

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...?

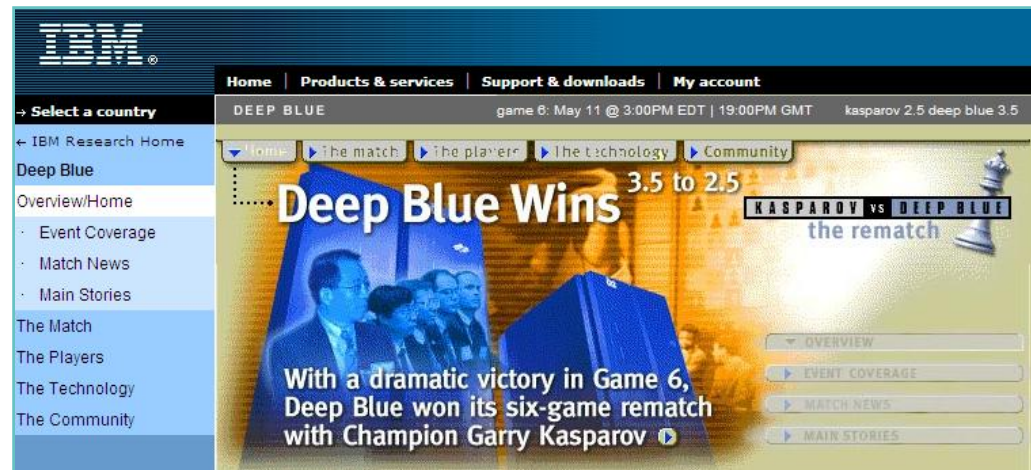
David Bechtold

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.



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



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.**”

whatever you're
THINKING
THINK
BIGGER

Let's look at
Big Data

in a different way... a perspective...

Byte : one grain of rice



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



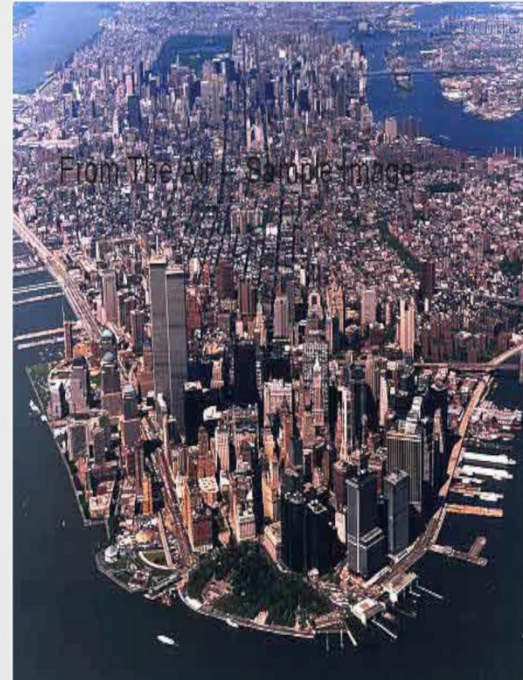
Gigabyte

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



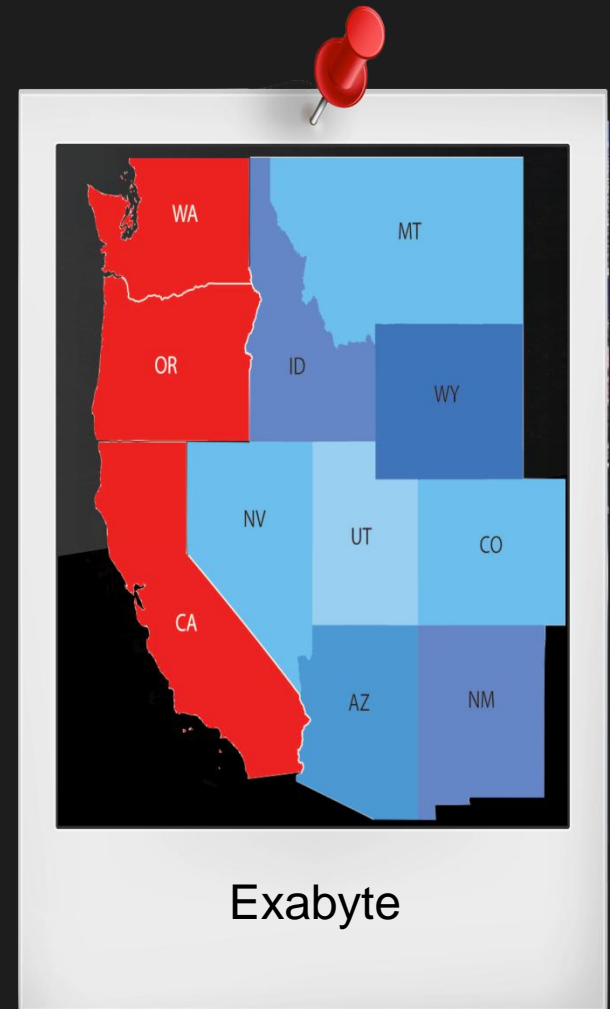
Terabyte

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



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

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!



Hobbyist



Desktop



Internet



Big Data

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!

facebook

YAHOO!

amazon.com

ebay

Google

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!



Hobbyist



Desktop



Internet



Big Data

The Future?

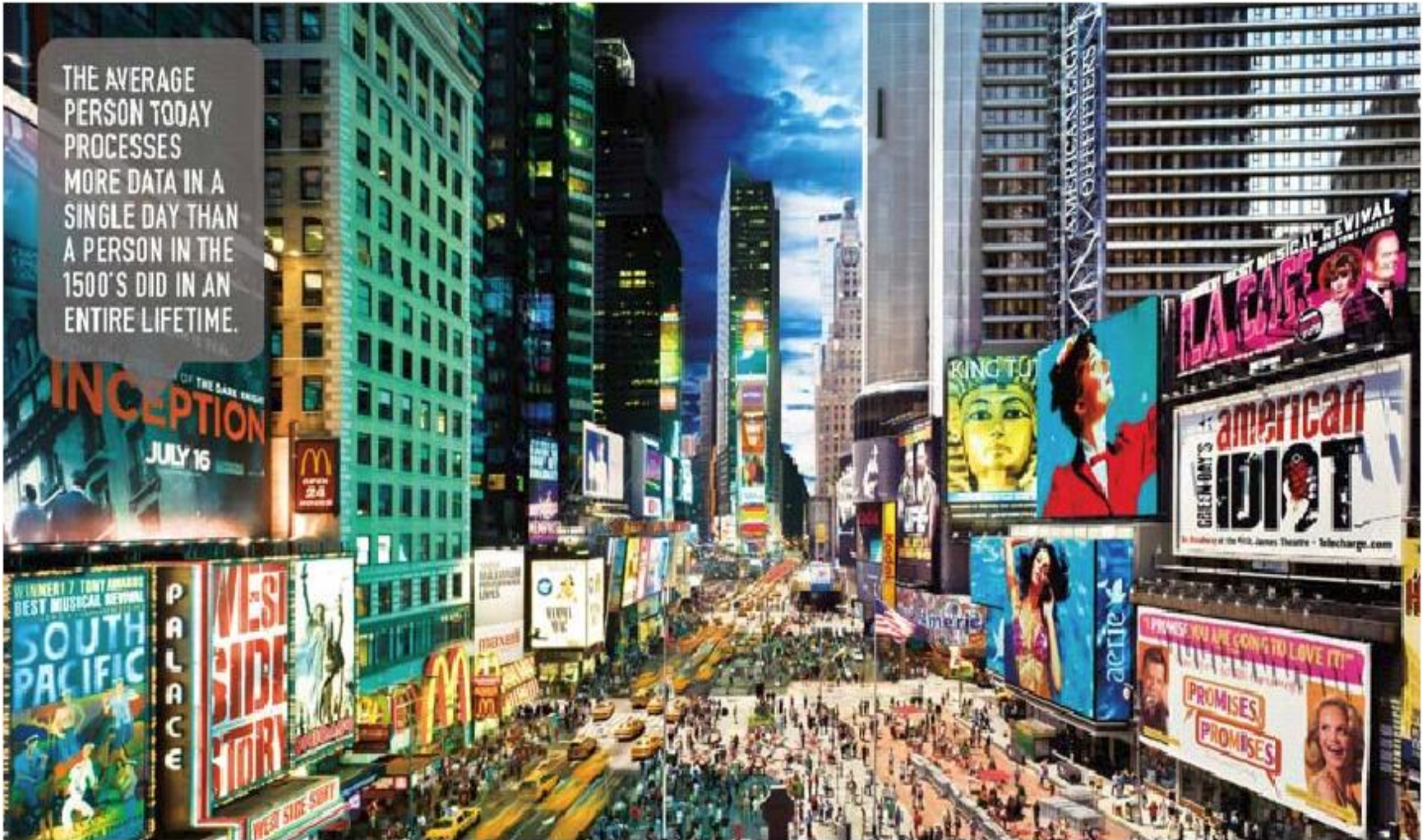
BIG DATA spans three dimensions: Volume, Velocity and Variety

- **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



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.



