IMPACT OF BIG DATA ANALYTICS ON BUSINESS, ECONOMY, HEALTH CARE AND SOCIETY

Key Note Address presented on 18th,Nov.15 at the

4th International Conference and Exhibition on Biometrics & Biostatistics , San Antonio, USA

BY

Ajit Kumar Roy

Central Agricultural University, Tripura, India

Contents

- ***Introduction & basics of big data***
- ***Potential value & benefit of big data analytics***
- ***Impact of big data on health care***
- ***Impact of big data on big science***
- ***Impact on finance***
- ***Impact of big data on energy sector***
- ***Big data for global development & Society***
- ***Big data and privacy concerns***

Why do we care about Big Data?

- ***Data is the new oil – we have to learn how to mine it!* Qatar – European Commission Report**
- ***$ 7 trillion economic value in 7 US sectors alone***
- ***New McKinsey 4th factor of production :Land, Labor, Capital, + Data***
- ***Gartner: hundreds of billions of GDP by 2020***
- ***Big Data: new driver for digital economy & society***
- ***An insurance firm with 5 terabytes of data on share drives pays $1.5 m per year***
- ***$90 B annually in sensitive devices***

Big Data Born

- Google, eBay, LinkedIn, and Facebook were built around Big Data from the beginning.
- Big Data could stand alone.
- Big Data analytics could be the only focus of analytics
- Big Data technology architectures could be the only architecture
- No need to reconcile or integrate Big Data with more traditional sources of data and the analytics performed upon them
- No merging Big Data technologies with their traditional IT infrastructures

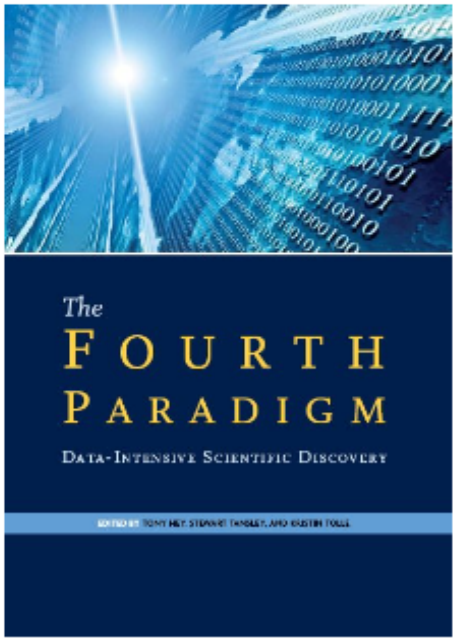
The Data Deluge – Wired 16.07



New kind of science?

The Fourth Paradigm: Data-Intensive Scientific Discovery

Presenting the first broad look at the rapidly emerging field of data-intensive science



Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets.

The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud computing technologies.

In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, the collection of essays expands on the vision of pioneering computer scientist Jim Gray for a new, fourth paradigm of discovery based on data-intensive science and offers insights into how it can be fully realized.

Download

- Full text, low resolution (6 ME
- Full text, high resolution (93 M
- By chapter and essay

Purchase from Amazon.com

- Paperback
- Kindle version

In the news

- Sailing on an Ocean of 0s and
  Magazine)
- A Deluge of Data Shapes a N
  Computing (*New York Times*)

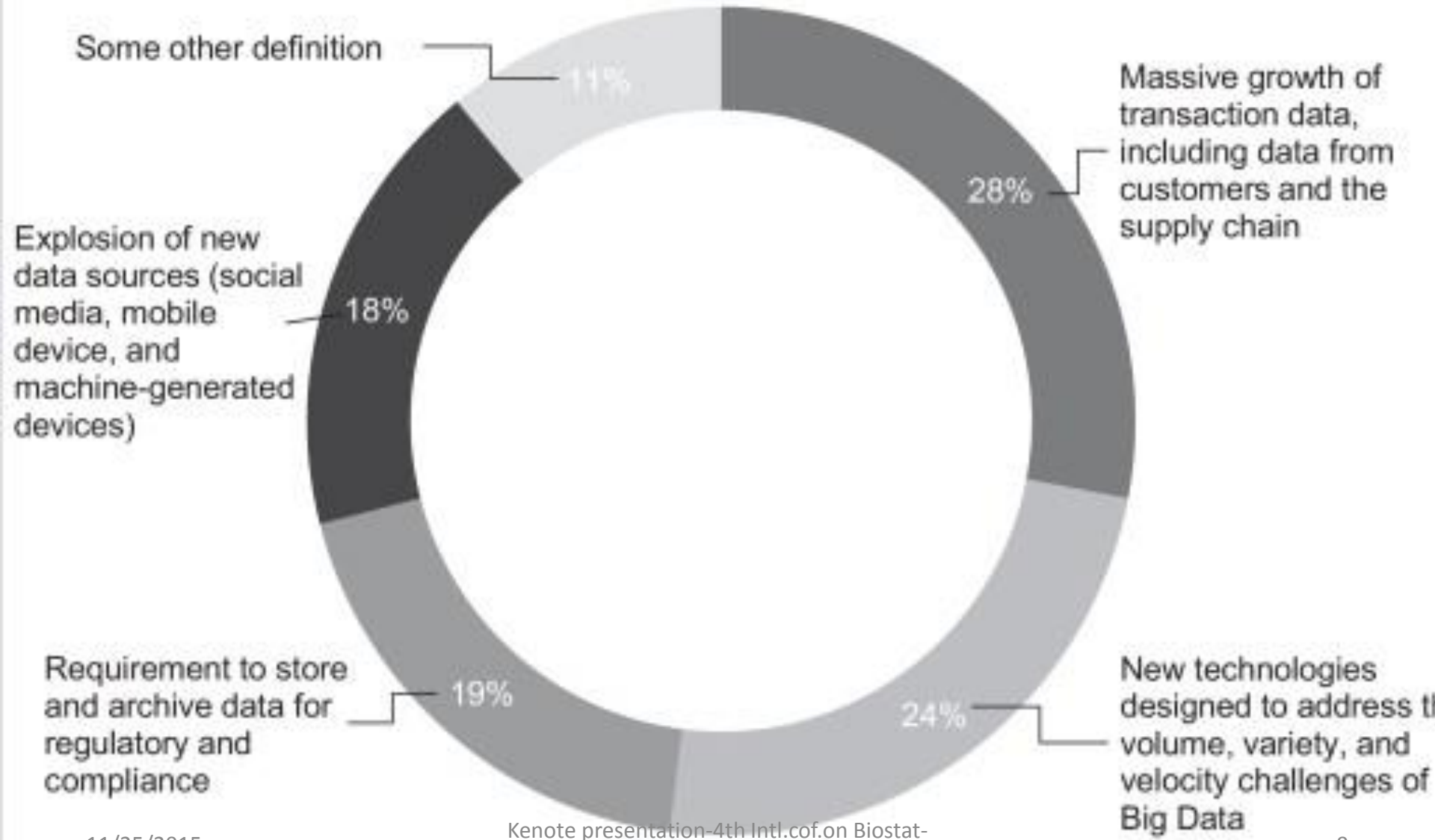
Potential strength of big data

- Frequent and potential in real time or with short lag
- No cost or low cost
- Often geolocated
- Usually with time stamp
- Potential / optional unique stable ID for matching / linking
- Potentially invaluable insights for longitudinal follow-up (including geolocation)
- Social interactions: ego-centric ties and full network
- Might allow to know more or collect info about life history and vital events
- Any individual attributes linkable?

What is Big Data?

"The collection, analysis and generation of insights from a wide variety of data sources in order to be able to improve business performance"

Definitions of big data based on an online survey of 154 global

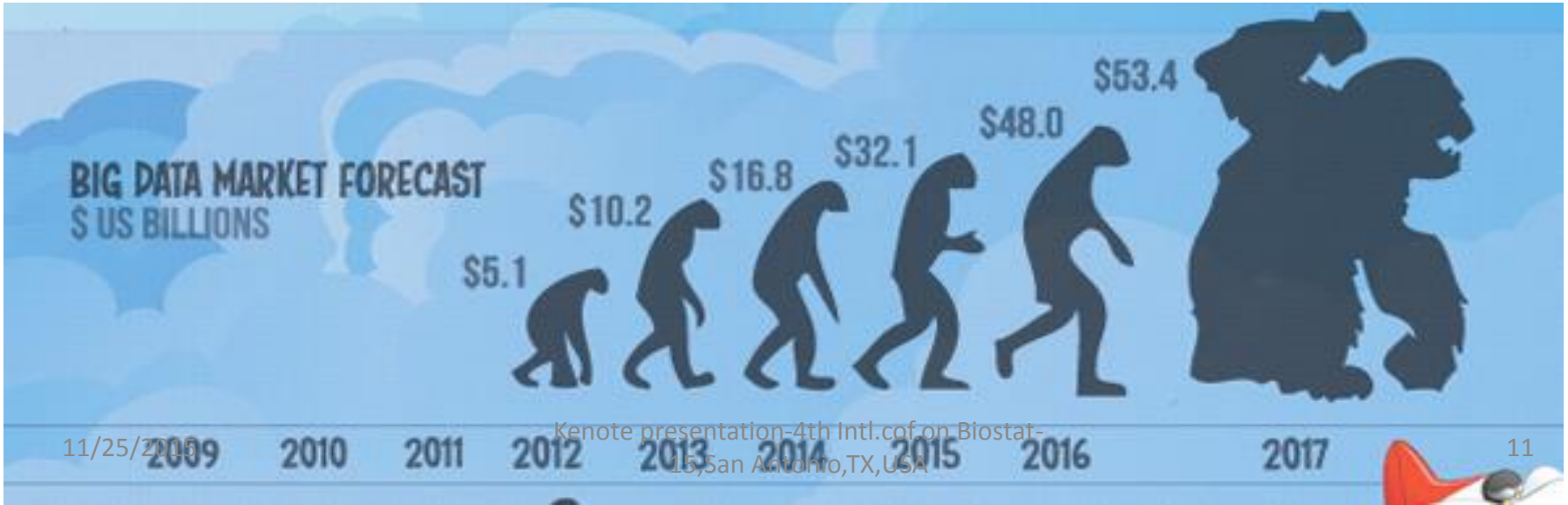


What is BIG DATA?

