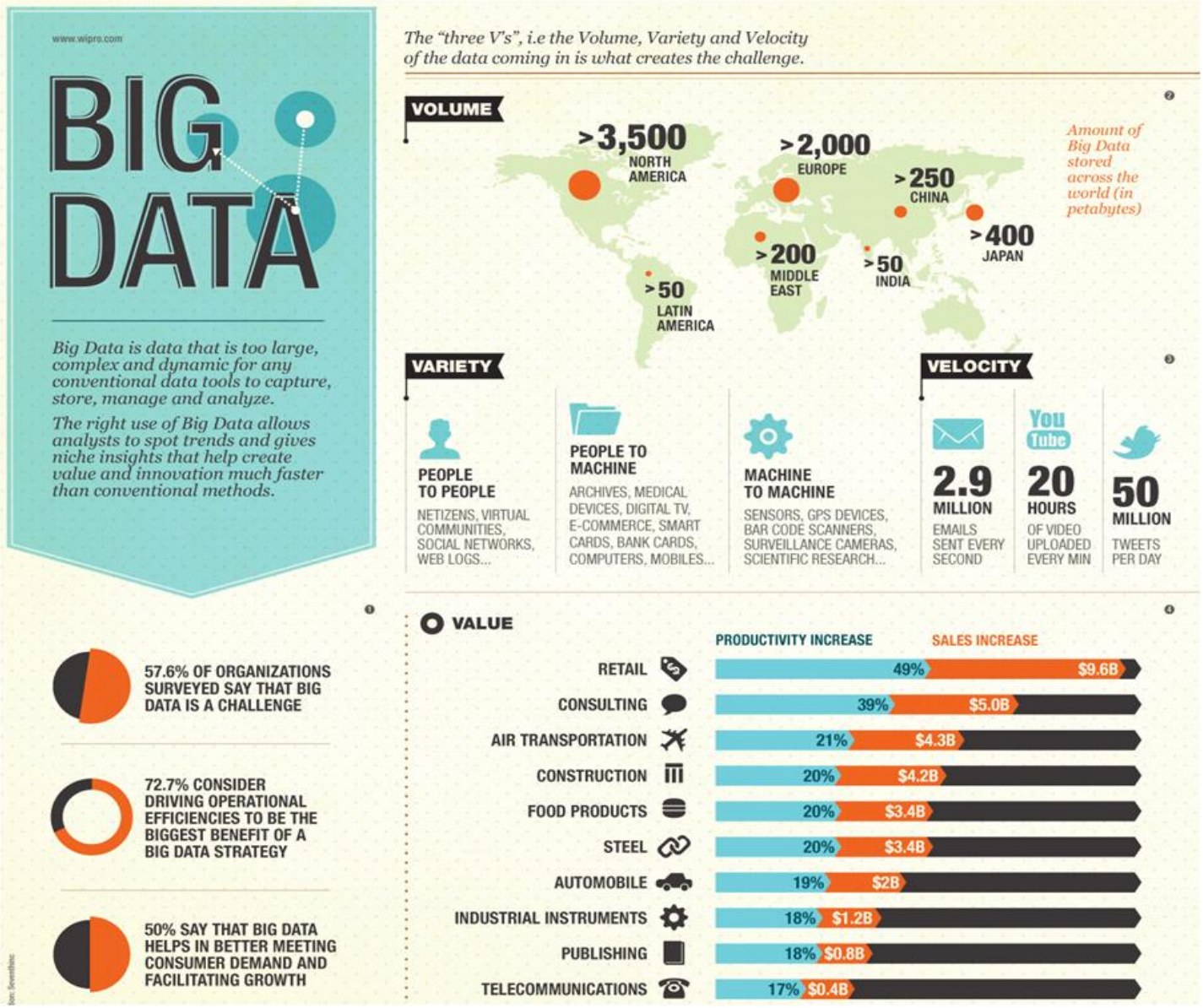


Big Data is not about the *size* of the data, it's about the *value* within the data.

Analyzing BIG DATA (ANALYTICS)… to obtain Value



Big Data –> Volume, Variety, Velocity = Value!