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...



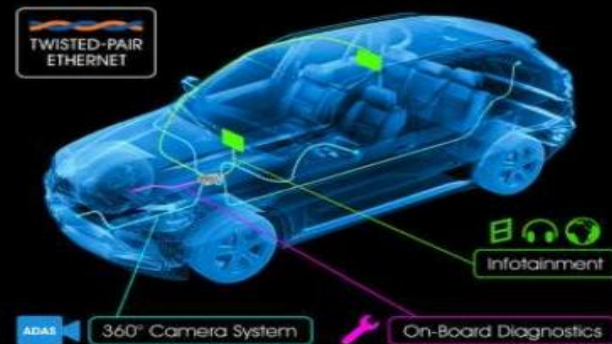
Data Velocity & Transparency...

DATA TRANSPARENCY,
AMPLIFIED BY TWITTER AND
FACEBOOK, HAS LED TO MASS
MOVEMENTS ON A SCALE
NEVER SEEN BEFORE.



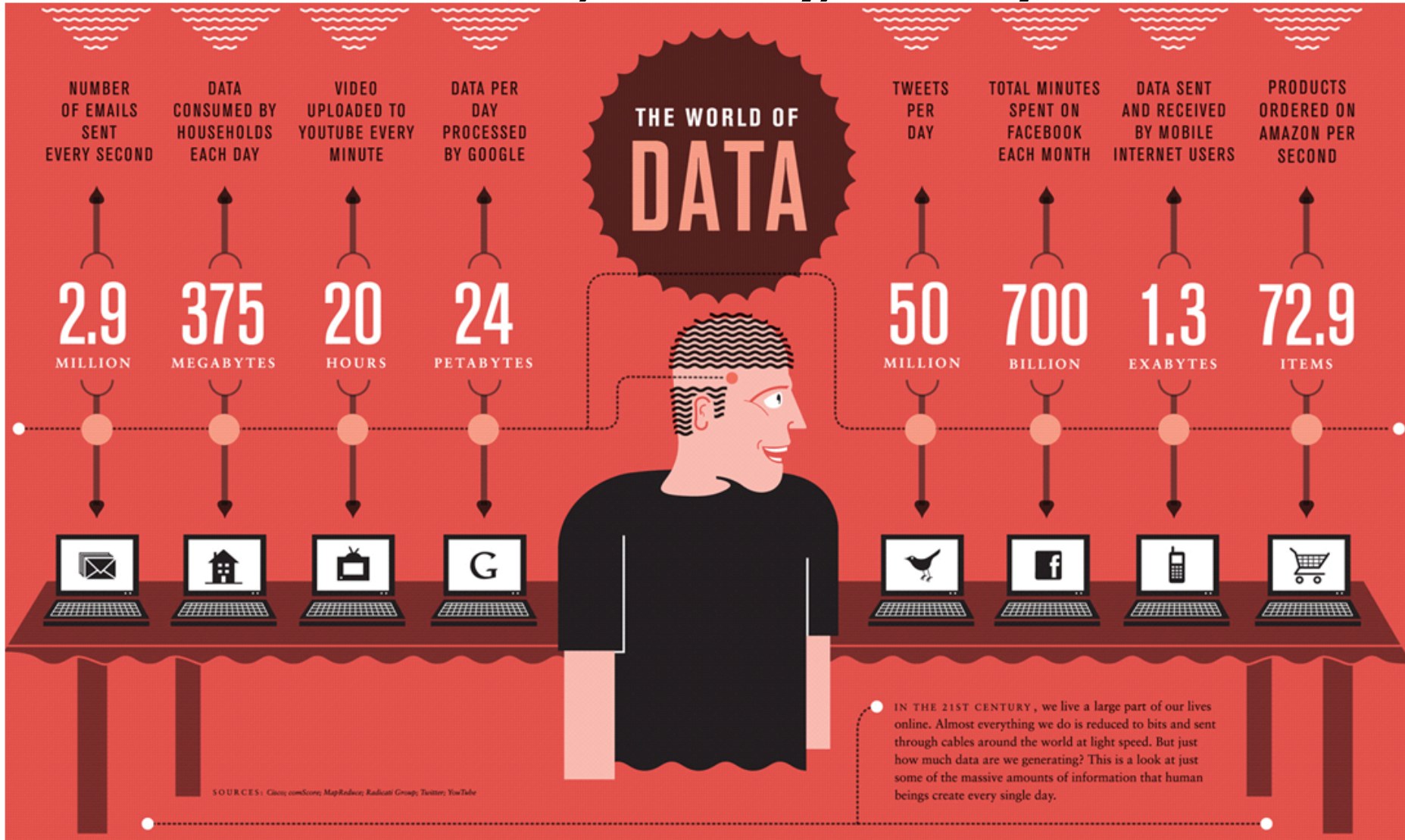
Data Variety...

New sources of data



A look at Data (sec, min, day... month)

Volume, Velocity, Variety



What happens when you search Google?

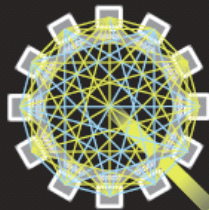
WHAT HAPPENS WHEN YOU GOOGLE?

Like most search engines, Google is continually “crawling” through the web, cataloging and storing billions of pages. When you search for something, the system calls up these cached pages so that it can respond to you quickly.

1 You enter a query.



2 It goes to the Google web server ...



... a network of more than 3 million computers linked together and connected to the Internet.

3 The query is sent to 1 million index servers.

These servers are like the index in a book: they indicate which pages contain the words that match your query and where those pages are stored in the document servers.



4 The query travels to document servers that retrieve the stored pages.



Using PageRank software (named after Larry Page), the system measures the importance of a page by solving an equation of more than 500 million variables and 2 billion terms. It then displays a ranking of the best-known and most-visited pages. Since these pages have the most complete information on the given subject, they are the ones you're likely to want.

5 Snippets are generated to describe each search result ...

6 ... which are then returned to you.



7 The whole process takes about half a second.



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

BIG DATA

IN A SINGLE DAY ONLINE

ENOUGH INFORMATION IS CONSUMED TO FILL
168 MILLION DVDS

294bn E-MAILS
ARE SENT

MINUTES SPENT
ON FACEBOOK **4.7M**

2 MILLION BLOG POSTS
ARE WRITTEN

VIDEO UPLOADED TO
YOUTUBE **864,000 HRS**

MORE IPHONES
ARE SOLD THAN BABIES BORN

What to make of all that data...?

GLOBAL LOGISTICS

Every day, thousands of packages are transported across towns, countries, and entire continents. What does it really take to send a box from Switzerland to South Africa on time? In addition to a global network of delivery people, you need the power of logistics to bring air, ground, and water transport in sync with the latest high tech solutions for tracking and receiving. Here's a snapshot of how the global reach of UPS logistics facilitates package delivery to just about every destination, every day.

A DAY IN THE LIFE OF THE UPS FLEET

15.6 MILLION PACKAGES & DOCUMENTS DELIVERED DAILY

218 JETS IN SERVICE

942 DOMESTIC AND 815 INTERNATIONAL FLIGHT SEGMENTS PER DAY

92,734 DELIVERY CARS, VANS, TRACTORS & MOTORCYCLES

1,928 ALTERNATIVE FUEL VEHICLES

26.2 MILLION TRACKING REQUESTS PER DAY

27,280 FREIGHT TRACTORS & TRAILERS

UPS WORLDPORT BY THE NUMBERS

UPS WORLDPORT IS LOCATED IN LOUISVILLE, KENTUCKY

Worldport recently has undergone a billion dollar upgrade, and is now one of the most advanced distribution hubs in the world.

AS THE HEADQUARTERS FOR THE AIR FLEET, THE HUB HAS 70 PLANE DOCKS AND TURNS OVER **130 AIRCRAFT DAILY**

SORTS 416,000 PKGS PER HOUR, OR ABOUT 115 PKGS PER SECOND.