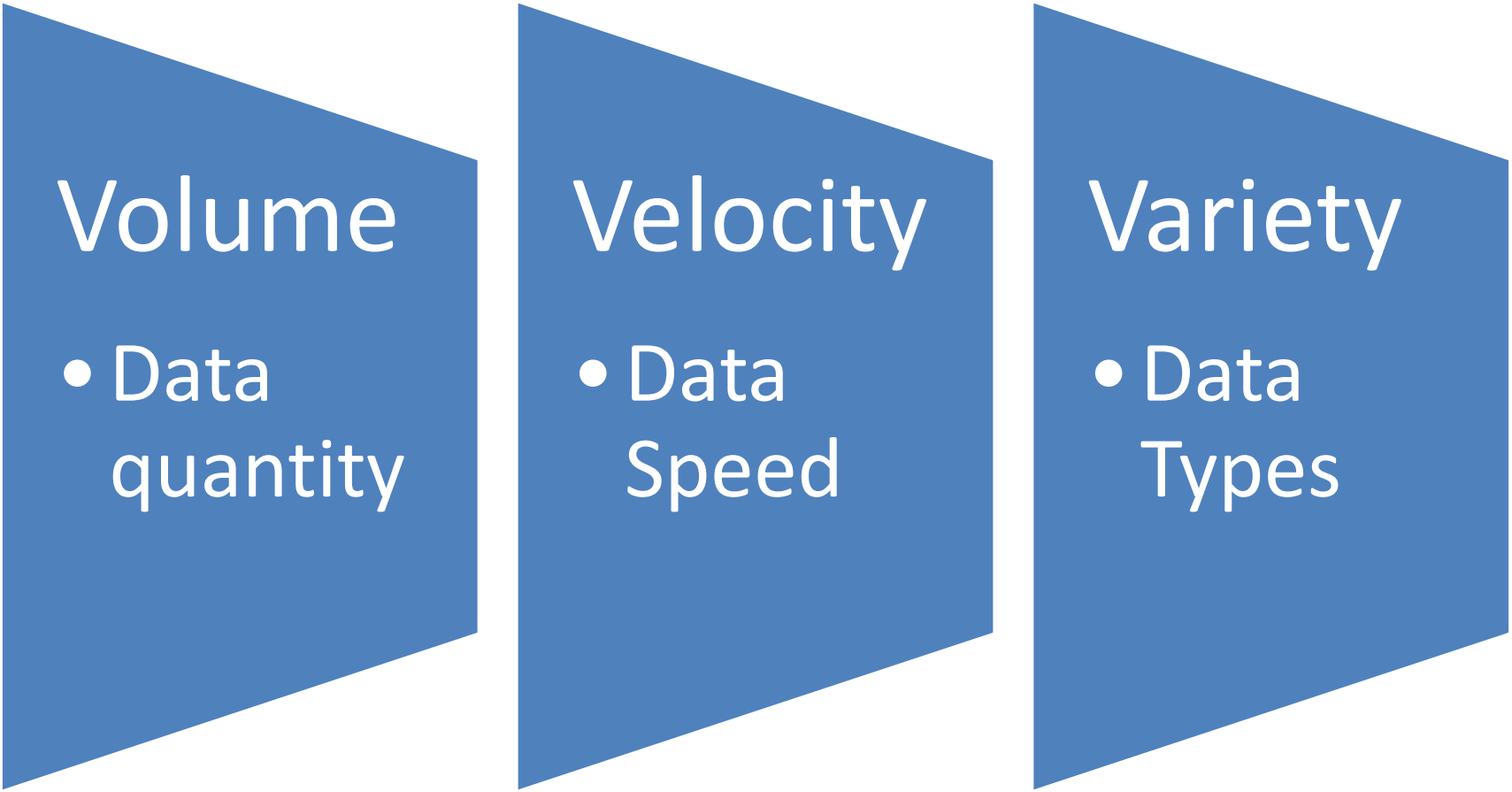
- 'Big Data' is similar to 'small data', but bigger in size
- but having data bigger it requires different approaches:
  - Techniques, tools and architecture
- an aim to solve new problems or old problems in a better way
- Big Data generates value from the storage and processing of very large quantities of digital information that cannot be analyzed with traditional computing techniques.

What is BIG DATA

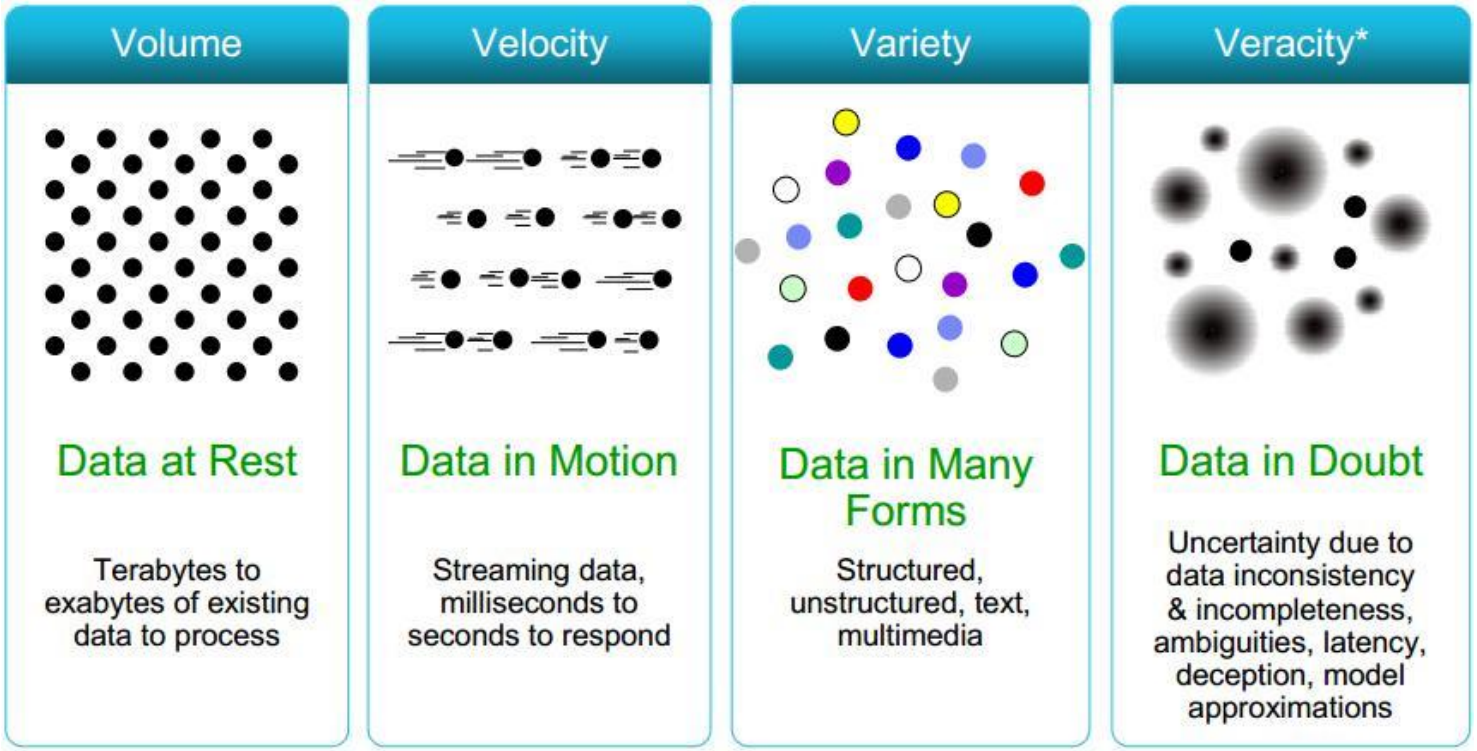
- Wal-Mart handles more than 1 million customer transactions every hour.
- Face book handles 40 billion photos from its user base.
- Decoding the human genome originally took 10 years to process; now it can be achieved in one week.



Three Characteristics of Big Data V3s

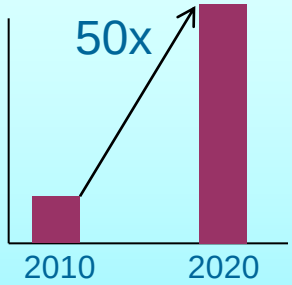


Some Make it 4V's



Four Characteristics of Big Data

Cost efficiently processing the growing **Volume**



35 ZB

Responding to the increasing **Velocity**



30 Billion RFID sensors and counting

Collectively Analyzing the broadening **Variety**



80% of the worlds data is unstructured



Establishing the **Veracity** of big data sources

1 in 3 business leaders don't trust the information they use to make decisions

1st Character of Big Data
Volume

•A typical PC might have had 10 gigabytes of storage in 2000.

•Today, Face book ingests 500 terabytes of new data every day.

•Boeing 737 will generate 240 terabytes of flight data during a single flight across the US.

• The smart phones, the data they create and consume; sensors embedded into everyday objects will soon result in billions of new, constantly-updated data feeds containing environmental, location, and other information, including video.

2nd Character of Big Data Velocity

- Clickstreams and ad impressions capture user behavior at millions of events per second
- high-frequency stock trading algorithms reflect market changes within microseconds
- machine to machine processes exchange data between billions of devices
- infrastructure and sensors generate massive log data in real-time
- on-line gaming systems support millions of concurrent users, each producing multiple inputs per second.

3rd Character of Big Data
Variety

- Big Data isn't just numbers, dates, and strings. Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media.

- Traditional database systems were designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure.

- Big Data analysis includes different types of data

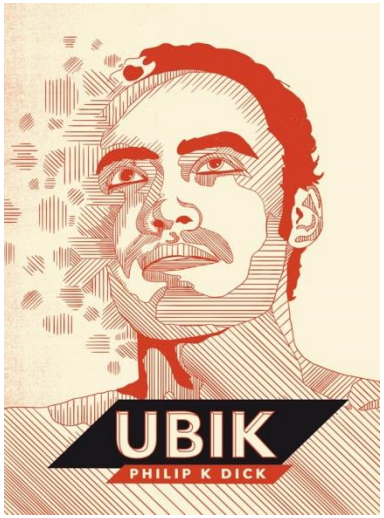
Sources of Big Data

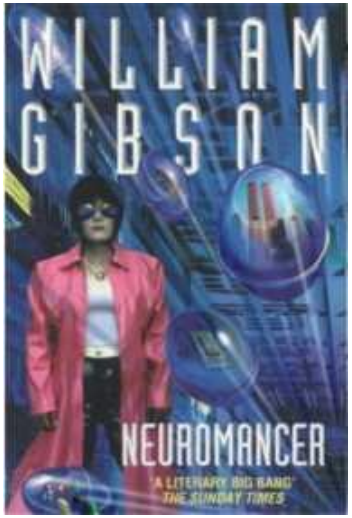
Social media, web analytics, log files, sensors, and such like all provide valuable information, while the cost of it solutions continues to drop and computer-processing power is increasing.

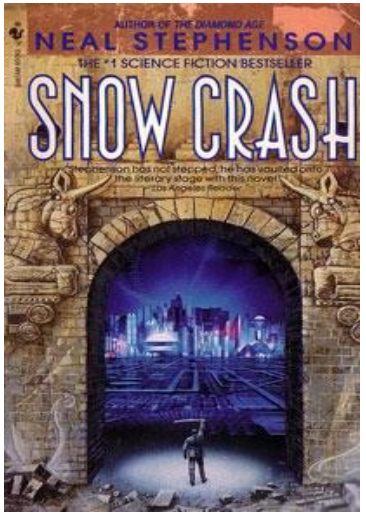
- The Internet of Things is also becoming an increasingly rich source of data. At this moment, there are 13 billion devices connected to the Internet and that will be 50 billion in 2020 says Cisco CTO Padmasree Warrior.
- IDC expects more than **1 billion sensors** to be connected to the Internet by that time. All the accompanying data flows can supply interesting insights that can aid better business decisions.

*Gartner: "IoT Installed Base Will Grow to **26 Billion Units** By 2020." That number might be too low.*

- Every mobile
- Every auto
- Every door
- Every room on every
- Every part,







- **Every sensor in every device … in every bed, chair or bracelet ... in every home, office, building or hospital room … in every city and village ... on Earth ...**

Storing Big Data

❖**Analyzing your data characteristics**

- Selecting data sources for analysis
- Eliminating redundant data
- Establishing the role of NoSQL

❖**Overview of Big Data stores**

- Data models: key value, graph, document, column-family
- Hadoop Distributed File System
- HBase
- Hive