Businesses & Industry are Embracing Big Data

Retail

- CRM – Customer Scoring
- Store Siting and Layout
- Fraud Detection / Prevention
- Supply Chain Optimization

Advertising & Public Relations

- Demand Signaling
- Ad Targeting
- Sentiment Analysis
- Customer Acquisition

Financial Services

- Algorithmic Trading
- Risk Analysis
- Fraud Detection
- Portfolio Analysis

Media & Telecommunications

- Network Optimization
- Customer Scoring
- Churn Prevention
- Fraud Prevention

Manufacturing

- Product Research
- Engineering Analytics
- Process & Quality Analysis
- Distribution Optimization

Energy

- Smart Grid
- Exploration

Government

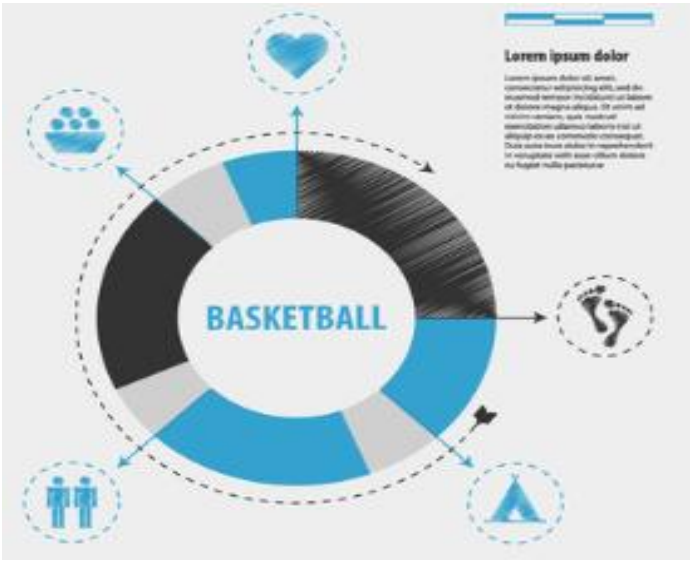
- Market Governance
- Counter-Terrorism
- Econometrics
- Health Informatics

Healthcare & Life Sciences

- Pharmaco-Genomics
- Bio-Informatics
- Pharmaceutical Research
- Clinical Outcomes Research

Practical Examples of how big data is used today to deliver real value

Improving Security and Law Enforcement: Security services use big data analytics to foil terrorist plots and detect cyber attacks. Police forces use big data tools to catch criminals and even predict criminal activity and credit card companies use big data analytics it to detect fraudulent transactions.



Improving Sports Performance: Most elite sports have now embraced big data analytics. Many use video analytics to track the performance of every player in a football or baseball game, sensor technology is built into sports equipment such as basket balls or golf clubs, and many elite sports teams track athletes outside of the sporting environment – using smart technology to track nutrition and sleep, as well as social media conversations to monitor emotional wellbeing.

Your Turn:
Volume, Variety, Velocity
What is the Value of this data?