1.5 MILLION PACKAGES DELIVERED ON AN AVERAGE DAY

WORLDPORT MEASURES **5,200,000 SQ. FT.** OR ABOUT 90 FOOTBALL FIELDS

HAS 155 MILES OF CONVEYOR BELTS

AIRPORTS AND HUBS WORLDWIDE

EUROPE: COLOGNE/BONN, GERMANY

Covers 60 countries and territories

ASIA PACIFIC: CHINA: SHANGHAI, SHENZHEN, HONG KONG

Covers 40 countries and territories

UPS EMPLOYS OVER **400,000** PEOPLE WORLDWIDE

COUNTRIES AND TERRITORIES COVERED **220**

PREDICTING MOTHER NATURE

A TEAM OF **5** METEOROLOGISTS FOLLOW THE ATMOSPHERIC CONDITIONS **24** HOURS A DAY.

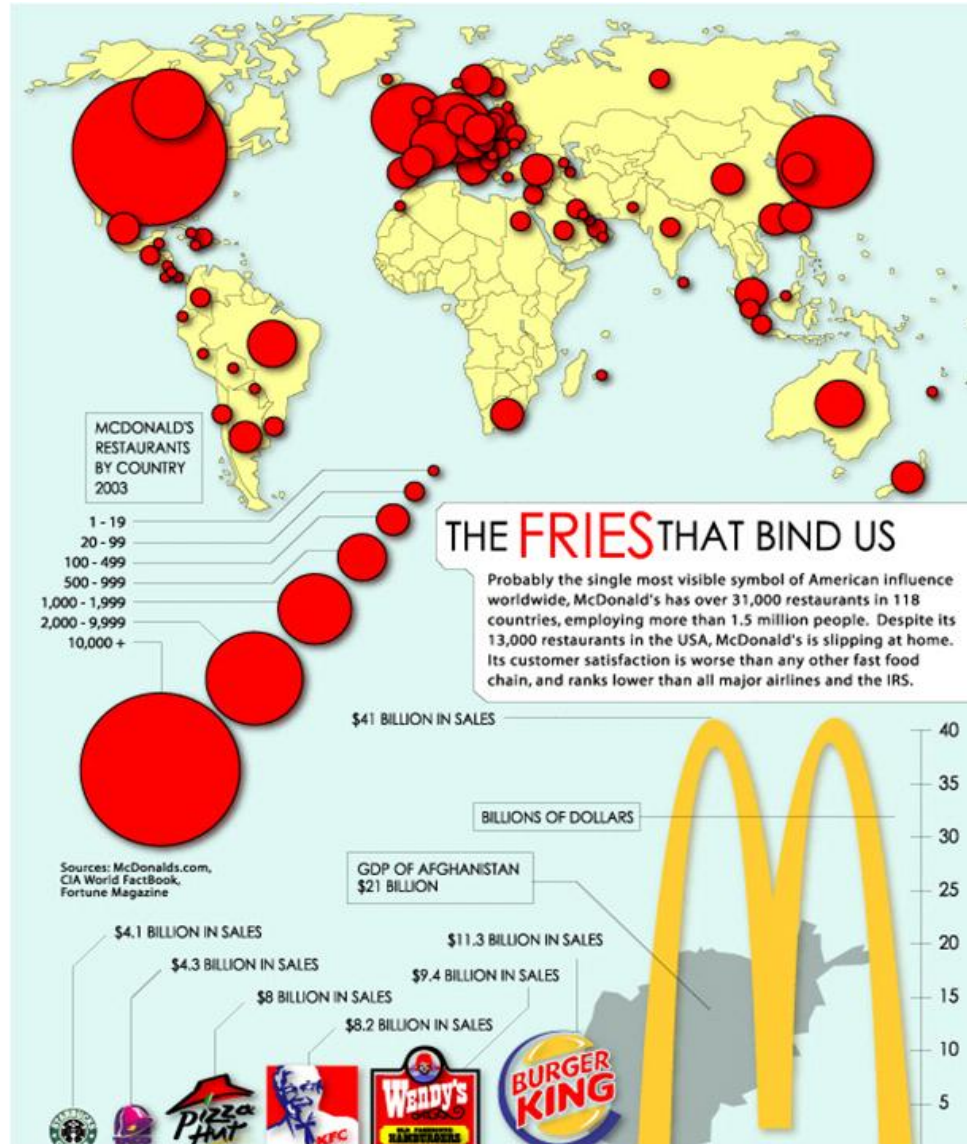
TO INSURE ON-TIME DELIVERY, THEY WORK WITH THE NATIONAL WEATHER SERVICE TO FORECAST WEATHER ALONG ROUTES AND WILL RE-ROUTE OR SWITCH MODES OF TRANSPORTATION TO AVOID BAD WEATHER.

UPS PLANES ARE EQUIPPED WITH SENSORS FOR REAL-TIME WEATHER DATA RETRIEVAL.

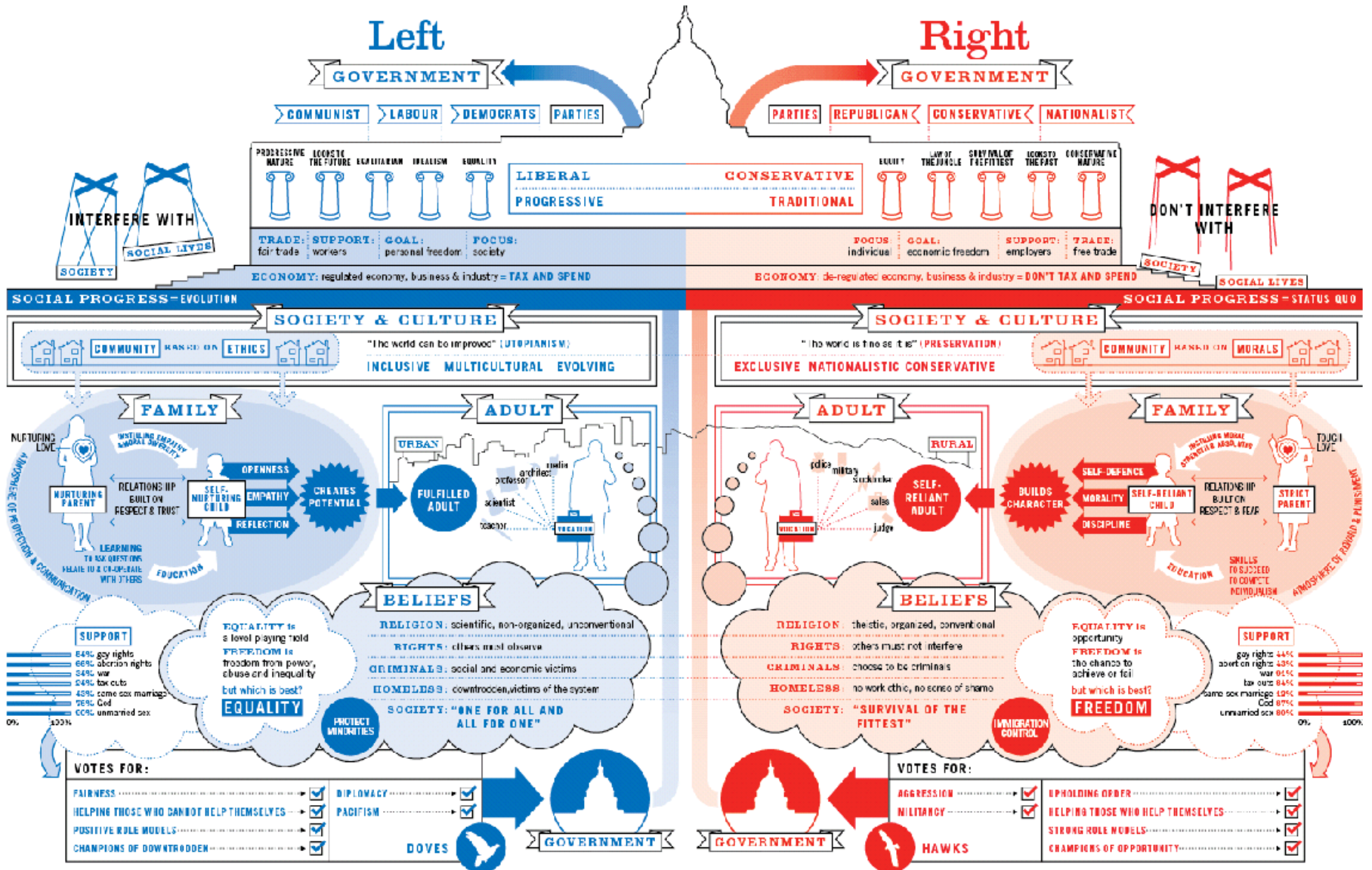
USE ON-SITE MODEL OF AN AIRPLANE WING TO PREDICT ICING AND FROST CONDITIONS IN THE AIR.