Processing Big Data

❖ **Integrating disparate data stores**

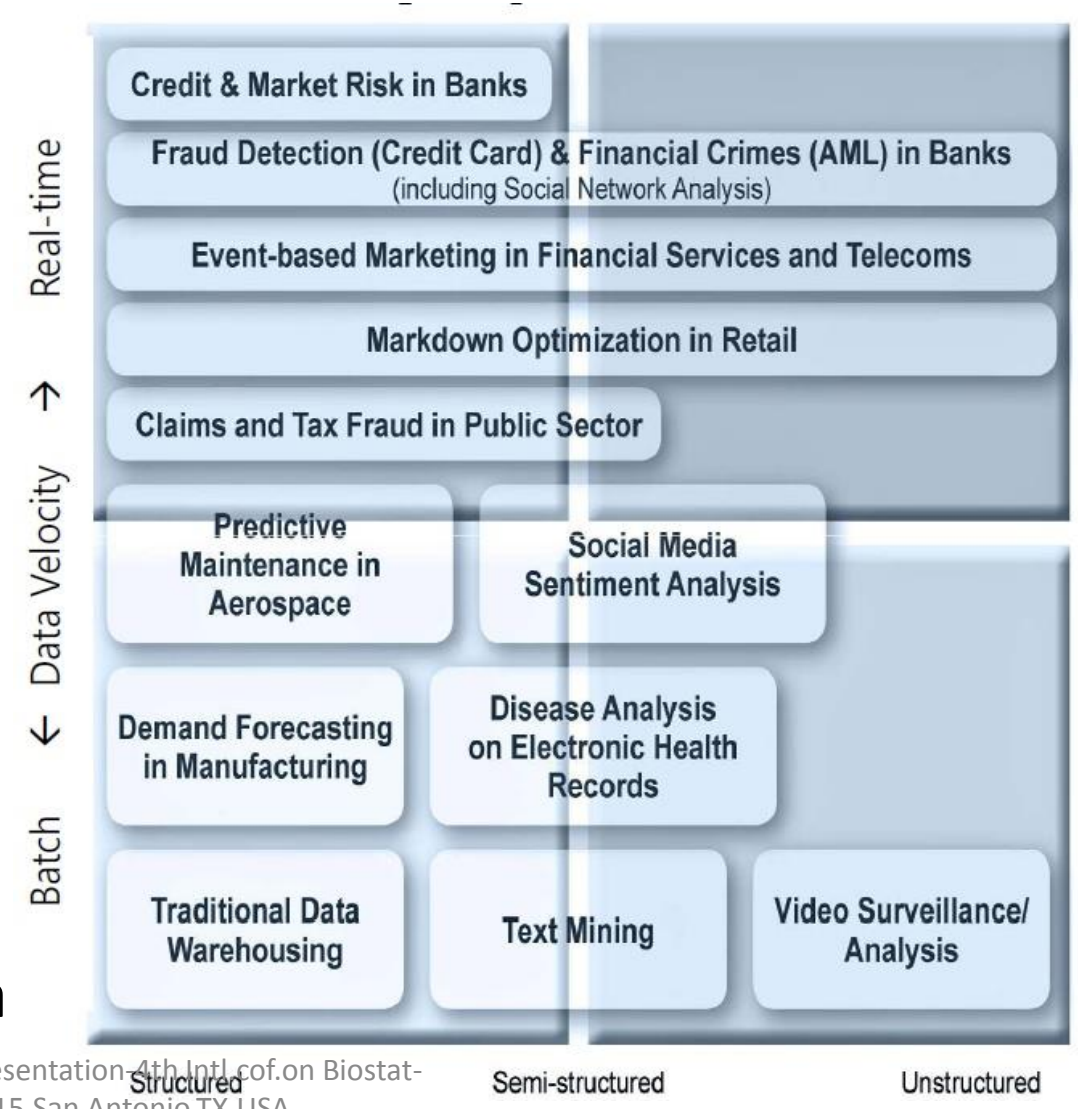
- Mapping data to the programming framework
- Connecting and extracting data from storage
- Transforming data for processing
- Subdividing data in preparation for Hadoop MapReduce

❖ **Employing Hadoop MapReduce**

- Creating the components of Hadoop MapReduce jobs
- Distributing data processing across server farms
- Executing Hadoop MapReduce jobs
- Monitoring the progress of job flows

The Structure of Big Data

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



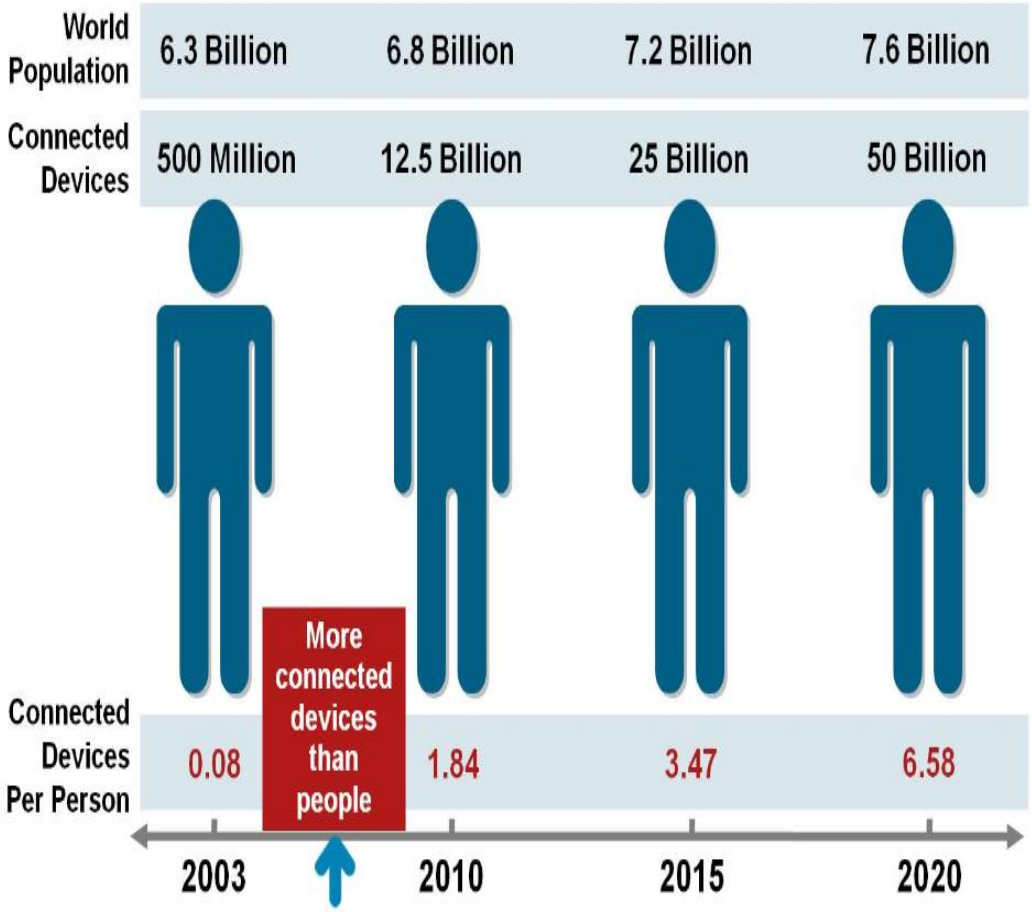
Why Big Data

- Growth of Big Data needs
  - Increase of storage capacities
  - Increase of processing power
  - Availability of data(different data types)
  - Every day we create 2.5 quintillion bytes of data; 90% of the data in the world today has been created in the last two years alone

Why Big Data

- FB generates 10TB daily
- Twitter generates 7TB of data Daily
- IBM claims 90% of today's stored data was generated in just the last two years.

Figure 1. The Internet of Things Was "Born" Between 2008 and 2009



Source: Cisco IBSG, April 2011

How Is Big Data Different?

1) Automatically generated by a machine
   (e.g. Sensor embedded in an engine)

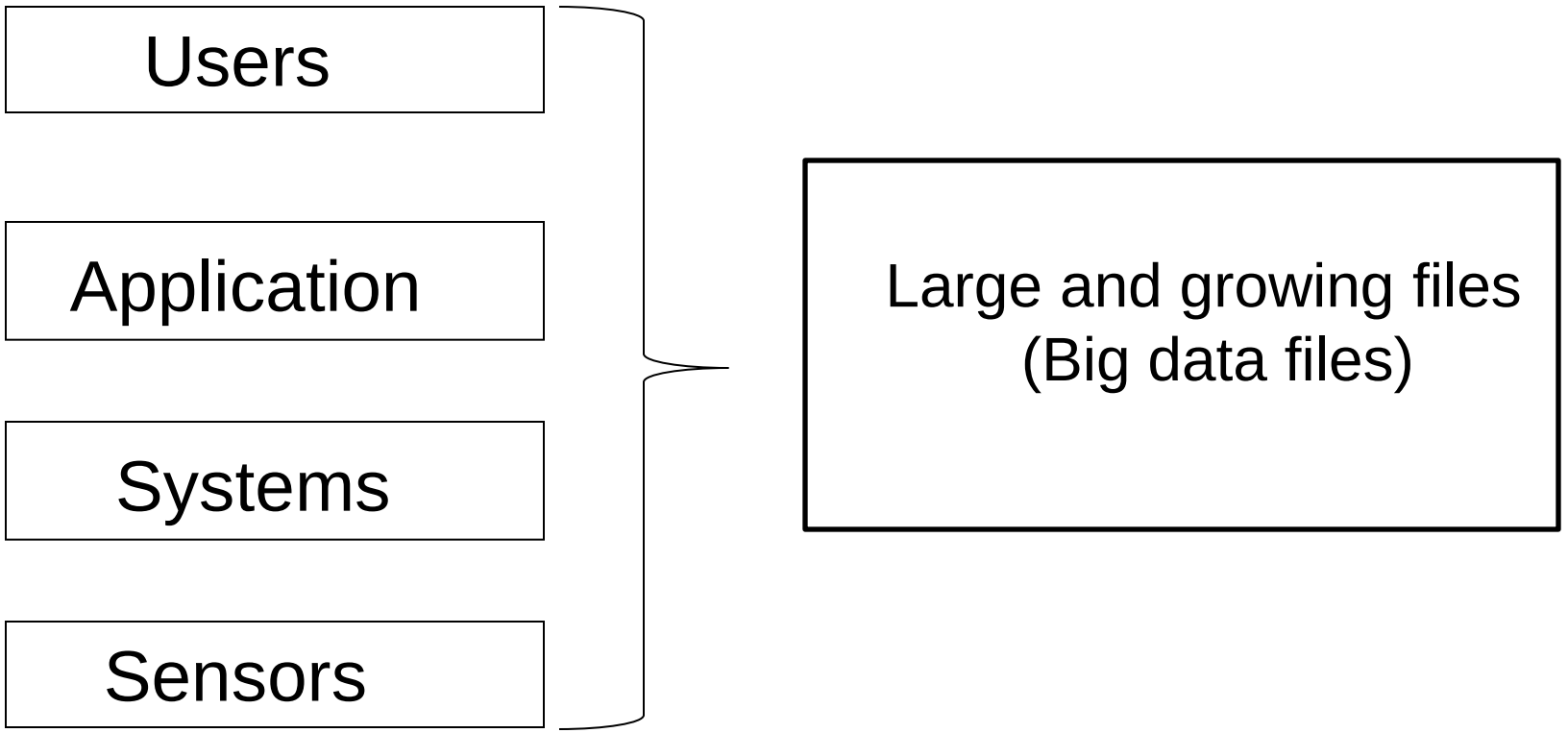
2) Typically an entirely new source of data
   (e.g. Use of the internet)



3) Not designed to be friendly
   (e.g. Text streams)

4) May not have much values
   - Need to focus on the important part

Big Data sources



What is collecting all these data?

Web Browsers

Microsoft's Internet Explorer



Mozilla's FireFox
(Non-profit foundation, used to be Netscape)



Google's Chrome



Apple's Safari



Time-Warner's AOL Explorer



Search Engines

Google's



Microsoft's



Yahoo's



IAC Search's



What is collecting all these data?

Smartphones & Apps

Apple's iPhone
(Apple O/S)



Samsung, HTC.
Nokia, Motorola
(Android O/S)



RIM Corp's Blackberry
(BlackBerry O/S)



Tablet Computers & Apps

Apple's iPad



Samsung's Galaxy



Amazon's Kindle Fire



What is collecting all these data?

Games Boxes and GPS Systems









Internet Service Providers









What is collecting all these data?

HDTV's and Blu-Ray Players with built-in Internet connectivity





Movie Rental Sites

NETFLIX

redbox

amazon instant video
Connecting

What is collecting all these data?

Hospitals & Other Medical Systems

Pharmacies
Laboratories
Imaging Centers
Emergency Medical Services (EMS)
Hospital Information Systems
Doc-in-a-Box
Electronic Medical Records
Blood Banks
Birth & Death Records

Banking & Phone Systems



verizon

Can you hear me now?
(Heh heh heh!)

Sprint

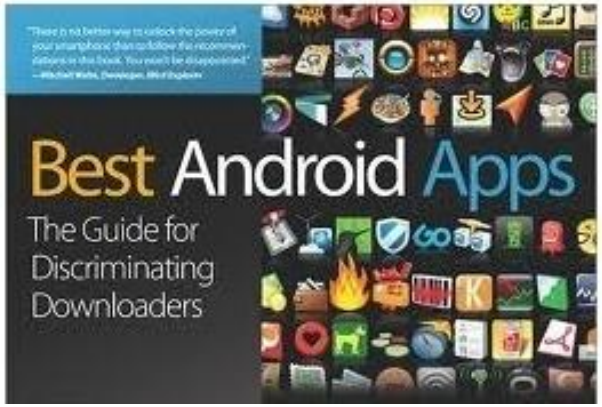
T · ·Mobile·

at&t

What is collecting all these data?

A real pain in the apps!





What are they collecting?