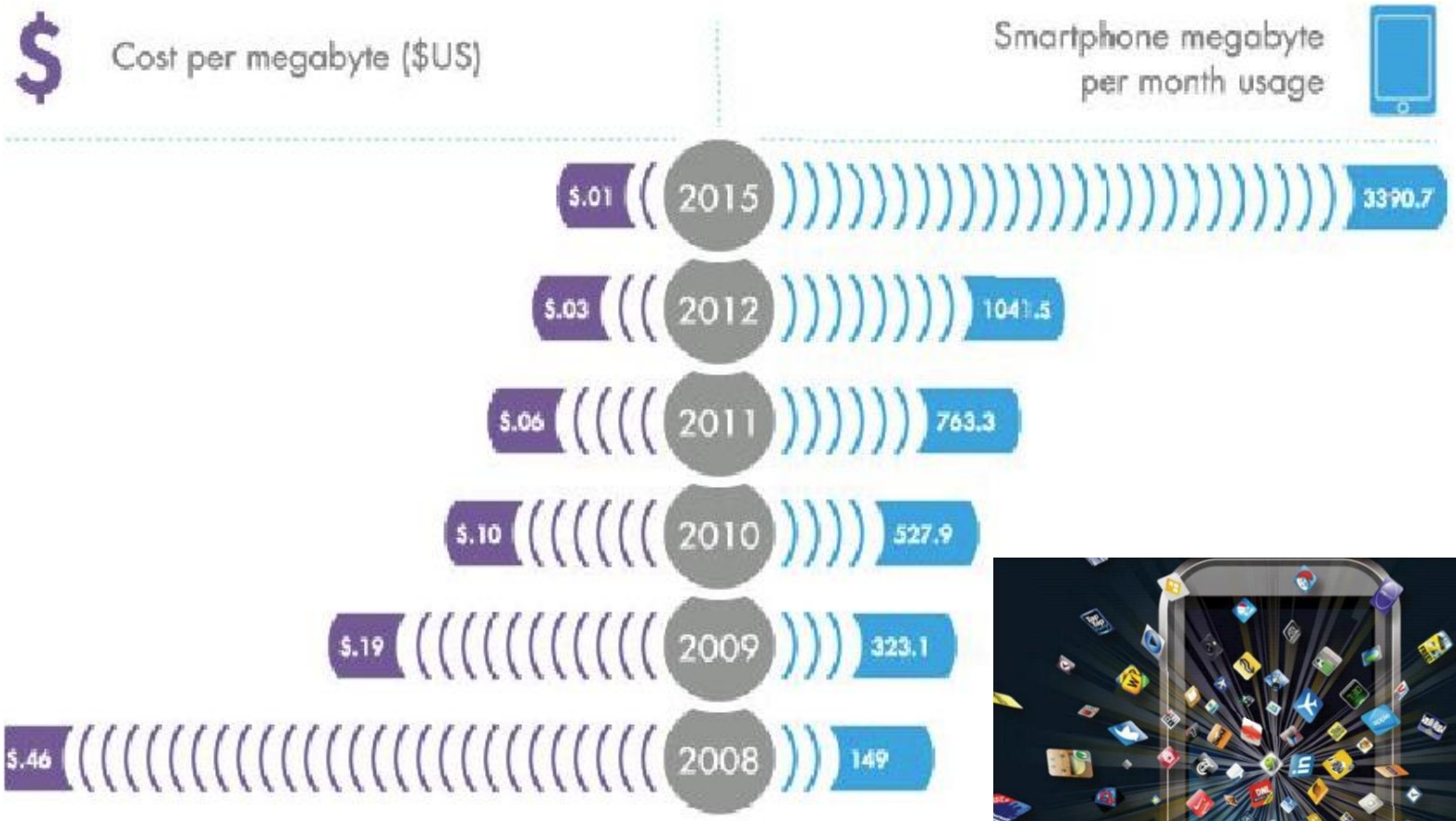
8	1	6
3	5	7
4	9	2

?