A COLLABORATION BETWEEN GOOD AND KIKI KARPUS, IN PARTNERSHIP WITH UPS WE/C/LOGISTICS

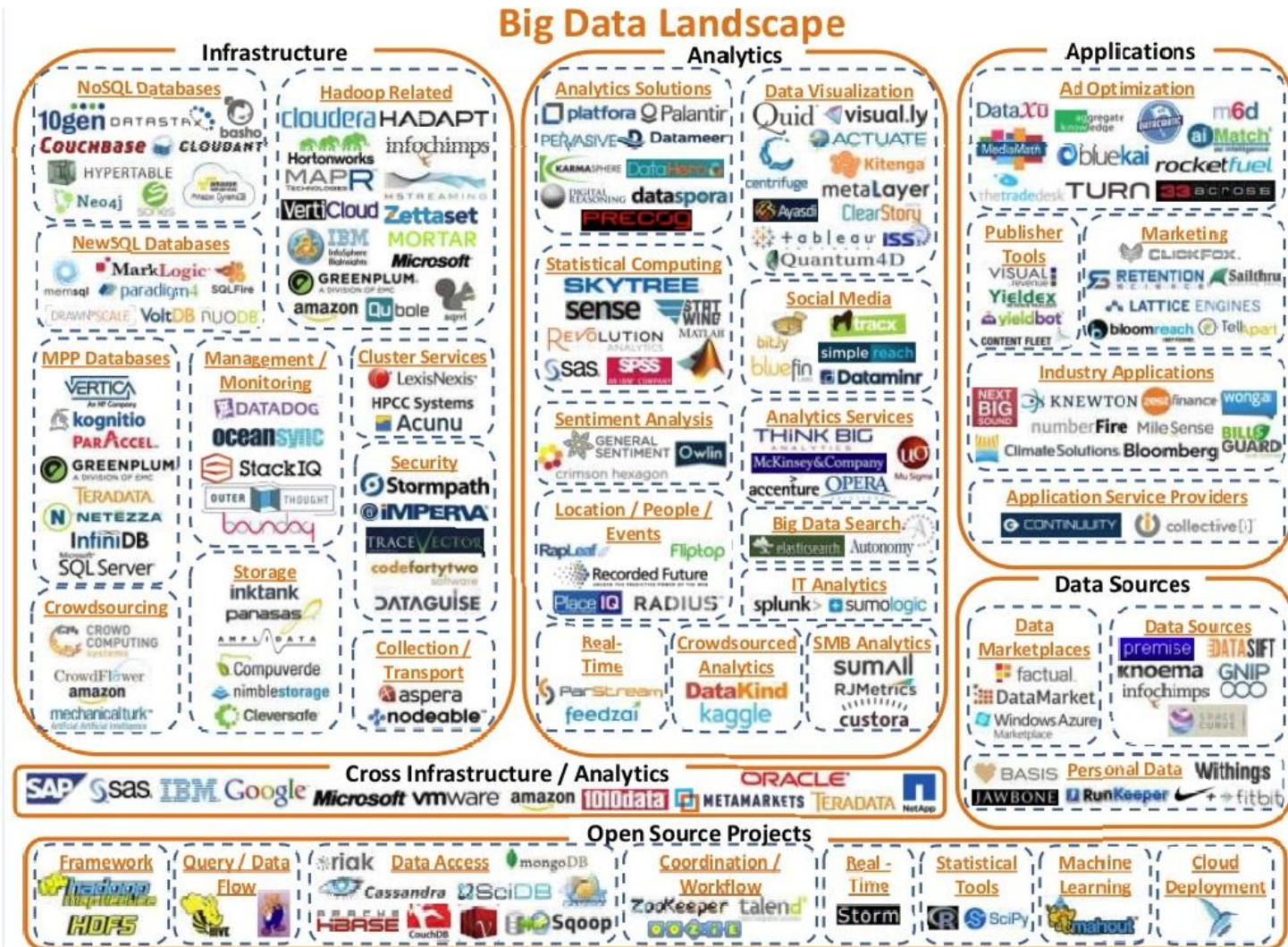
What to make of all that data...?



What to make of all that data...?



The BIG DATA landscape



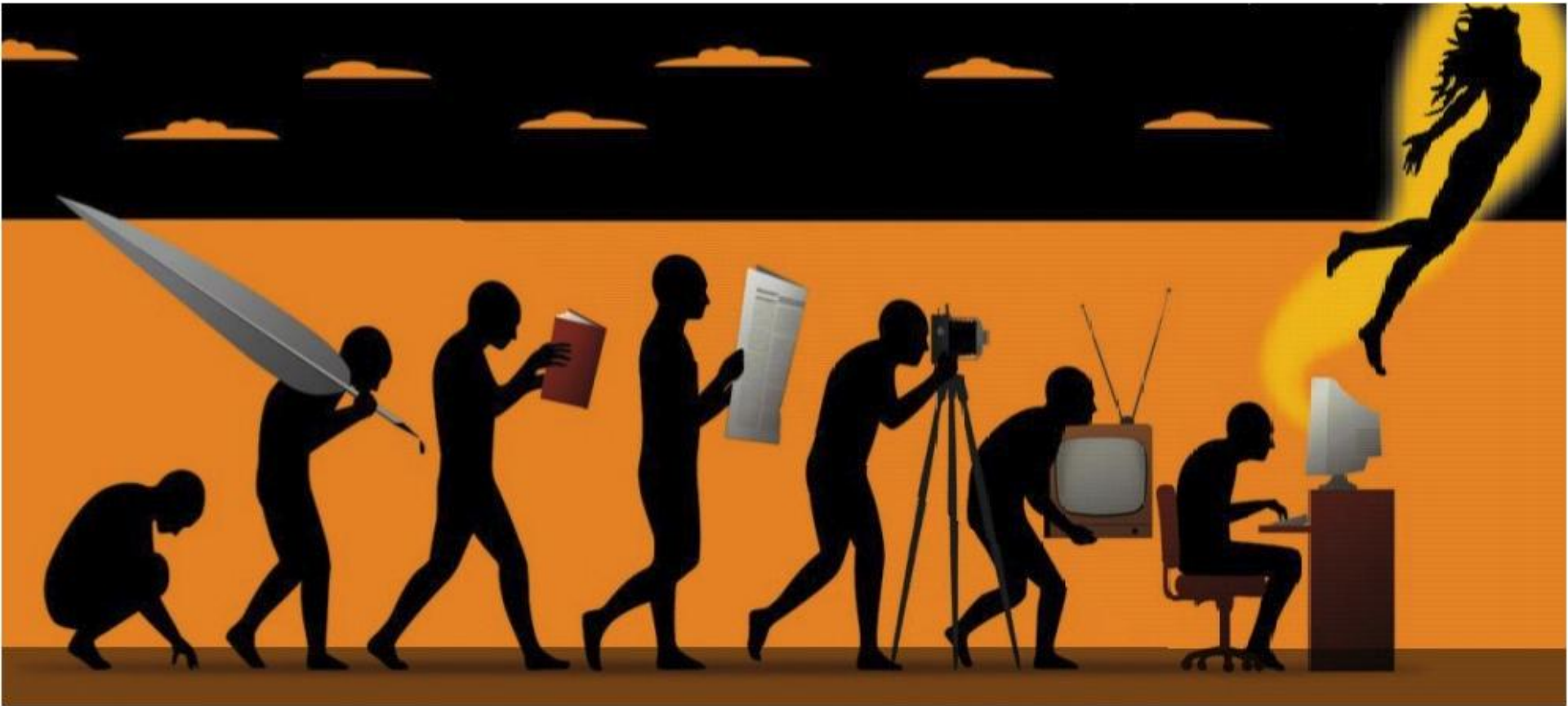
How did we get here...?