- Restaurant reservations (Open Table)
- Weather in L.A. in 3 days (Weather+)
- Side effects of medications (MedWatcher)
- 3-star hotels in New Orleans (Priceline)
- Which PC should I buy and where (PriceCheck)

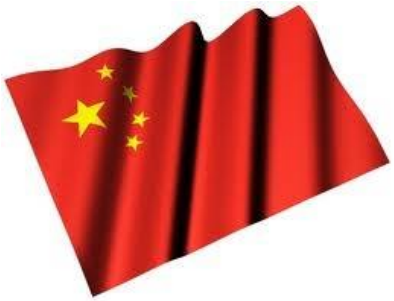
Who is collecting all of this data?

Government Agencies

IRS

Department of the Treasury
Internal Revenue Service





Big Pharmaceutical Companies

Pfizer

BAYER

Roche

MERCK

AstraZeneca

gsk GlaxoSmithKline

(Hey, I didn't say which government!)

Who is collecting all this data?

Consumer Products Companies



Big Box Stores



Who is collecting what?

Credit Card Companies

What data are they getting?

Airline ticket

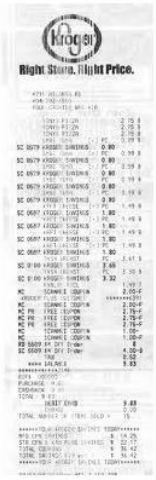


Restaurant check



1	Cab Sauv Corvina	£15.25
1	Cab Sauv Glass	£4.00
20	Net Sales Total	£73.45
1	SUCK MY D[illegible] F[illegible] FACE	£0.00
1	FISH CAKES	£4.95
2	1/2 Wings Starter	£7.90
1	MELON AND PARMA HAM	£3.95
1	Calamari	£4.95
1	Garlic brd starter	£2.25
3	Meatball Starter	£17.85
1	Aub & Feta Starter	£3.95

Grocery Bill



Hotel Bill



Big Data is Everywhere

6 Billion
Flickr Photos



28 Million
Wikipedia Pages

900 Million
Facebook Users

72 Hours a Minute
YouTube


The New York Times
SundayReview

WORLD U.S. N.Y. / REGION BUSINESS TEC

NEWS ANALYSIS

The Age of Big Data

By STEVE LOHR

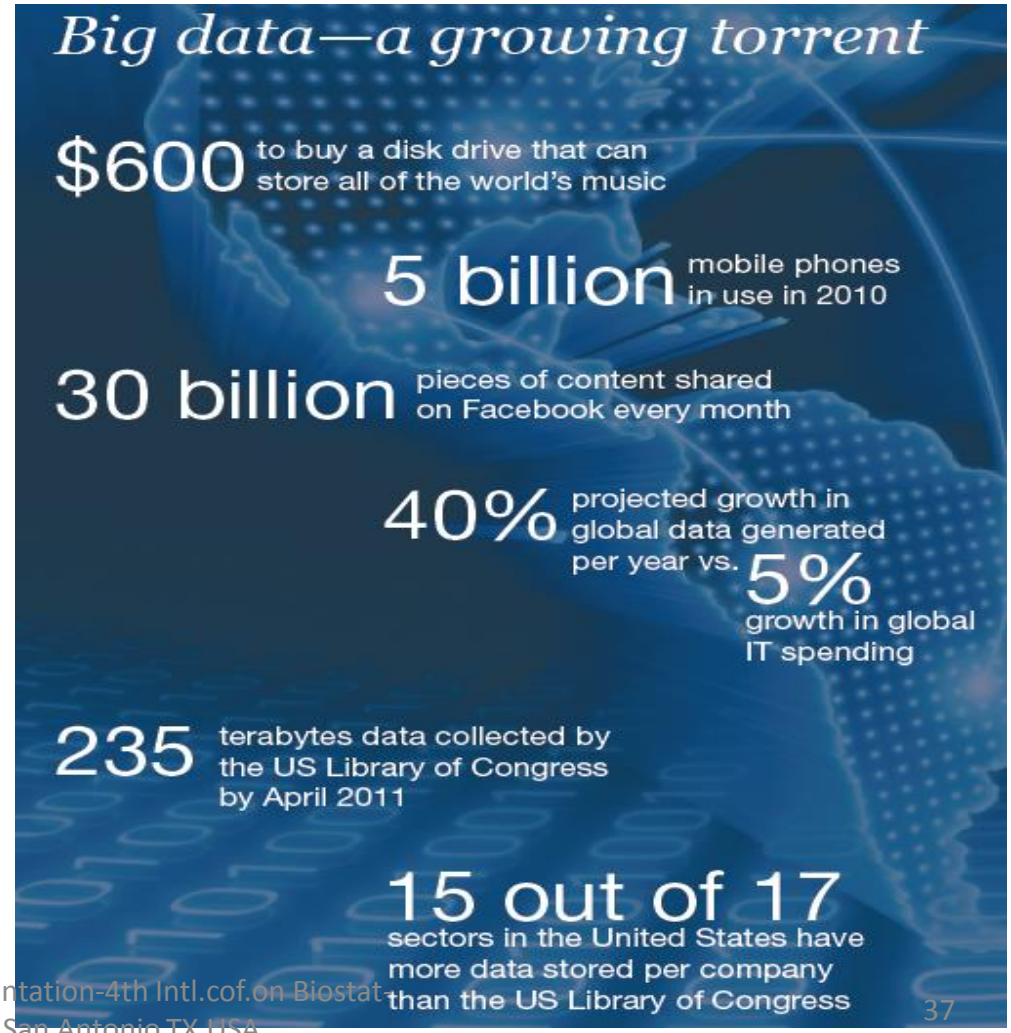
Published: February 11, 2012

"…growing at 50 percent a year…"

"… data a new class of economic asset, like currency or gold."

Data generation points Examples

- Mobile Devices
- Microphones
- Readers/Scanners
- Science facilities
- Programs/ Software
- Social Media
- Cameras



IBM on 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, and Cell phone GPS signals to name a few. This data is big data." –IBM on big data.*

Estimates of Big Data volume

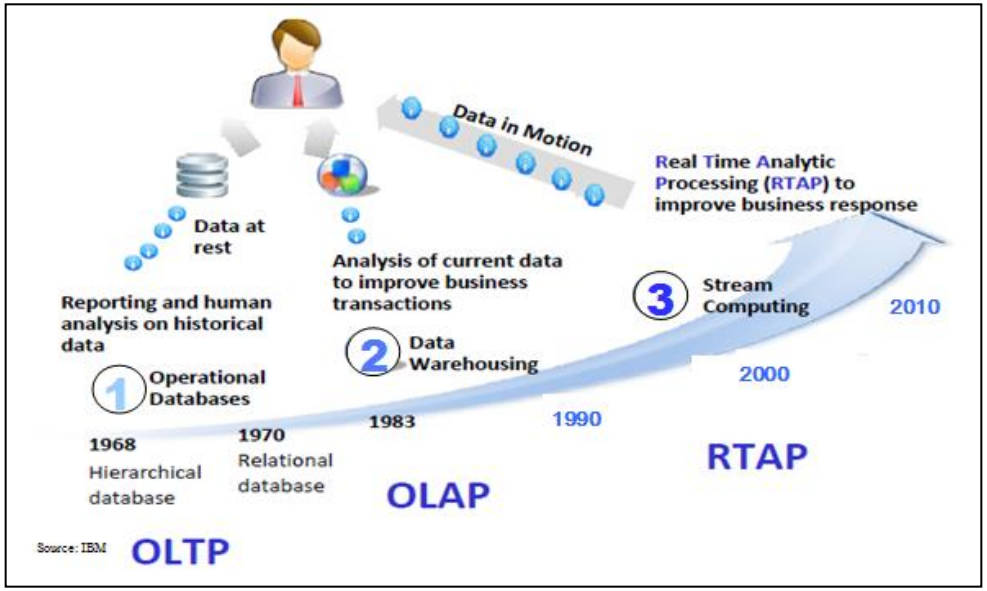
- In 2020, there will be 35 zettabytes of digital data. That represents a stack of DVD's that would reach half way from the Earth to Mars.
- Facebook has 70 petabytes and 2700 multiprocessor nodes.
- The Bing search engine has 150 petabytes and 40,000 nodes.
- The simplest response comes from Forrester Research and is as follows:
- *Big Data: Techniques and Technologies that Make Handling Data at Extreme Scale Economical.*

Let Us Guess the volume of data

In 2011,Hilbert and Lopez estimated that if all the data used in the world today were written to CD-ROMs and the CD-ROMs piled up in a single stack, the stack would stretch all the way from the Earth to the Moon and a quarter of the way back again.

A report by the International Data Corporation in 2010 estimated that by the year 2020 there will be 35 Zettabytes (ZB) of digital data created per annum.

Harnessing Big Data



- **OLTP:** Online Transaction Processing (DBMSs)
- **OLAP:** Online Analytical Processing (Data Warehousing)
- **RTAP:** Real-Time Analytics Processing (Big Data Architecture & technology)

The Model Has Changed…

- **The Model of Generating/Consuming Data has Changed**

Old Model: Few companies are generating data, all others are consuming data



New Model: all of us are generating data, and all of us are consuming data

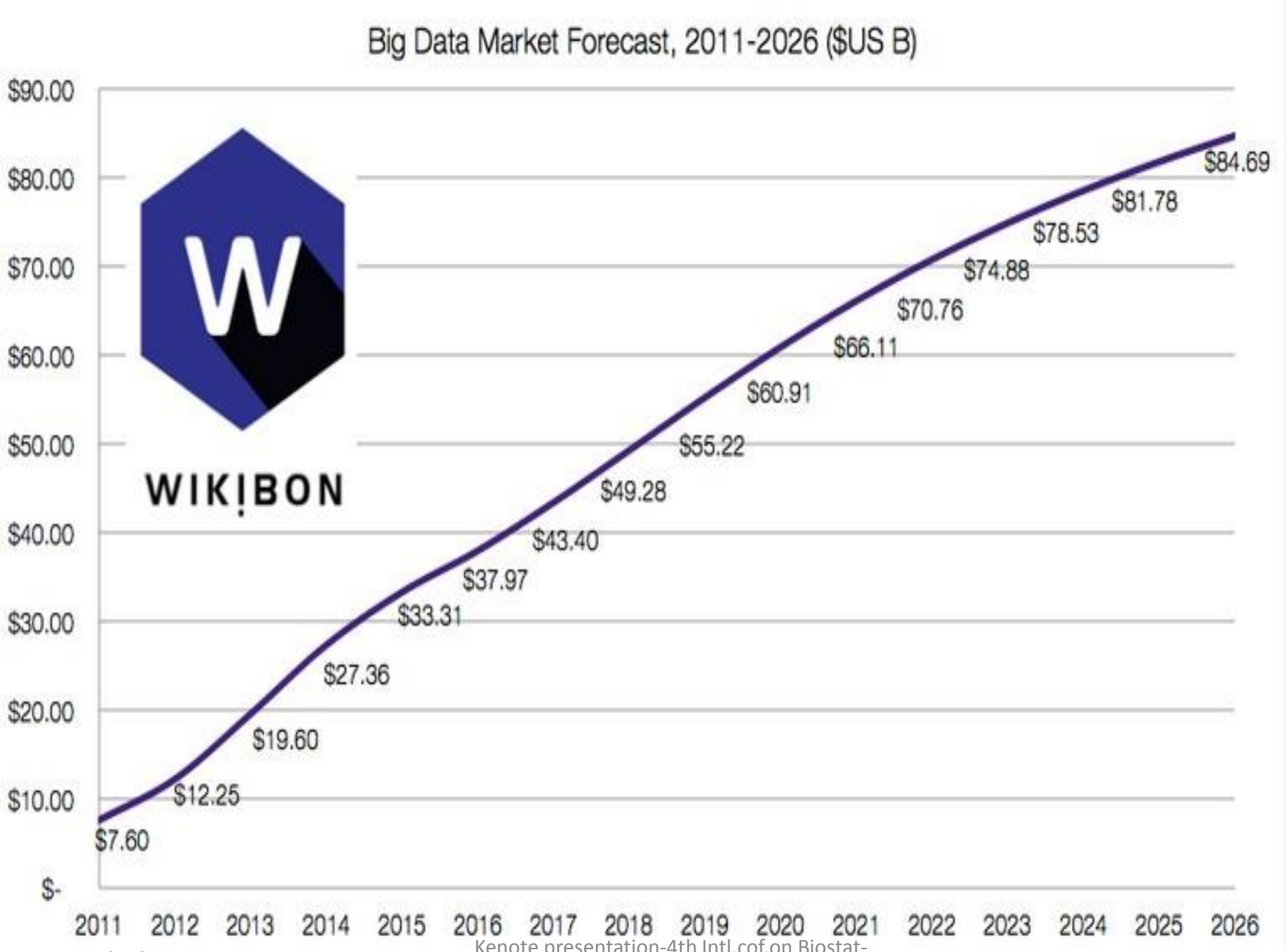


Big Data Analytics

- Examining large amount of data
- Appropriate information
- Identification of hidden patterns, unknown correlations
- Competitive advantage
- Better business decisions: strategic and operational
- Effective marketing, customer satisfaction, increased revenue