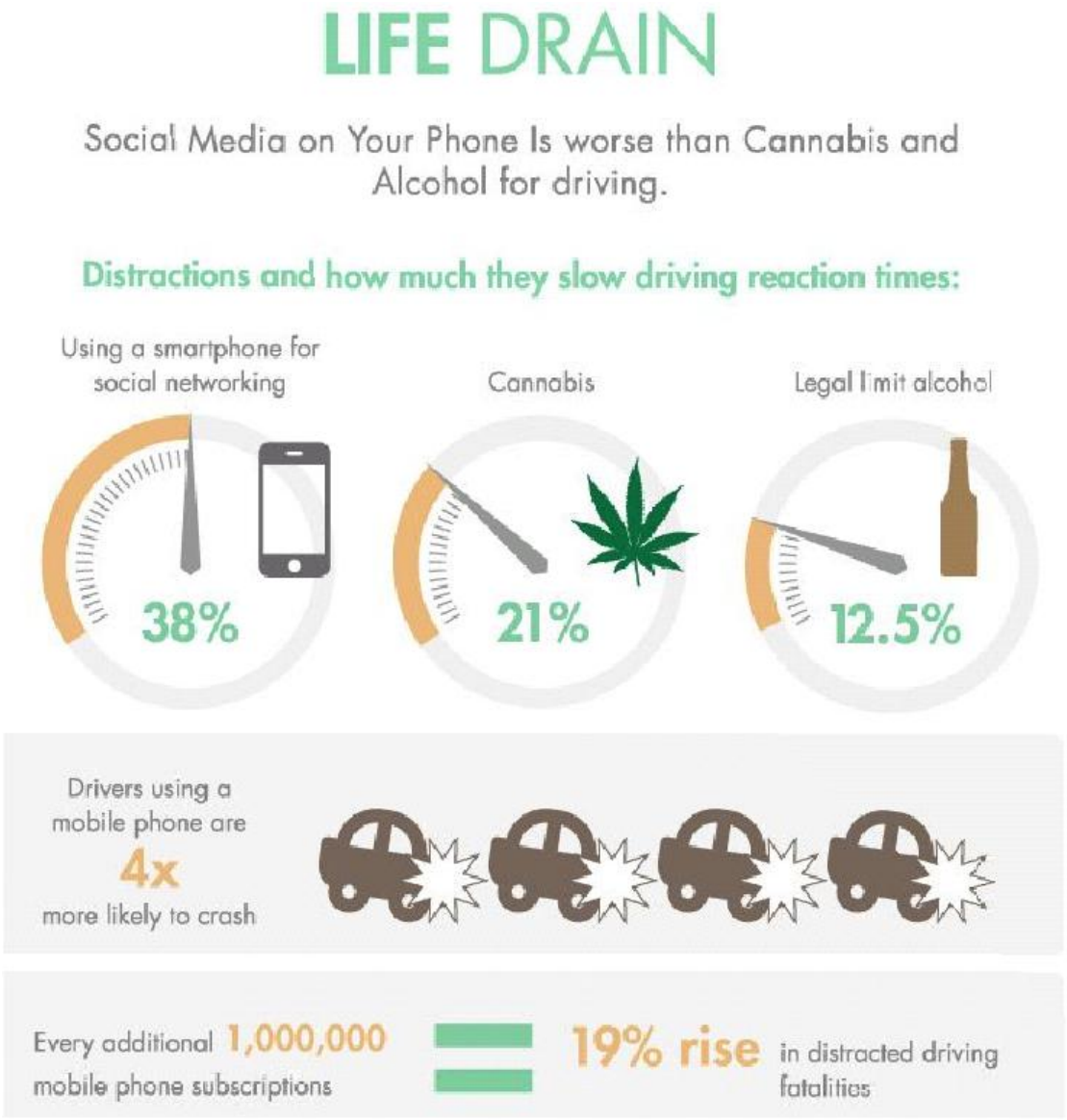
Volume, Variety, Velocity
What is the Value of this data?