Volume

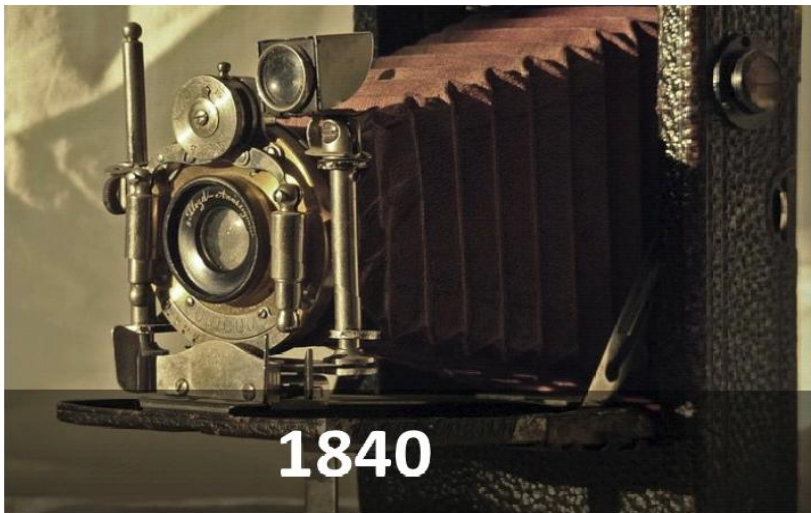
Velocity

Variety

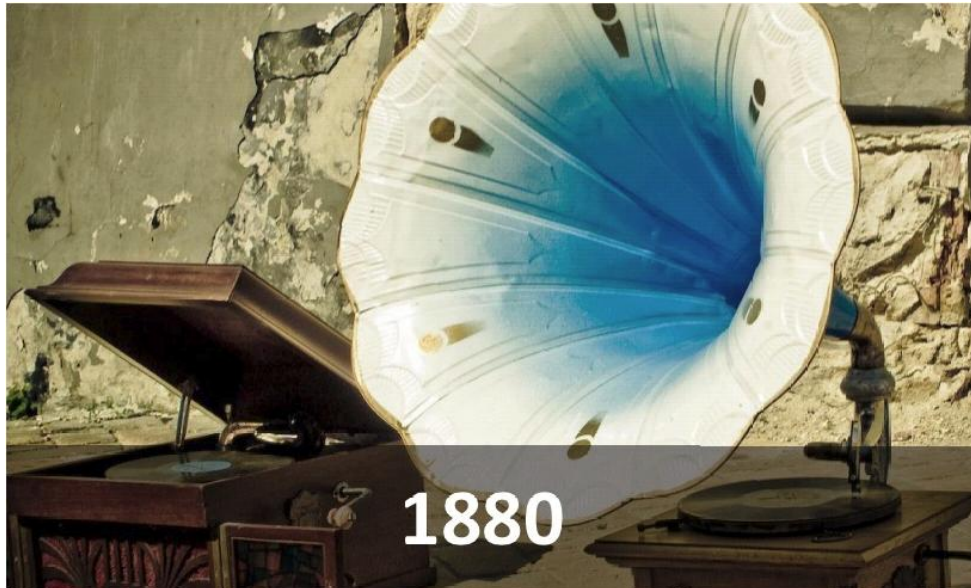
The rise of communication over time



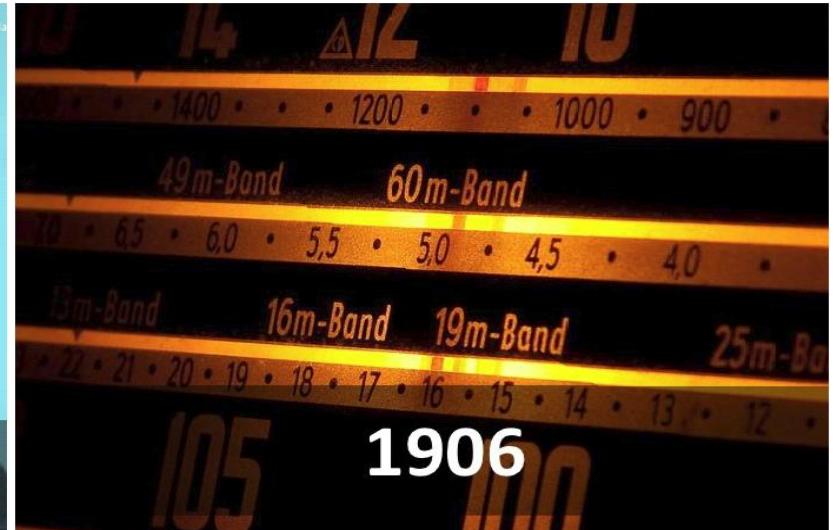
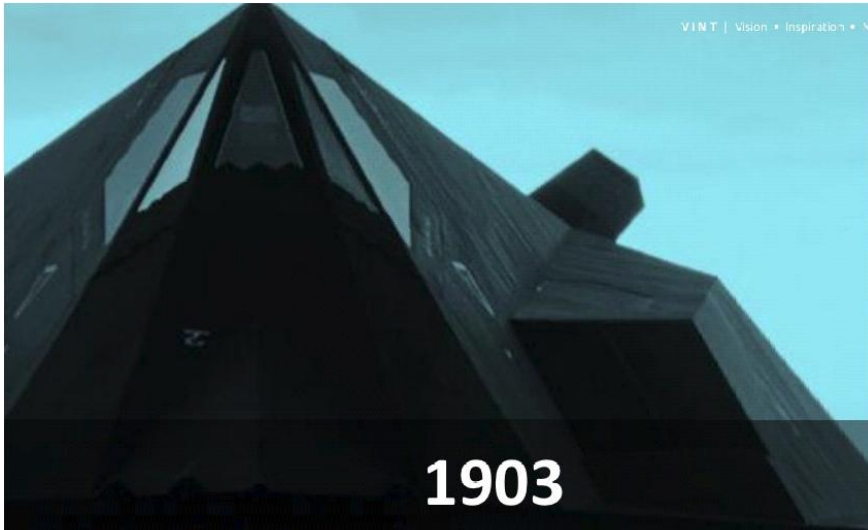
Data generation & communication over time



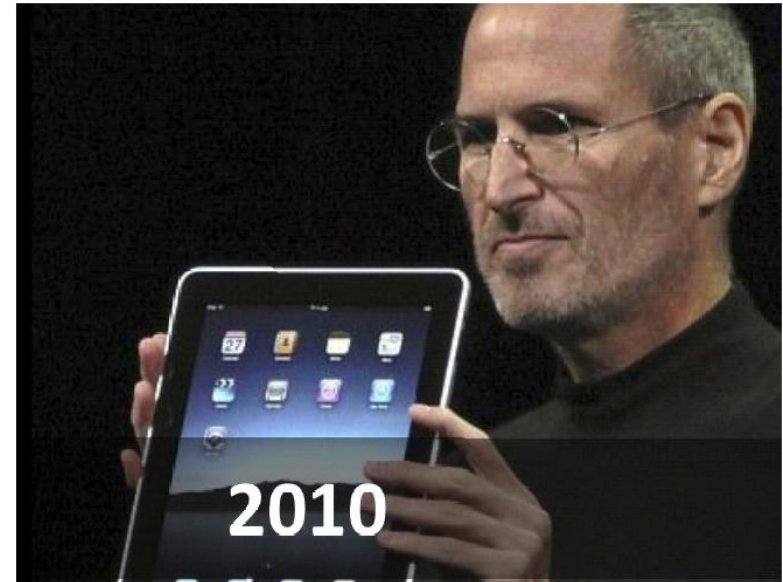
Data generation & communication over time