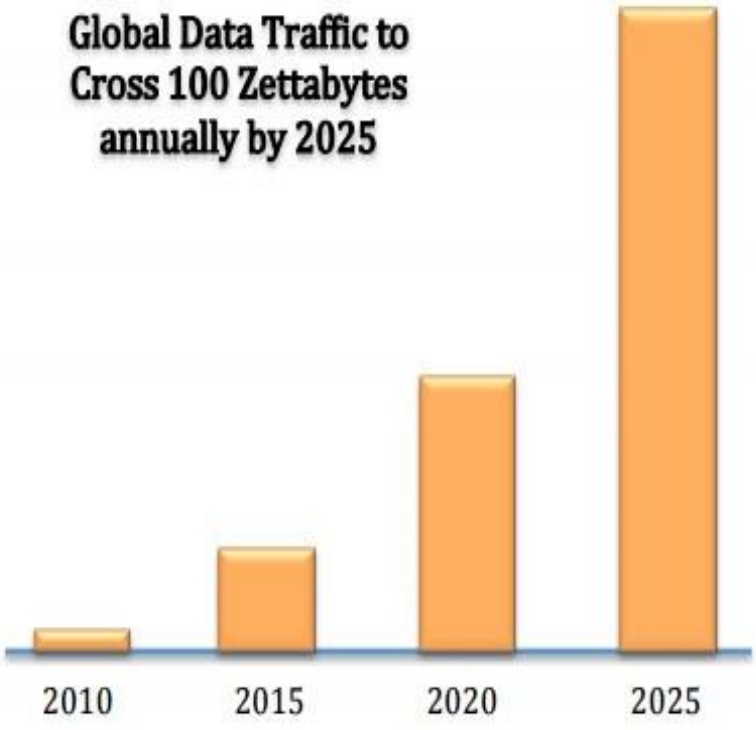
Types of tools used in Big-Data

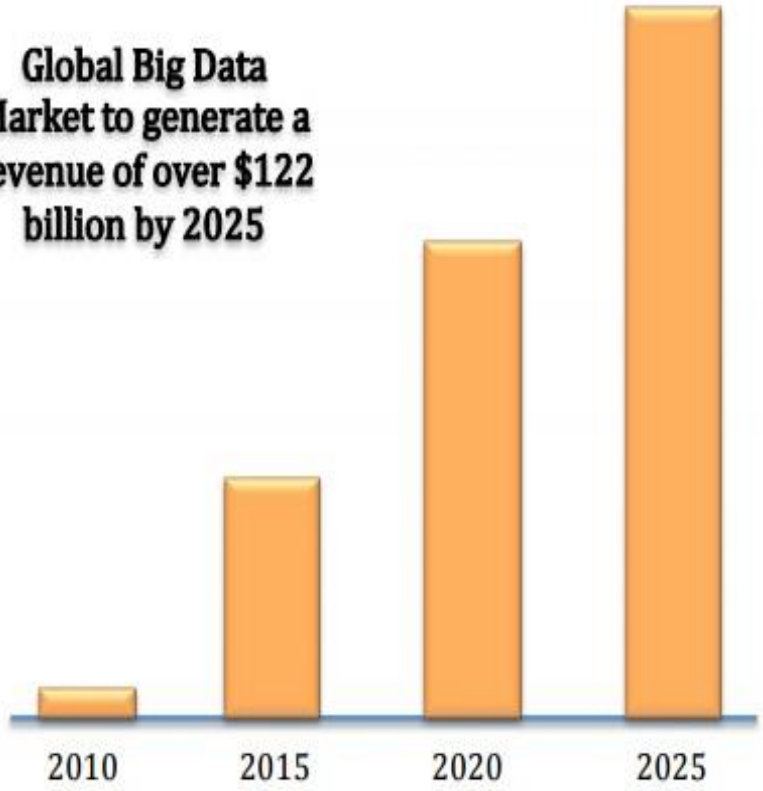
- Where processing is **hosted**?
  - Distributed Servers / Cloud (e.g. Amazon EC2)
- Where data is **stored**?
  - Distributed Storage (e.g. Amazon S3)
- What is the **programming model**?
  - Distributed Processing (e.g. MapReduce)
- How data is **stored & indexed**?
  - High-performance schema-free databases (e.g. MongoDB)
- What operations are performed on data?
  - Analytic / Semantic Processing


Big Data Market Forecast, 2011-2026 ($US B)
$90.00
$80.00
$70.00
$60.00
$50.00
$40.00
$30.00
$20.00
$10.00
$-
WIKIBON
$7.60
$12.25
$19.60
$27.36
$33.31
$37.97
$43.40
$49.28
$55.22
$60.91
$66.11
$70.76
$74.88
$78.53
$81.78
$84.69
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026