	8	1	6	15
	3	5	7	15
	4	9	2	15
15	15	15	15	15

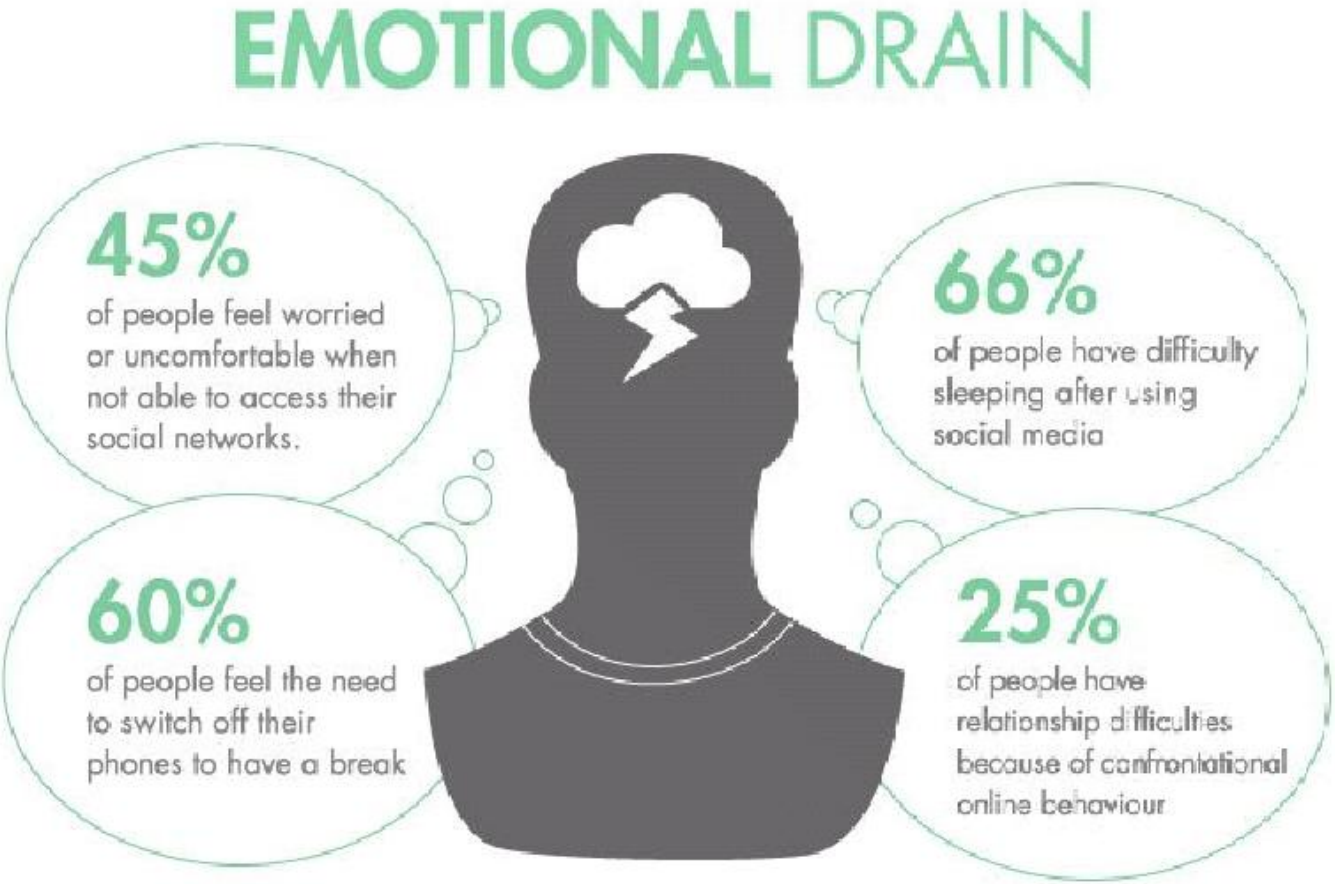
Do you have a healthy relationship with your smart phone?



Do you have a healthy relationship with your smart phone?



Do you have a healthy relationship with your smart phone?



A study has found that turning off mobile phones and avoiding the internet can leave people suffering from symptoms similar to those seen in drug addicts trying to go 'cold turkey'.

Data Exhaust – What about <u>your</u> history?

.

What does your Social Profile say about you?

Employers are watching: what does your social media profile say about you?

Your Social profile…

In addition to all the historic data analysts have at their disposal, social media is offering recruiters a rich new vein of real-time data.

Our blogs, websites, Twitter rants and LinkedIn profiles reveal as much - if not more - about us than a semi-fictionalized resume.

The days of keeping your personal and professional profiles separate are over

Social media is a great platform for individuals to demonstrate their expertise, experience and enthusiasm for their field of specialism.

However, candidates need to be conscious of the online reputation they are building and the data trail they are leaving behind

A growing number of tech companies are offering tools that can sift through masses of social media data and spot patterns of behavior and sentiment.

It's all about reputation. If people can't manage their own reputations, how are they going to protect the reputations of their future employers?

Does your job success depend more on data than your resume?

Nearly half of new recruits turn out to be duds within 18 months, according to a recent study, while two-thirds of hiring managers admit they've often chosen the wrong people

The main reason for failure is not because applicants didn't have the requisite skills, but because their personalities clashed with the company's culture

Employers are now resorting to big data analytics and other new methods to help make the fraught process of hiring and firing more scientific and effective

For job hunters, this means success is now as much to do with your online data trail as your finely crafted resume

Game for a job?