Data generation & communication over time

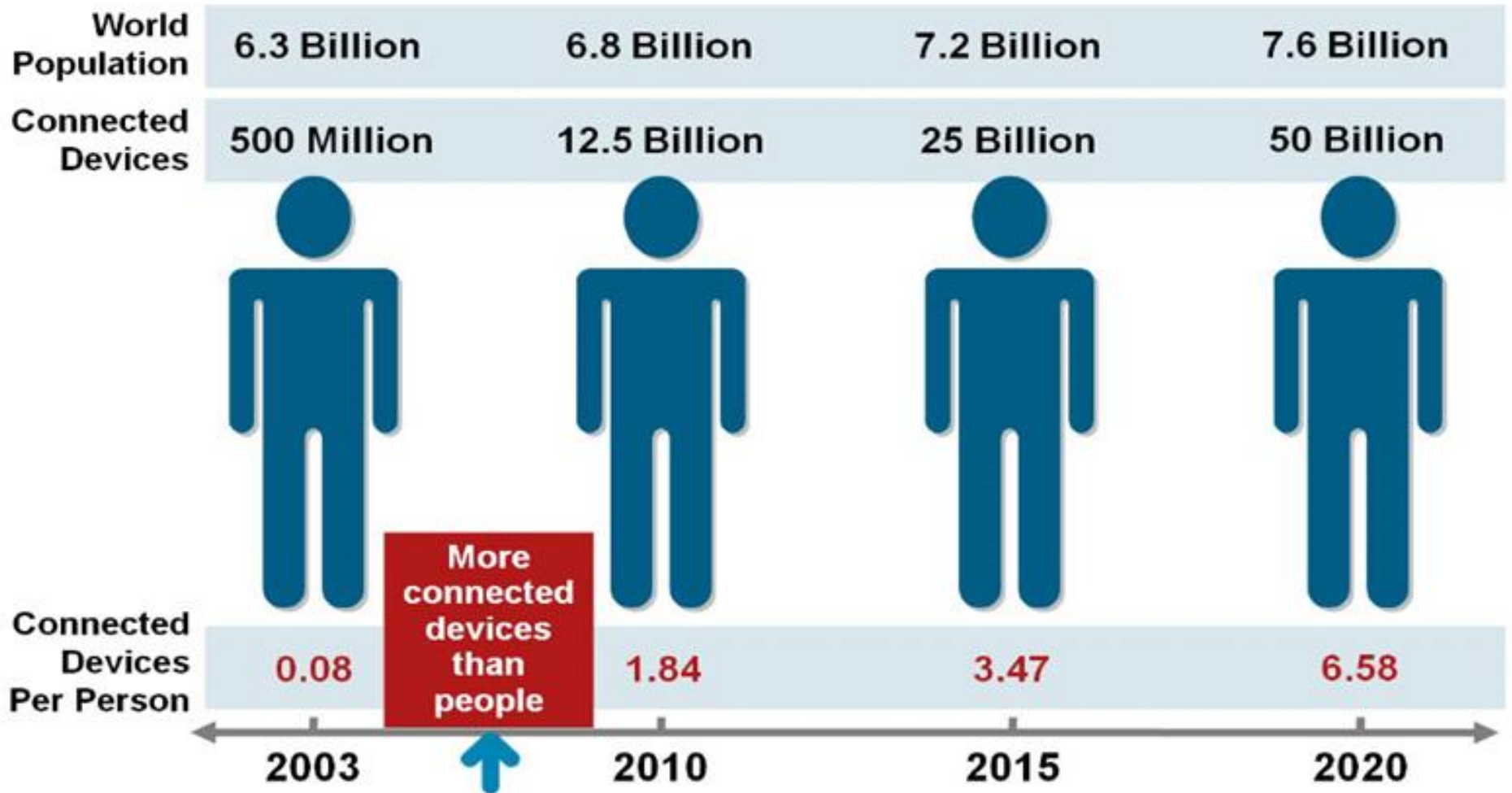


Data generation & communication over time



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

Driving Data Generation – Example:





15

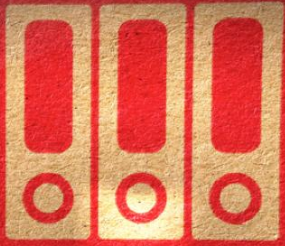
Need-to-Know

Facts

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



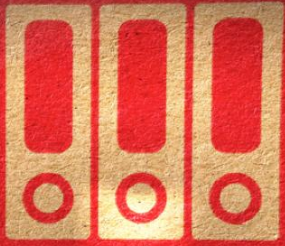
**Fact
1**



Big Data

Fact 2

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



Big Data

Fact

3

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

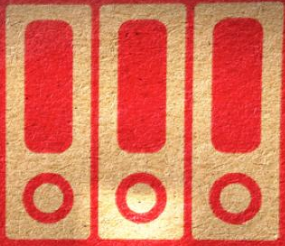
**Every minute
we send 204 million
emails, generate 1,8
million Facebook likes,
send 278 thousand
Tweets, and up-load
200,000 photos to
Facebook**



Big Data

Fact

4



Big Data

Fact

5

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

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.

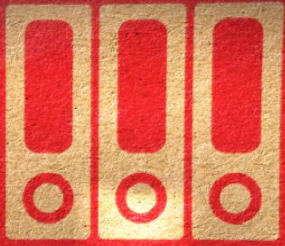


**Fact
6**

**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
7**



Big Data

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.**

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



**Fact
9**

**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
10**

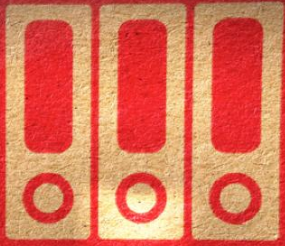


Big Data

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.**



Big Data

Fact

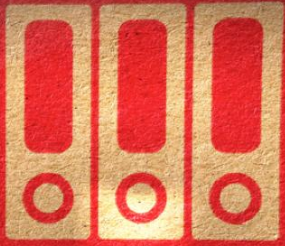
12

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.

**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
13**



Big Data

Fact

14

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

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



**Fact
15**

How is Big Data different?