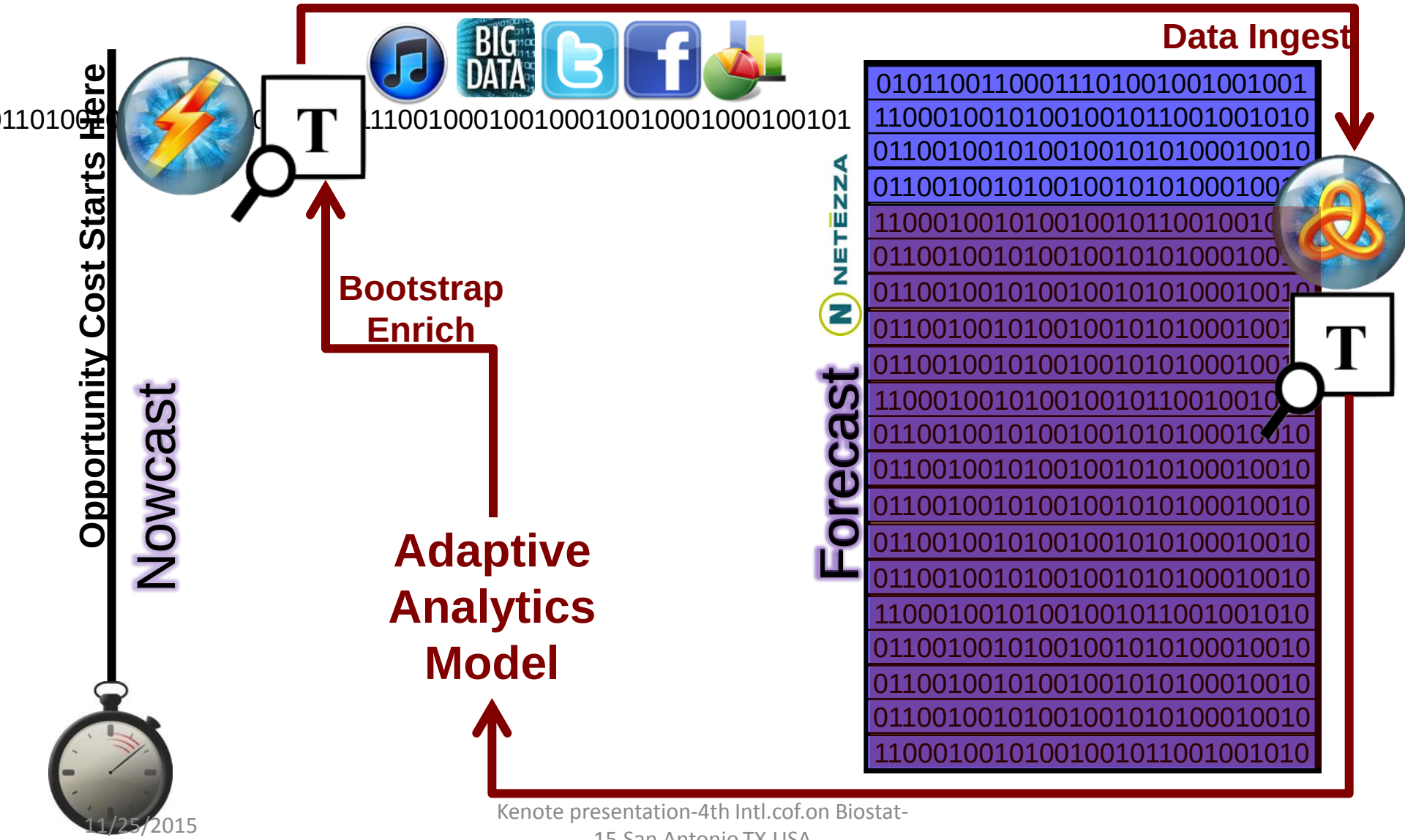
Sub Trend: Big Data Analytics

90% of the data in the world has been created in the last two years alone





Analytic With *Data-In-Motion & Data At Rest*



Application Of Big Data analytics

Smarter Healthcare



Multi-channel sales



Homeland Security



Telecom



Traffic Control



Trading Analytics



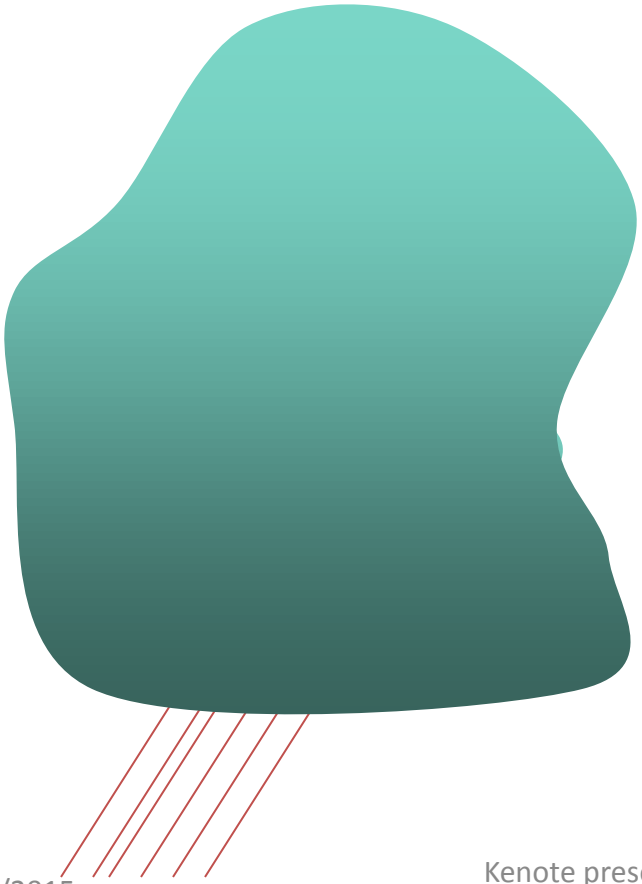
Manufacturing



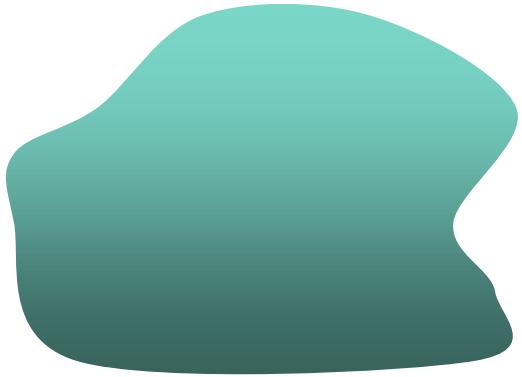
Search Quality



Identifying Storms by Satellite-Nowcasting

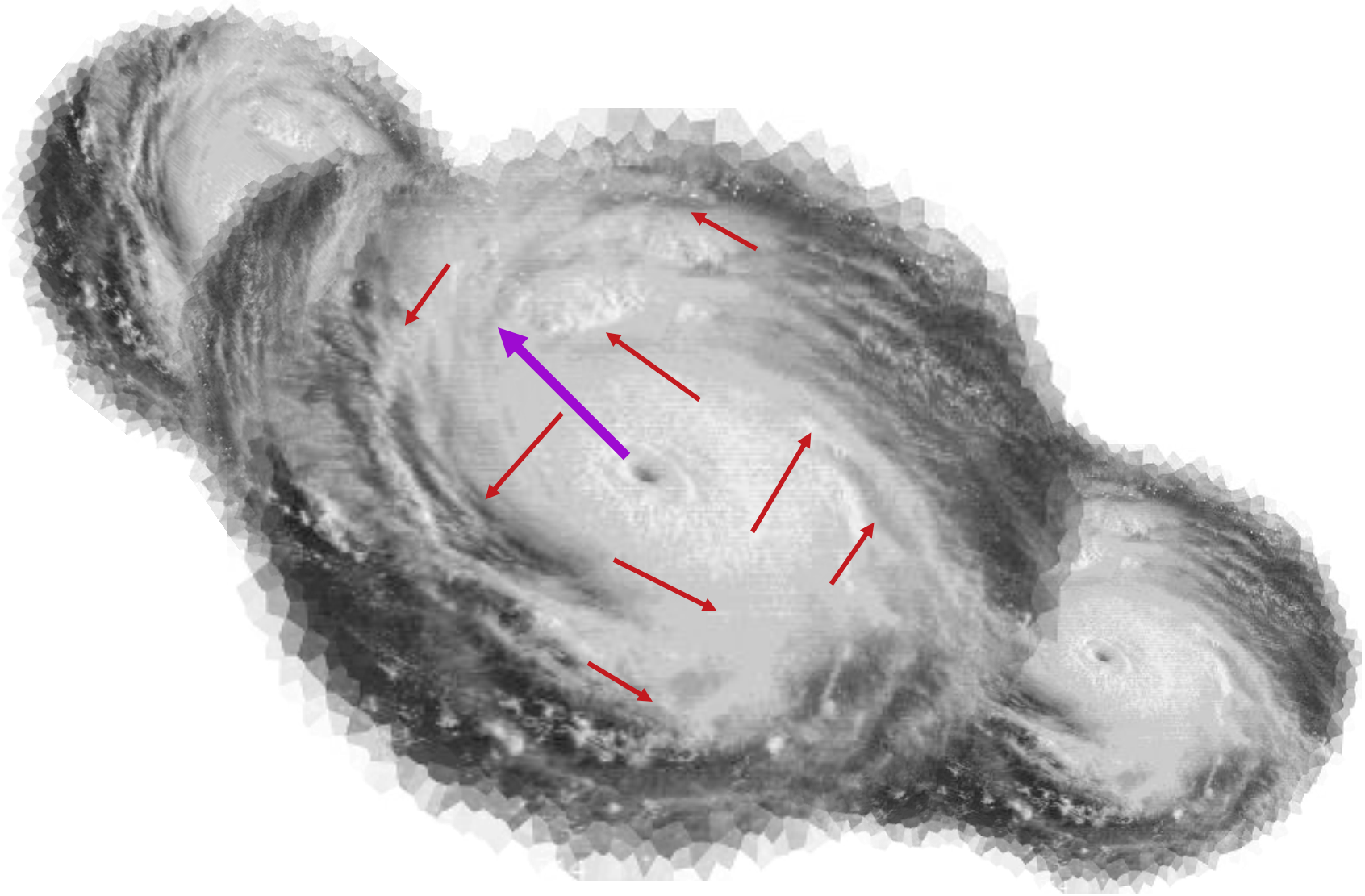


- Rapid growth: vertical and horizontal
- Rounded tops
- High local variability in cloud top altitudes.