Recruitment technology firm Electronic Insight doesn't even bother to look at your skills and experience when analyzing resumes

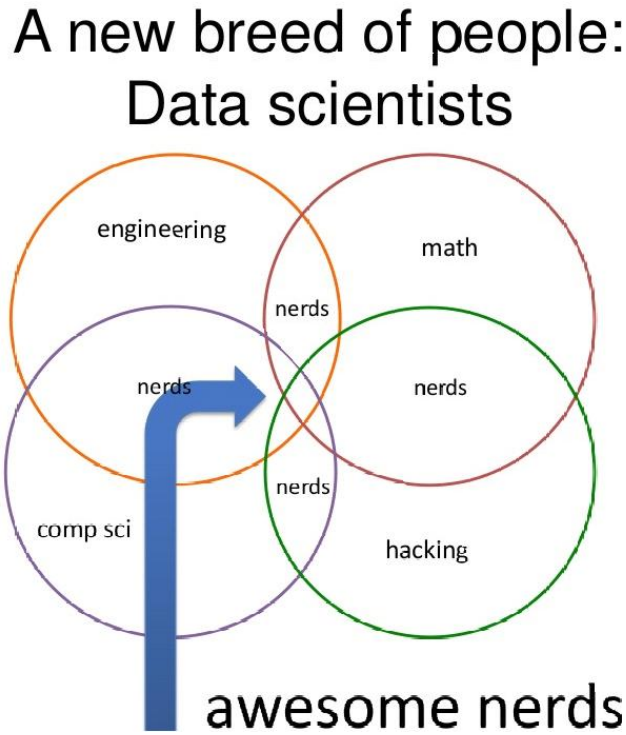
Companies such as Silicon Valley start-up Knack are even developing games as a way of assessing the suitability of job candidates.

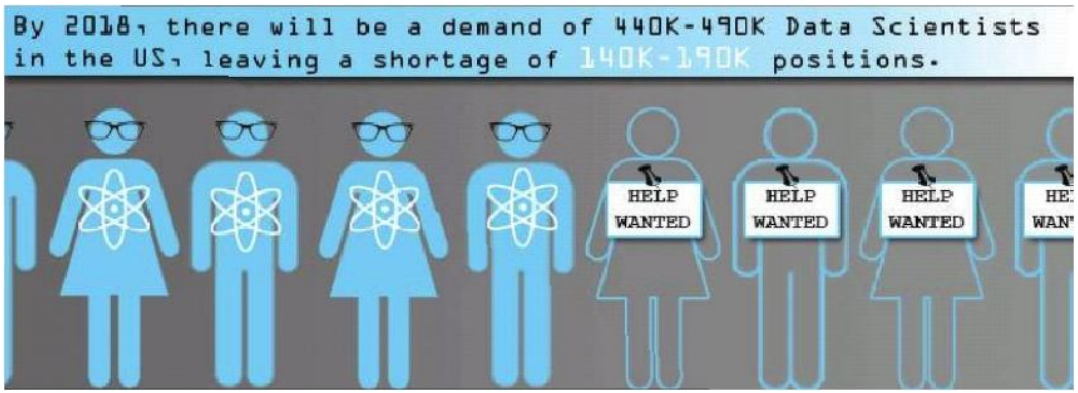
While applicants play an online game designed to reveal their personality, emotional maturity and problem-solving skills, hundreds of pieces of information are being collected in the background and analyzed by data scientists.

Gamification is definitely coming in… as games can tell if you're a risk taker or innovator and they appeal to today's gaming culture generation

New Talent is needed to address BIG DATA challenges…

Data Scientist… Nerd Talent Shortage





Real World Examples…….

NFL – Analytics at work
https://www.youtube.com/watch?v=AbYW1kPfnvs&list=PLUbRx39vvOvhCl4deCOJhhYQ3cdp8qoc9
(7:55 – 18:00)
https://www.youtube.com/watch?v=AbYW1kPfnvs&list=PLUbRx39vvOvhCl4deCOJhhYQ3cdp8qoc9#t=475

Cognitive Computing / Analytics – Watson
https://www.youtube.com/watch?v=Y_cqBP08yuA
https://www.youtube.com/watch?v=np1sJ08Q7lw

IBM History 100 x 100
https://www.youtube.com/watch?v=39jtNUGgmd4

Thank You.


Variety
Velocity
Volume
Structured &
Unstructured
Batch
Big
Data
Structured
Streaming Data
Zettabytes
Terabytes