❖ Structured

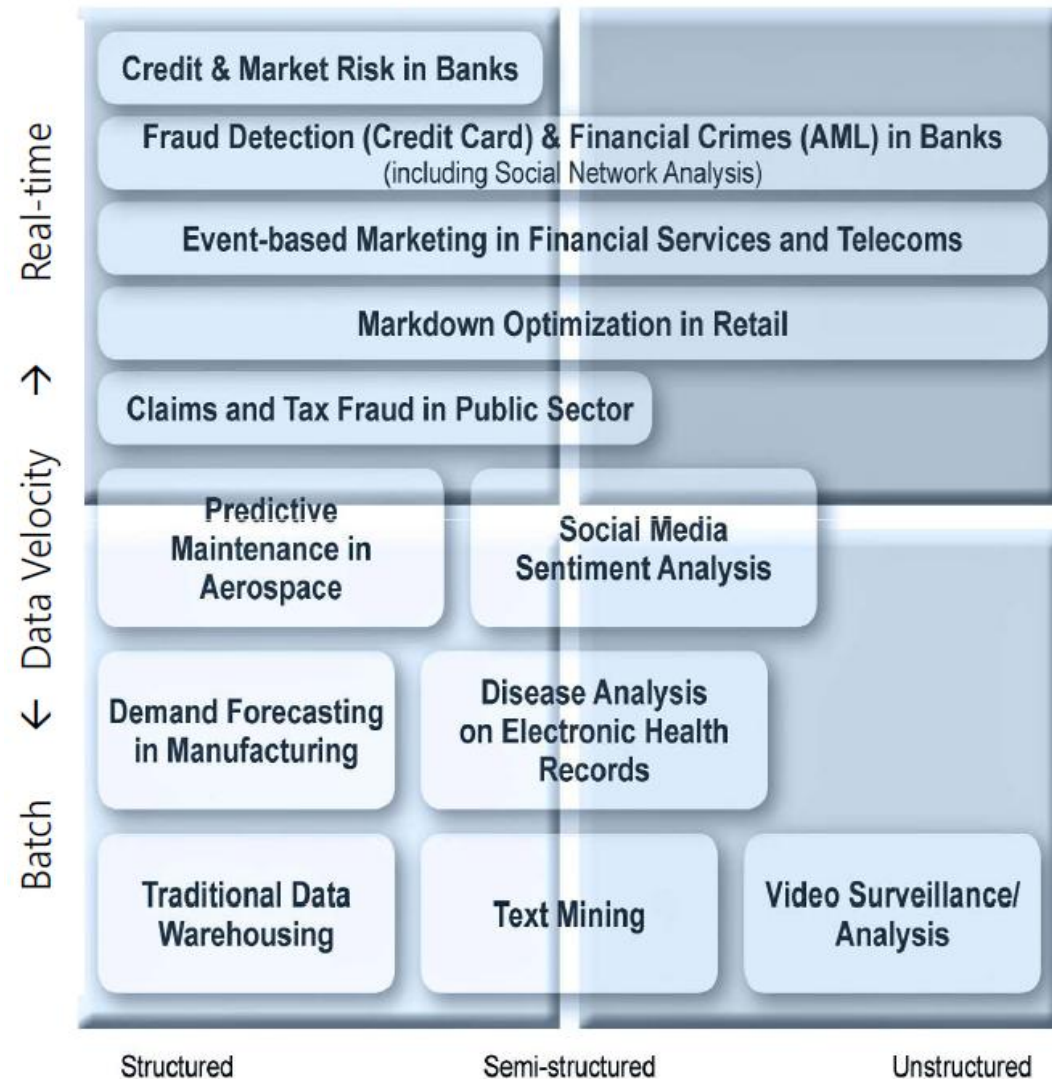
- Most traditional data sources


❖ Semi-structured

- Many sources of big data

❖ Unstructured

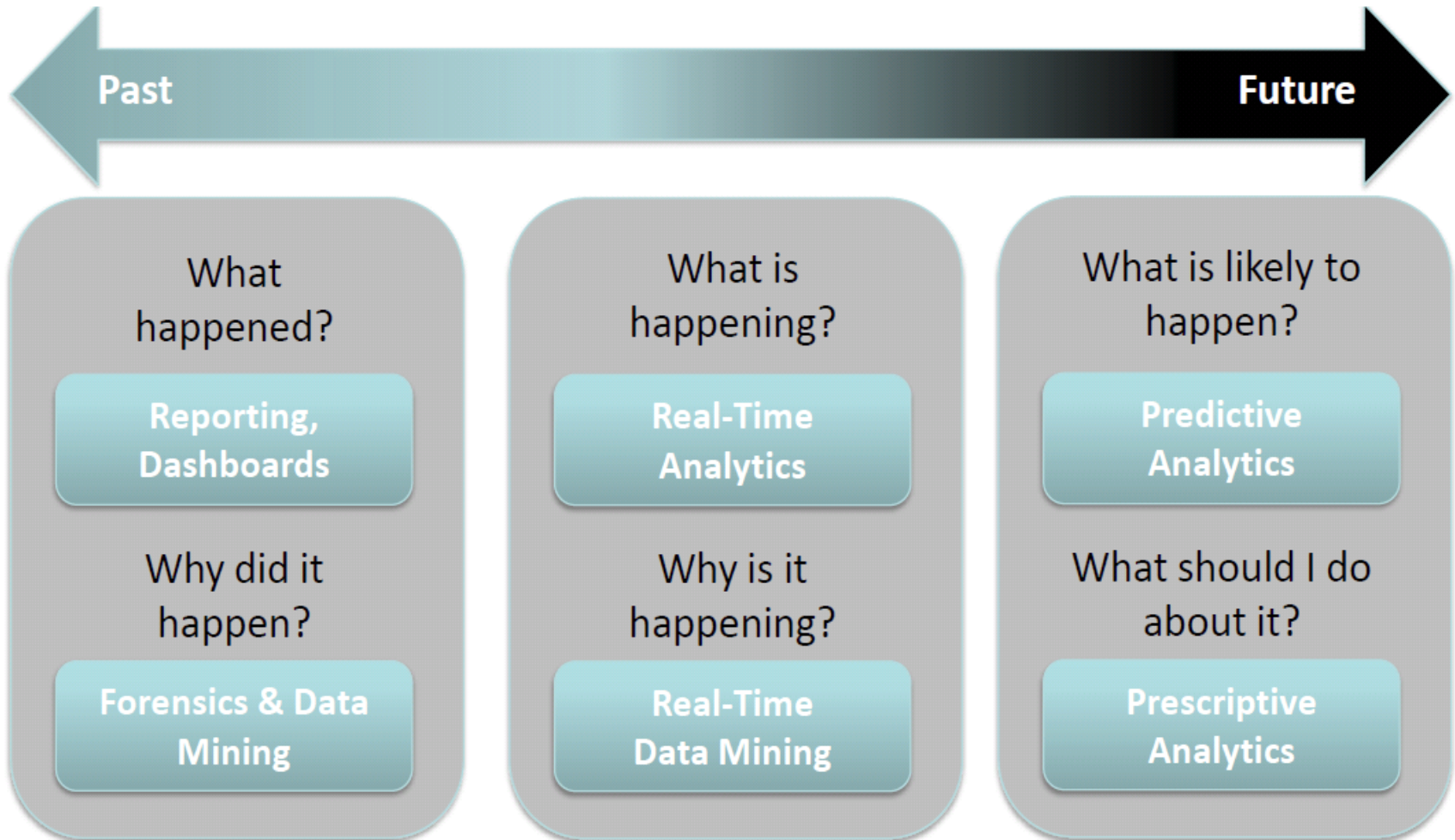
- Video data, audio data



A yellow measuring tape is shown, partially unrolled, with a metal blade visible on the left. The tape has markings for inches, with the numbers 1, 2, and 3 clearly visible. The text "FIBER-GLASS" is printed on the tape. The background is a dark, solid color.

**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!

www.wipro.com

BIG DATA

Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze.

The right use of Big Data allows analysts to spot trends and gives niche insights that help create value and innovation much faster than conventional methods.

The "three V's", i.e the Volume, Variety and Velocity of the data coming in is what creates the challenge.

VOLUME



VARIETY



PEOPLE TO PEOPLE

NETIZENS, VIRTUAL COMMUNITIES, SOCIAL NETWORKS, WEB LOGS...



PEOPLE TO MACHINE

ARCHIVES, MEDICAL DEVICES, DIGITAL TV, E-COMMERCE, SMART CARDS, BANK CARDS, COMPUTERS, MOBILES...



MACHINE TO MACHINE

SENSORS, GPS DEVICES, BAR CODE SCANNERS, SURVEILLANCE CAMERAS, SCIENTIFIC RESEARCH...

VELOCITY



2.9 MILLION

EMAILS SENT EVERY SECOND



20 HOURS

OF VIDEO UPLOADED EVERY MIN



50 MILLION

TWEETS PER DAY

VALUE



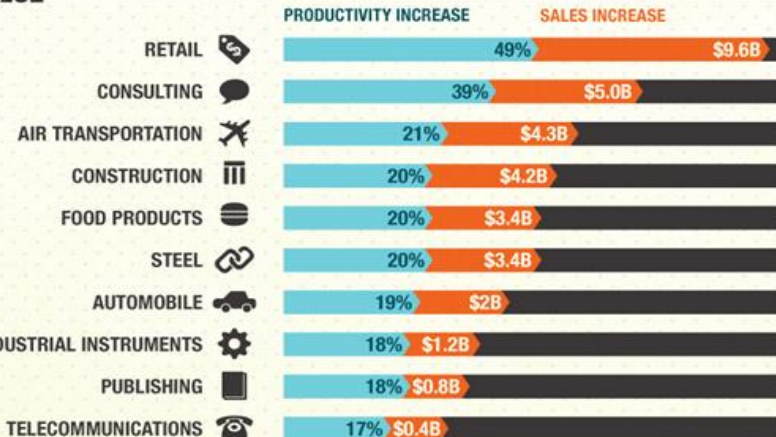
57.6% OF ORGANIZATIONS SURVEYED SAY THAT BIG DATA IS A CHALLENGE



72.7% CONSIDER DRIVING OPERATIONAL EFFICIENCIES TO BE THE BIGGEST BENEFIT OF A BIG DATA STRATEGY



50% SAY THAT BIG DATA HELPS IN BETTER MEETING CONSUMER DEMAND AND FACILITATING GROWTH



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

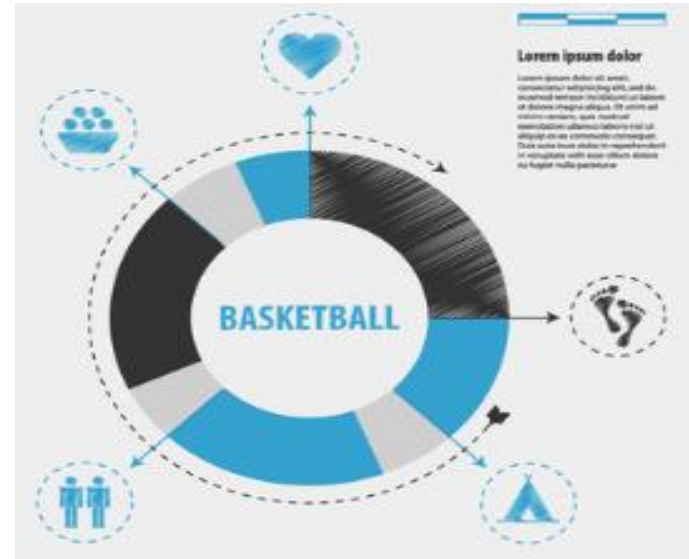
- Pharmaco-Genomics
- Bio-Informatics
- Pharmaceutical Research
- Clinical Outcomes Research