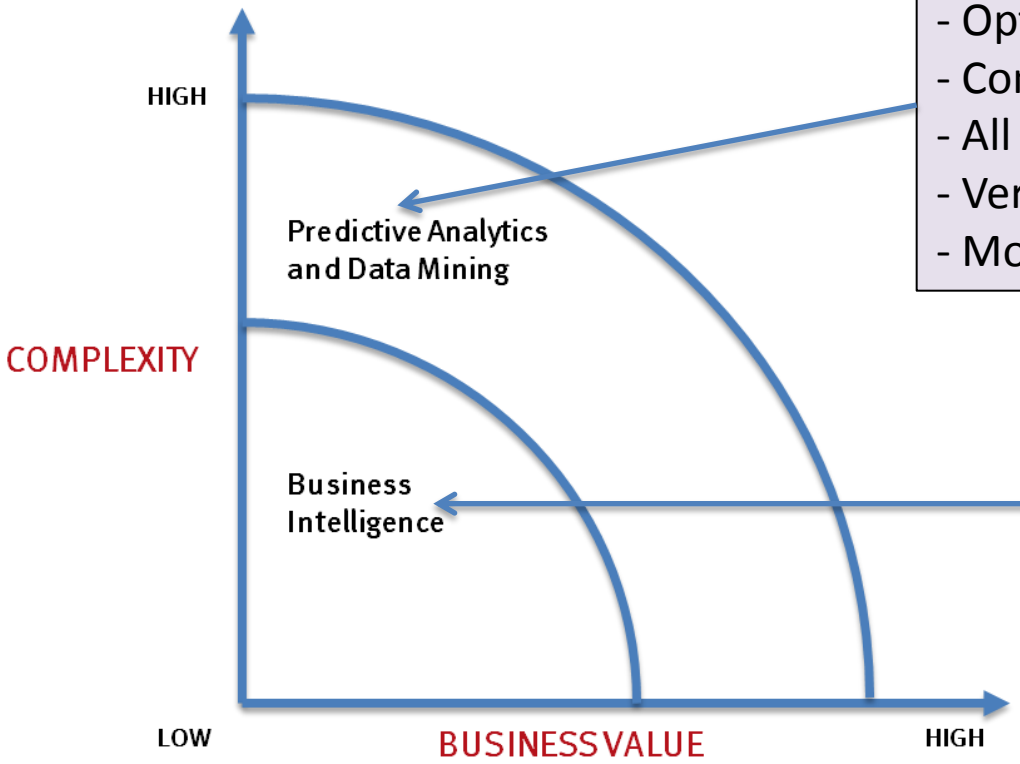


Improving Storm Forecasting 2

Correlated K-means



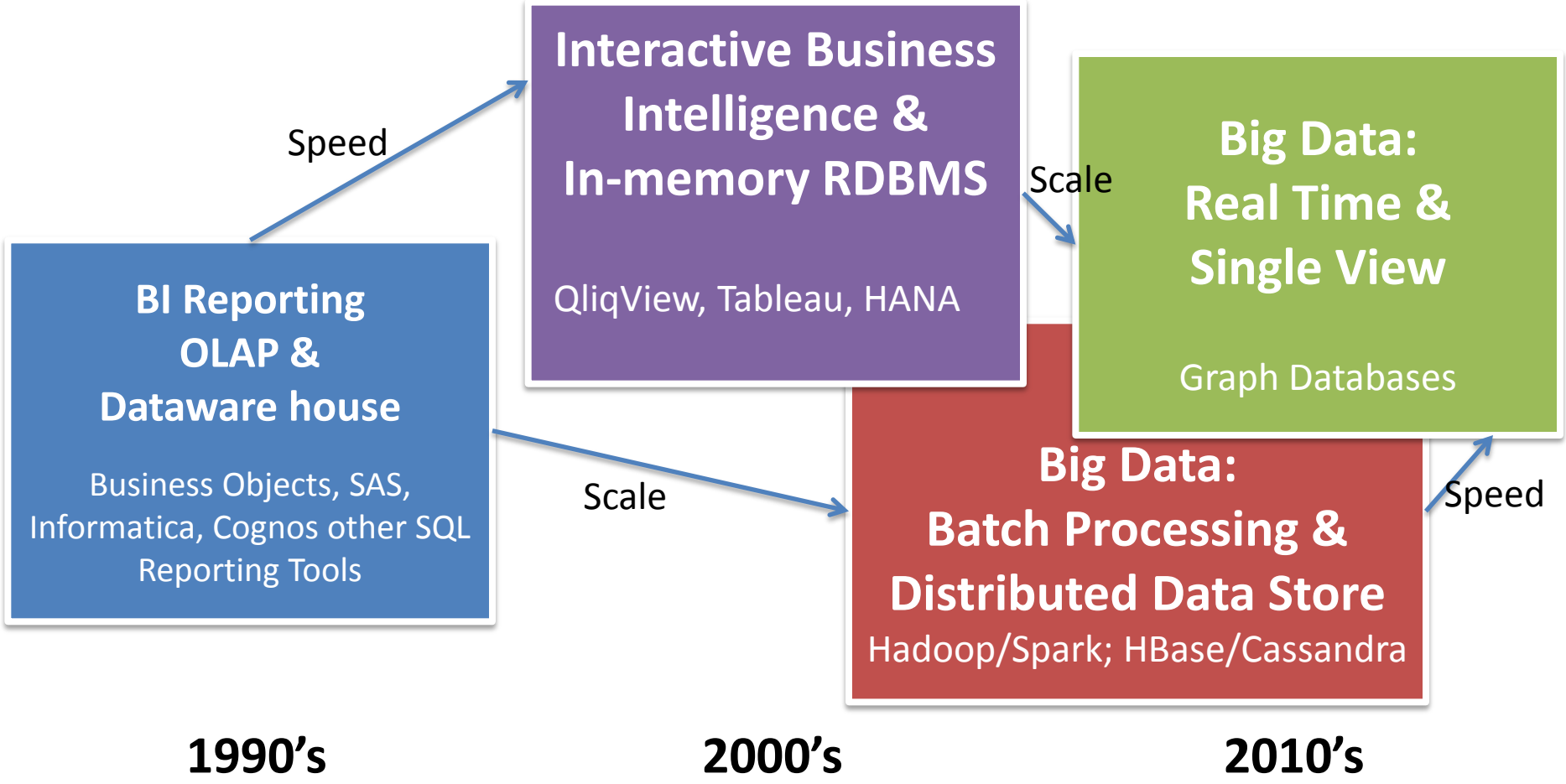
What's driving Big Data



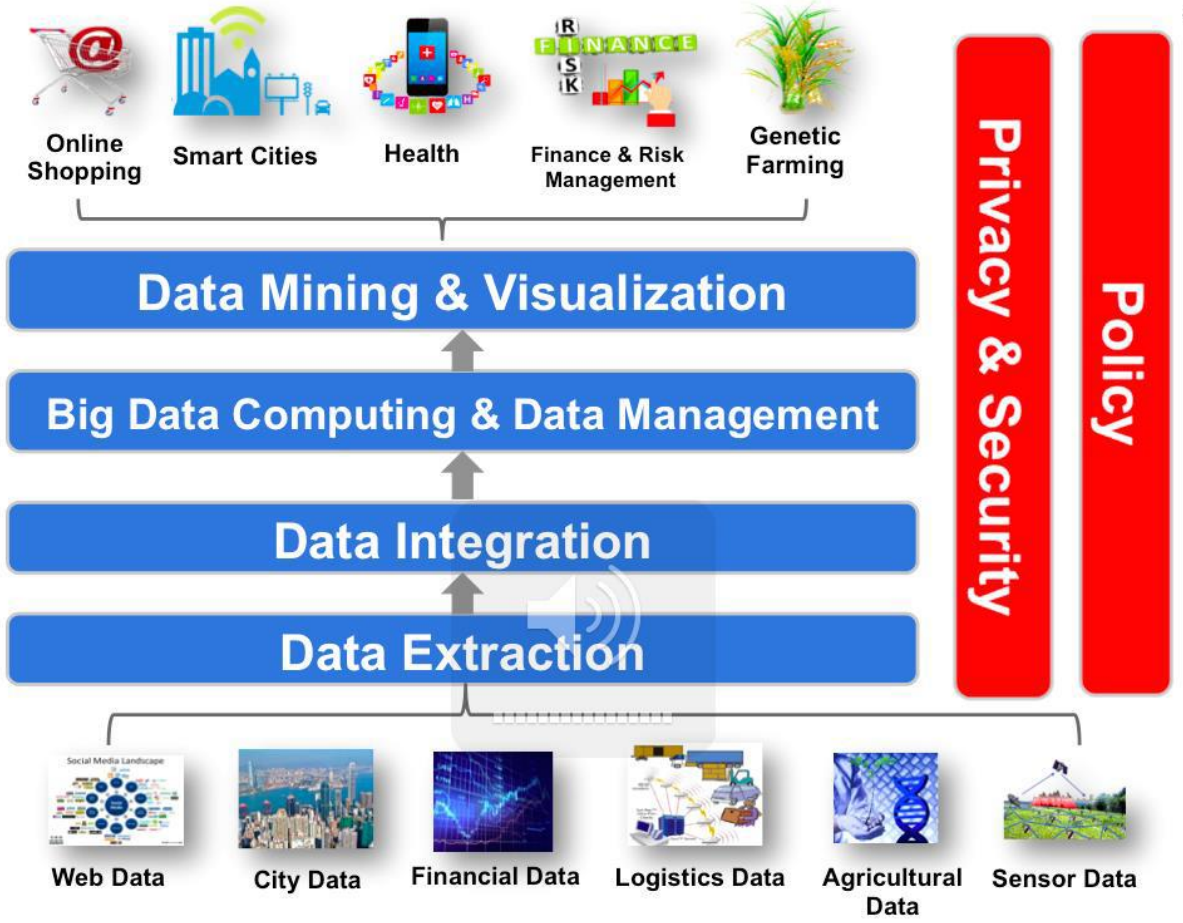
- Optimizations and predictive analytics
- Complex statistical analysis
- All types of data, and many sources
- Very large datasets
- More of a real-time

- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

THE EVOLUTION OF BUSINESS INTELLIGENCE



Big Data Analytics Objectives



The Techniques and Methods of Analytics

Learning analytics draws upon techniques from a number of established fields:

- *Statistics*
- *Artificial Intelligence*
- *Machine Learning*
- *Data mining*
- *Social Network Analysis*
- *Text Mining and Web Analytics*
- *Operational Research*
- *Information Visualization*

The Power of Big Data

- Big Data can bring "**big values" to our life in** almost every aspects.
- Technologically, Big Data is bringing about changes in our lives because it allows **diverse and heterogeneous data to be fully integrated and analyzed to help us make decisions.**
- Today, with the Big Data technology, **thousands of data from seemingly unrelated areas can help support important decisions. This is the power of Big Data.**

Big Data Usage

E-Commerce and Market Intelligence

- –Recommender system
- –Social media monitoring and analysis
- –Crowd-sourcing systems
- –Social and virtual games

E-Government and Politics

- –Ubiquitous government services
- –Equal access and public services
- –Citizen engagement

Science & Technology

- –S&T innovation
- –Hypothesis testing
- –Knowledge discovery

Big Data Usage

Smart Health and Wellbeing

- –Human and plant genomics
- –Healthcare decision support
- –Patient community analysis

Security and Public Safety

- –Crime analysis
- –Computational criminology
- –Terrorism informatics
- –Open-source intelligence
- –Cyber security

Big Data Use Cases



Big Data Exploration
Find, visualize, understand all big data to improve decision making



Enhanced 360° View of the Customer
Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources



Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis
Analyze a variety of machine data for improved business results



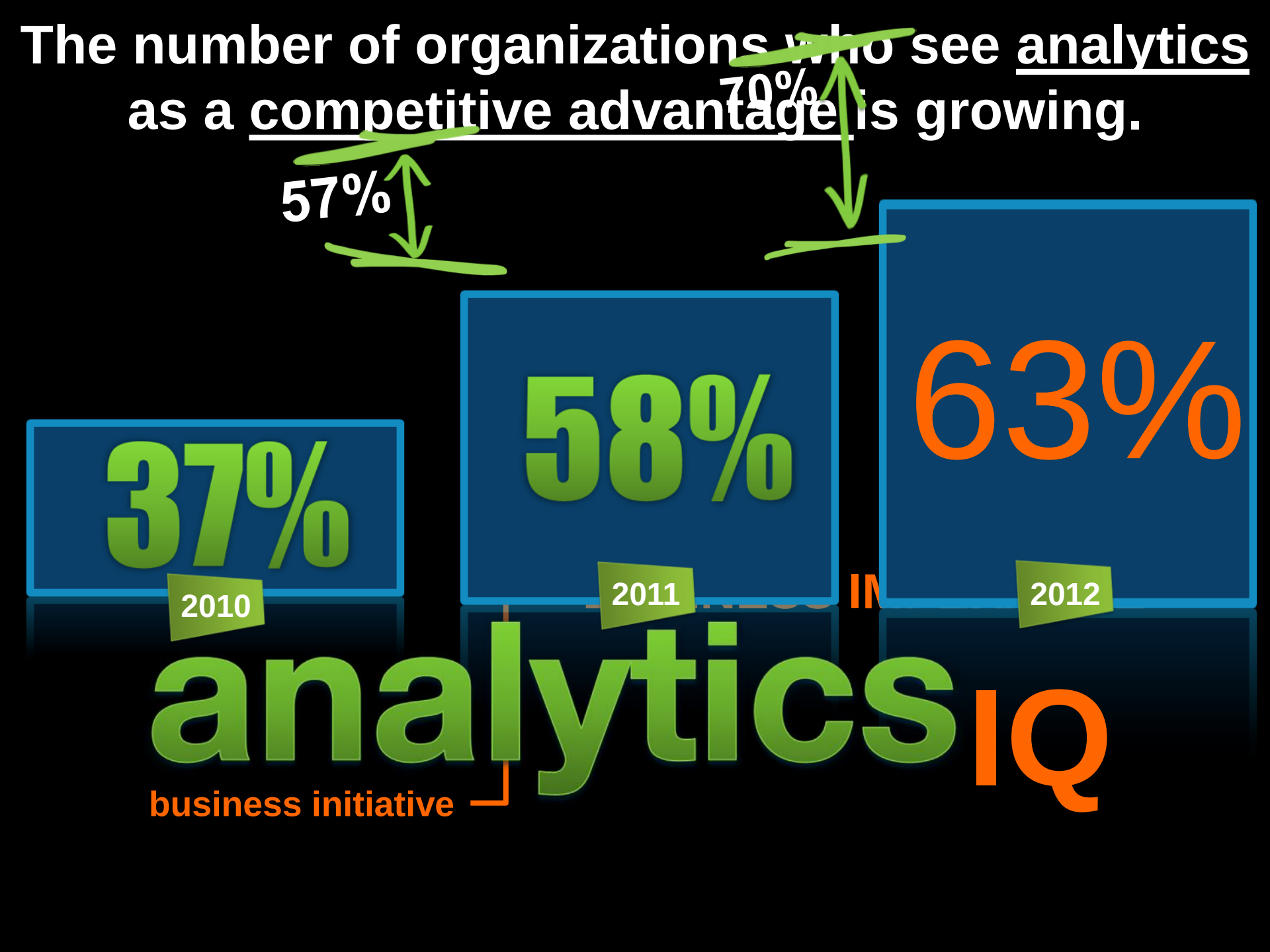
Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency

12 Big Data, Integration & Governance

McKinsey Global Institute in 2011 provided a list of the top 10 common techniques applicable across a range of industries.

Techniques

- ***Testing***
- ***Cluster Analysis***
- ***Classification***
- ***Data Mining***
- ***Network analysis***
- ***Predictive modelling***
- ***Sentiment analysis***
- ***Statistics***
- ***Visualization***


The number of organizations who see analytics as a competitive advantage is growing.
57%
70%
37%
58%
63%
2010
2011
2012
analytics
IQ
business initiative

Studies show that organizations competing on analytics outperform their peers

substantially o**utperform**

220%

1.6x Revenue Growth

2.5x Stock Price Appreciation

2.0x EBITDA Growth

IMPACT OF BIG DATA IN LIFE SCIENCES and HEALTH CARE

Big data can also be seen in the life sciences where big sets of data such as

- ✓ genome sequencing,
- ✓ clinical data and patient data

are analyzed and used to advance breakthroughs in science in research.

Human Genome Project

In biomedicine the Human Genome Project is determining the sequences of the three billion chemical base pairs that make up human DNA.

Genome biology and the study of cell circuits - the most important Big Data challenges of our time.

- A fundamental Big Data development in the field of healthcare is the ambition of the Broad Institute, an initiative of MIT and Harvard, to expand the Human Genome Project, which was eventually rounded off in 2003.
- Broad Institute is currently researching the cell mutations that cause cancer, the molecular structure of the viruses, bacteria etc. that are responsible for infectious illnesses, and the possibilities of their use in the development of medicines.

Digital healthcare data expected to reach 25,000 petabytes in 2020

- ***Standard medical practice is moving from relatively ad-hoc and subjective decision making to evidence-based healthcare.***
- *Healthcare Data is becoming more complex*
- ***In 2012, worldwide digital healthcare data was estimated to be equal to 500 petabytes***

Sources of Big Data in Healthcare

- *Structured EHR Data*
- *– Unstructured Clinical Notes*
- *– Medical Imaging Data*
- *– Genetic Data*
- *– Other Data (Epidemiology & Behavioral)*

Overall Goals of Big Data Analytics in Healthcare

- *Take advantage of the massive amounts of data and provide right intervention to the right patient at the right time.*
- *Personalized care to the patient.*
- *Potentially benefit all the components of a healthcare system i.e., provider, payer, patient, and management.*

Impact on Health care

- ***Google Flu Trends***: a service that predicts and locates outbreaks of the flu by making use of information— *aggregate search queries*.
- Of course, early detection of disease, when followed by rapid response, can reduce the impact of both seasonal and pandemic influenza.

Global Viral Forecasting Initiative (GVFI)

- The San Francisco-based Global Viral Forecasting Initiative (GVFI) uses advanced data analysis on information mined from the Internet to identify comprehensively the locations, sources and drivers of local outbreaks before they become global epidemics- GVFI‟s Chief Innovation Officer, Lucky Gunasekara, says.
- Public health offers one of the most compelling areas where the analysis of mobile and Internet data could lead to huge public gains.

DIAGNOSTICS

- In 2011, the Watson supercomputer, developed by IBM and named after its legendary CEO Thomas Watson, beat human jeopardy champions in a TV show.
- Health has been the first area of commercial application of this technology for IBM.
- Watson's ability to understand questions and context, and to rifle through 200 million pages of data and provide precise responses in just seconds, can help a physician treating a patient to consider all related texts, reference materials, prior cases, and the latest knowledge in journals and medical literature.

PATIENT CARE PROCESSES: Different actors are already entering this arena with offerings based on big data.

- Accenture targets the 4 % chronically ill patients that tie up more than 60 % of the hospital resources Through a service that monitors the chronically ill patients that can live at home,
- Avoiding readmission to hospital is a huge business case for the healthcare system and also for the patients in the form of improved life quality.
- Monitoring patients' conditions is made on the basis of medical and lab records, as well as self-reporting of key parameters from instruments available to the patients.
- One example of an entry as a data aggregator is the Premier Healthcare Alliance in the US teaming up with IBM In a pilot project with local hospital

Big data analytics in the home care of chronically ill patients

In the area of patient monitoring and management, many hospitals have started to apply big data analytics in the care of chronically ill patients to reduce cost per patient and to avoid unnecessary admissions to hospitals.

IBM, through their 'Smarter Cities' initiatives, has support home care of chronically ill patients (e.g. Stavanger Hospital in Norway).

Mobilising Data to Deal with an Epidemic

- In the wake of Haiti‟s devastating 2010 earthquake, researchers at ***the Karolinska Institute and Columbia Universit***y demonstrated that mobile data patterns could be used to understand the
- movement of refugees and the consequent health risks posed by these movements.
- cholera outbreak struck the country and the same team used mobile data to track the movement of people from affected zones.
- demonstrates how mobile data analysis could revolutionize disaster and emergency responses.