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

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

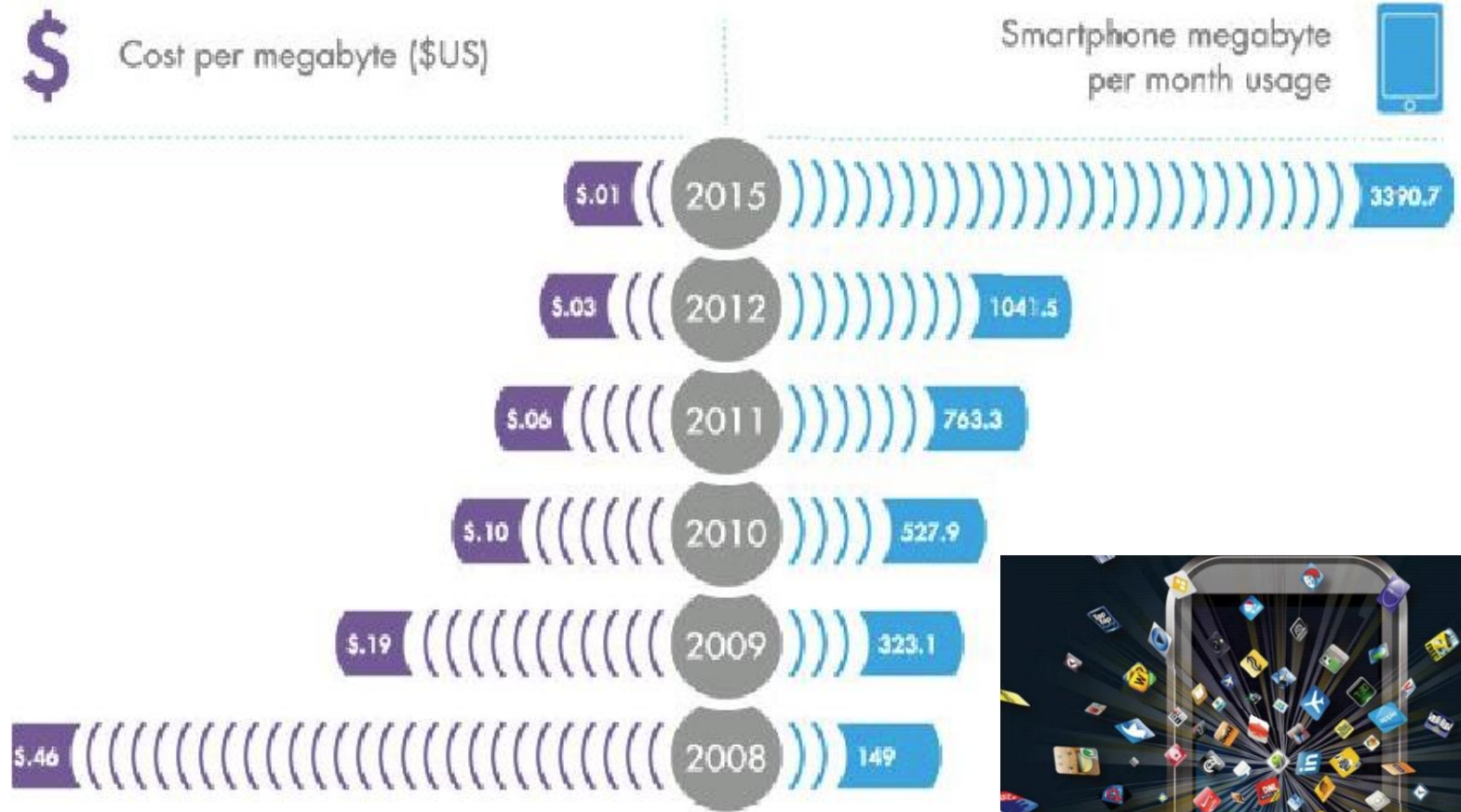
?

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?

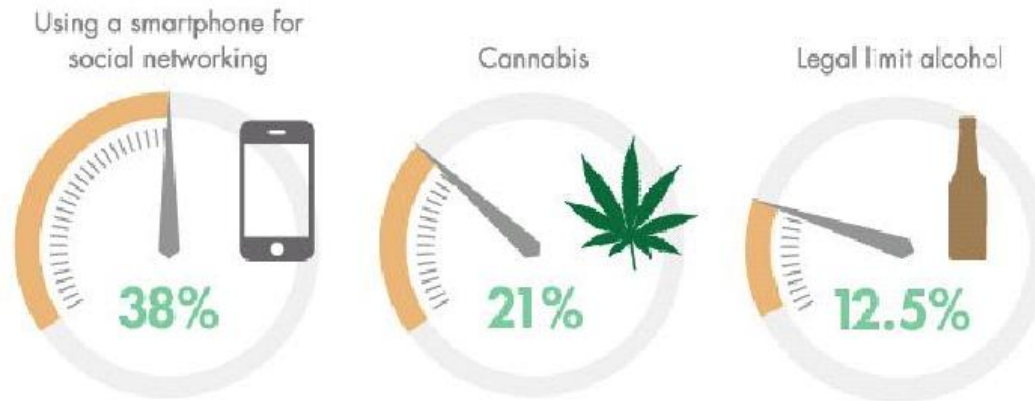


Do you have a healthy relationship with your smart phone?

LIFE DRAIN

Social Media on Your Phone Is worse than Cannabis and Alcohol for driving.

Distractions and how much they slow driving reaction times:



Drivers using a mobile phone are **4x** more likely to crash



Every additional **1,000,000** mobile phone subscriptions



19% rise in distracted driving fatalities

Do you have a healthy relationship with your smart phone?

EMOTIONAL DRAIN



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 your history?



EACH OF US NOW LEAVES A TRAIL OF DIGITAL EXHAUST, AN INFINITE STREAM OF PHONE RECORDS, TEXTS, BROWSER HISTORIES, GPS DATA, AND OTHER INFORMATION, THAT WILL LIVE ON FOREVER.

What does your Social 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?



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

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?

Games



Balloon Brigade

Fill colorful water balloons with water and lob them onto a legion of fiery imps



Wasabi Waiter

Run an upscale sushi bar and serve your customers the best



Crazy Maze

Play Crazy Maze and try to find your way through mind-boggling maze!

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



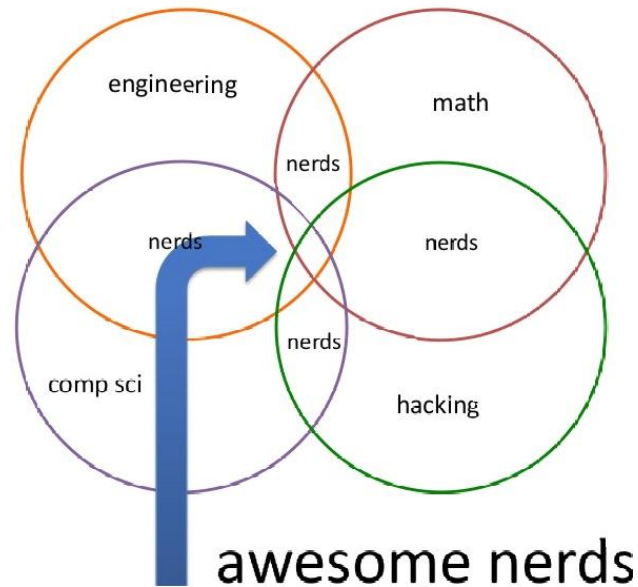
HUMAN BODY is the **CONTROLLER**



ITUNES

Data Scientist... Nerd Talent Shortage

A new breed of people:
Data scientists



By 2018, there will be a demand of 440K-490K Data Scientists in the US, leaving a shortage of 140K-190K positions.



Real World Examples.....

NFL – Analytics at work

<https://www.youtube.com/watch?v=AbYW1kPfnvs&list=PLURx39vvOvhCl4deCOJhhYQ3cdp8qoc9>

(7:55 – 18:00)

<https://www.youtube.com/watch?v=AbYW1kPfnvs&list=PLURx39vvOvhCl4deCOJhhYQ3cdp8qoc9#t=475>

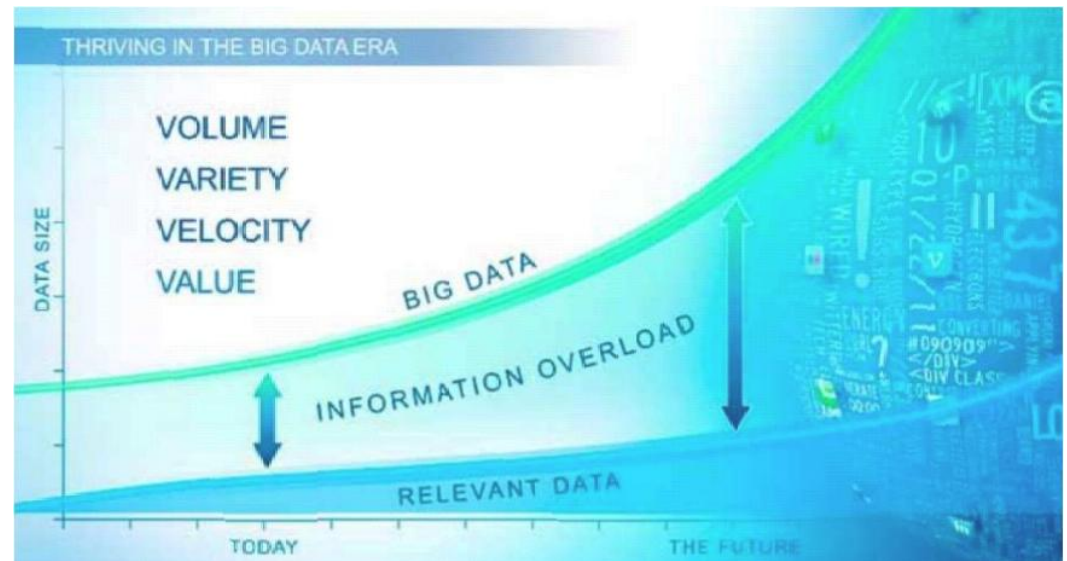
Cognitive Computing / Analytics – Watson

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.