Heritage Health Prize - Improve health care win $3,000,000

- **Over $30 billion was spent on unnecessary hospital admissions.**
- **Goals:**
  - Identify patients at high-risk and ensure they get the treatment they need.
  - Develop algorithms to predict the number of days a patient will spend in a hospital in the next year.
- **Outcomes:**
- Health care providers can develop new strategies to care for patients before its too late reduces the number of unnecessary hospitalizations.
- Improving the health of patients while decreasing the costs of care.

Penalties for Poor Care - 30-Day Readmissions

- Hospitalizations account for more than 30% of the 2 trillion annual cost of healthcare in the United States.
- Around 20% of all hospital admissions occur within 30 days of a previous discharge.
- not only expensive but are also potentially harmful, and most importantly, they are often preventable.
- Medicare penalizes hospitals that have high rates of readmissions among patients with heart failure, heart attack, and pneumonia.
- Identifying patients at risk of readmission can guide efficient resource utilization and can potentially save millions of healthcare dollars each year.
- Effectively making predictions from such complex hospitalization data will require the development of novel advanced analytical models.

White House unveils BRAIN Initiative

- The US President unveiled a new bold $100 million research initiative designed to revolutionize our understanding of the human brain.
- BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative.
- Find new ways to treat, cure, and even prevent brain disorders, such as Alzheimer's disease, epilepsy, and traumatic brain injury.
- *"Every dollar we invested to map the human genome returned $140 to our economy... Today, our scientists are mapping the human brain to unlock the answers to Alzheimer's."* ***-- President Barack Obama, 2013 State of the Union.***
- advances in "Big Data" that are necessary to analyze the huge amounts of information that will be generated; and increased understanding of how thoughts, emotions, actions and memories are represented in the brain." :
- http://www.whitehouse.gov/infographics/brain-initiative

- ***"Big data technologies, together with the sensors that ride on the "Internet of Things," pierce many spaces that were previously private".***

(White House Report, 2014)

Big Data is developing most rapidly in the world of Big Science

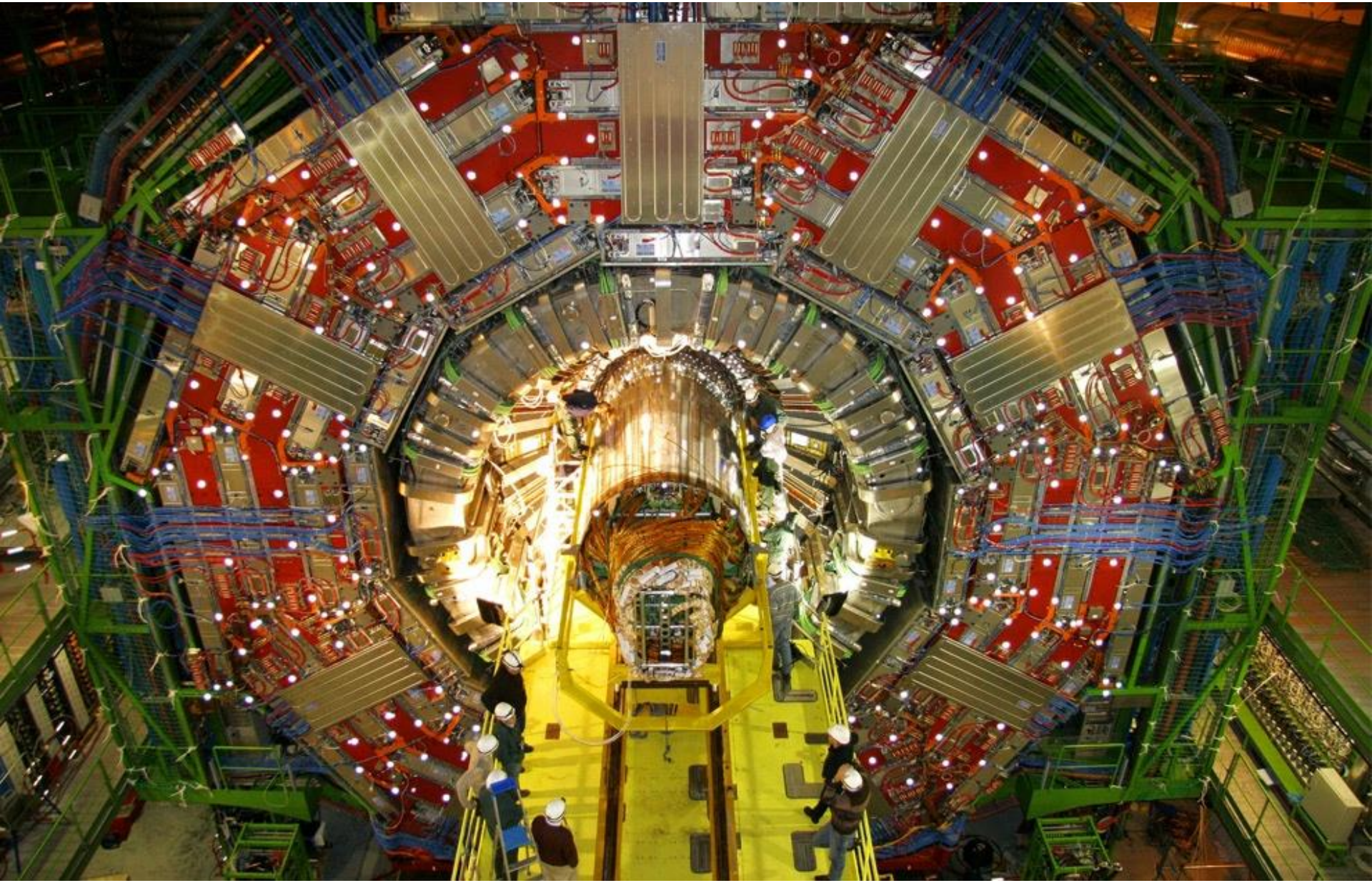
- In 10 years,2800 radio telescopes in the Square Kilometer Area project (ska), the largest Big Science project ever, will generate

1 billion gigabytes of data daily. That is equal to the entire Internet on a weekday in 2012.

- ***In 2012, Google processes* a total of 5 petabytes or 5000 terabytes per hour.**

Large Hadron Collider

- Another example of Big Data is the Large Hadron Collider, at the European Organisation for Nuclear Research (CERN), which has
- 150 million sensors and is creating
- 22 petabytes of data in 2012.



CERN's Large Hydron Collider (LHC) generates 15 PB a year

The Earthscope

- The Earthscope is the world's largest science project.
- Designed to track North America's geological evolution, this observatory records data over 3.8 million square miles, amassing 67 terabytes of data.
- It analyzes seismic slips in the San Andreas fault, sure, but also the plume of magma underneath Yellowstone and much, much more. ***(http://www.msnbc.msn.com/id/44363598/ns/technology_and_science-future_of_technology/#.TmetOdQ--uI)***



International Virtual Observatory Alliance

The Fourth Paradigm, is that the volume of data available to us is so large, now often termed Big Data, that it is both *presenting many new opportunities for analysis as well as requiring new modes of thinking*, for example in the International Virtual Observatory Alliance (http://virtualobservatory.org) and in citizen science.

The Virtual Astronomical Observatory (VAO)

- The Virtual Observatory (VO) is an international effort to bring a large-scale electronic integration of astronomy data, tools, and services to the global community
- VAO is the U.S. NSF- and NASA-funded VO effort that seeks to put efficient astronomical tools in the hands of U.S. astronomers, students, educators, and public outreach leaders.
- These tools will make use of data collected by the most powerful telescopes in the world.

Square Kilometre Array (SKA) to produce one exabyte data/day

One clear example of Big Data is the Square

Kilometre Array (SKA) planned to be constructed in South Africa and Australia.

When the SKA is completed in 2024 it will produce in excess of one exabyte of raw data per day, which is more than the entire daily internet traffic at present.

(www.skatelescope.org)

SKA

The SKA is a 1.5 billion Euro project that will have more than

3000 receiving dishes to produce a combined information collecting area of one square kilometre, and

will use enough optical fibre to wrap twice around the Earth.

Huge Data from Satellites in orbit

In Earth observation there are over 200 satellites in orbit continuously collecting data about the atmosphere and the land, ocean and ice surfaces of planet Earth with pixel sizes ranging from 50 cm to many tens of kilometres.

World Data System (WDS)

Approved members of the World Data System so far include

I. *centres for Antarctic data (Hobart),*
II. *climate data (Hamburg),*
III. *ocean data (Washington DC),*
IV. *environment data (Beijing)*
V. *solid Earth physics data (Moscow)*
VI. *plus the International Laser Ranging Service and the International Very Long Baseline Interferometry Service.*
VII. *It is anticipated that the WDS will comprise over 100 centres and networks of active, professional data management.*

The objectives of the World Data System (WDS)

- *Enable universal and equitable access to quality-assured scientific data, data services, products and information;*
- *Ensure long term data stewardship; Foster compliance to agreed-upon data standards and conventions;*
- *Provide mechanisms to facilitate and improve access to data and data products*

Improvements in professional data management will result in better science

Big Data presents science with many challenges, but at the same time presents many opportunities to influence how science grows and develops for the better.

"Big Data" Coming From-------

12+ TBs
of tweet data
every day

30 billion RFID
tags today
(1.3B in 2005)

4.6 billion
camera
phones
world
wide

? TBs of
data every
day

100s of millions of GPS enabled
devices
sold
annually

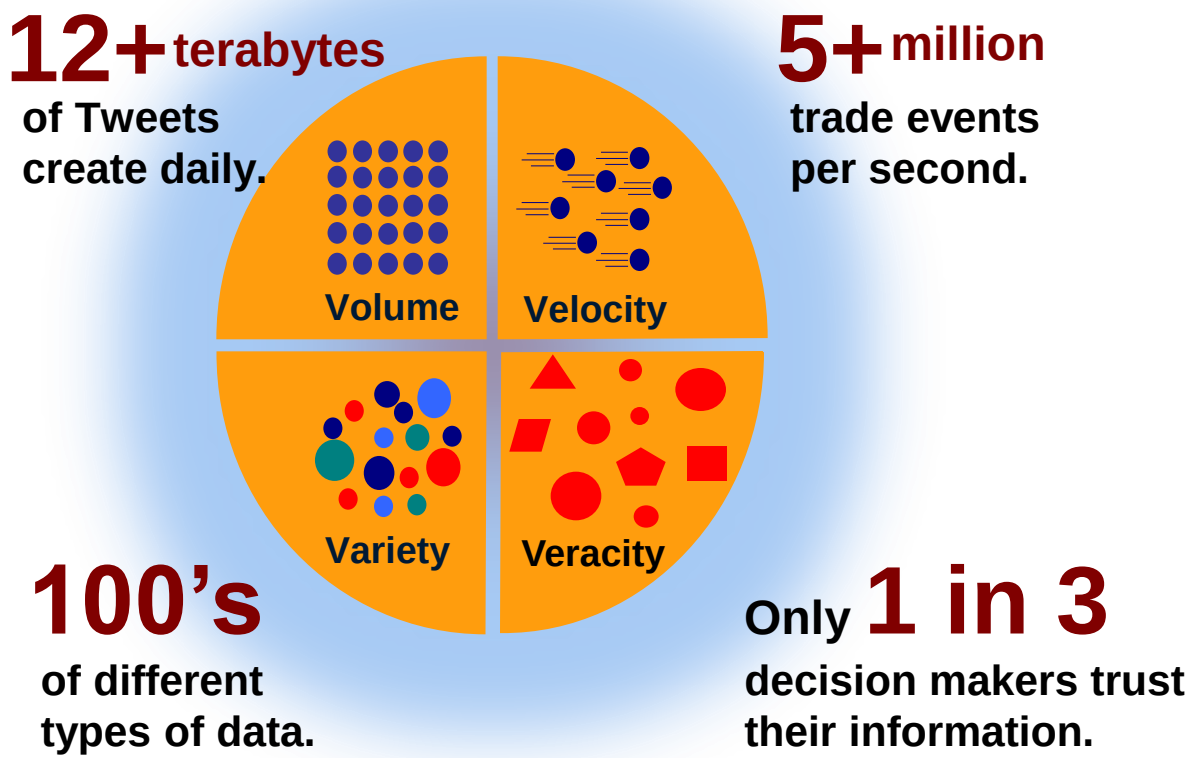
25+ TBs
of
log data
every day

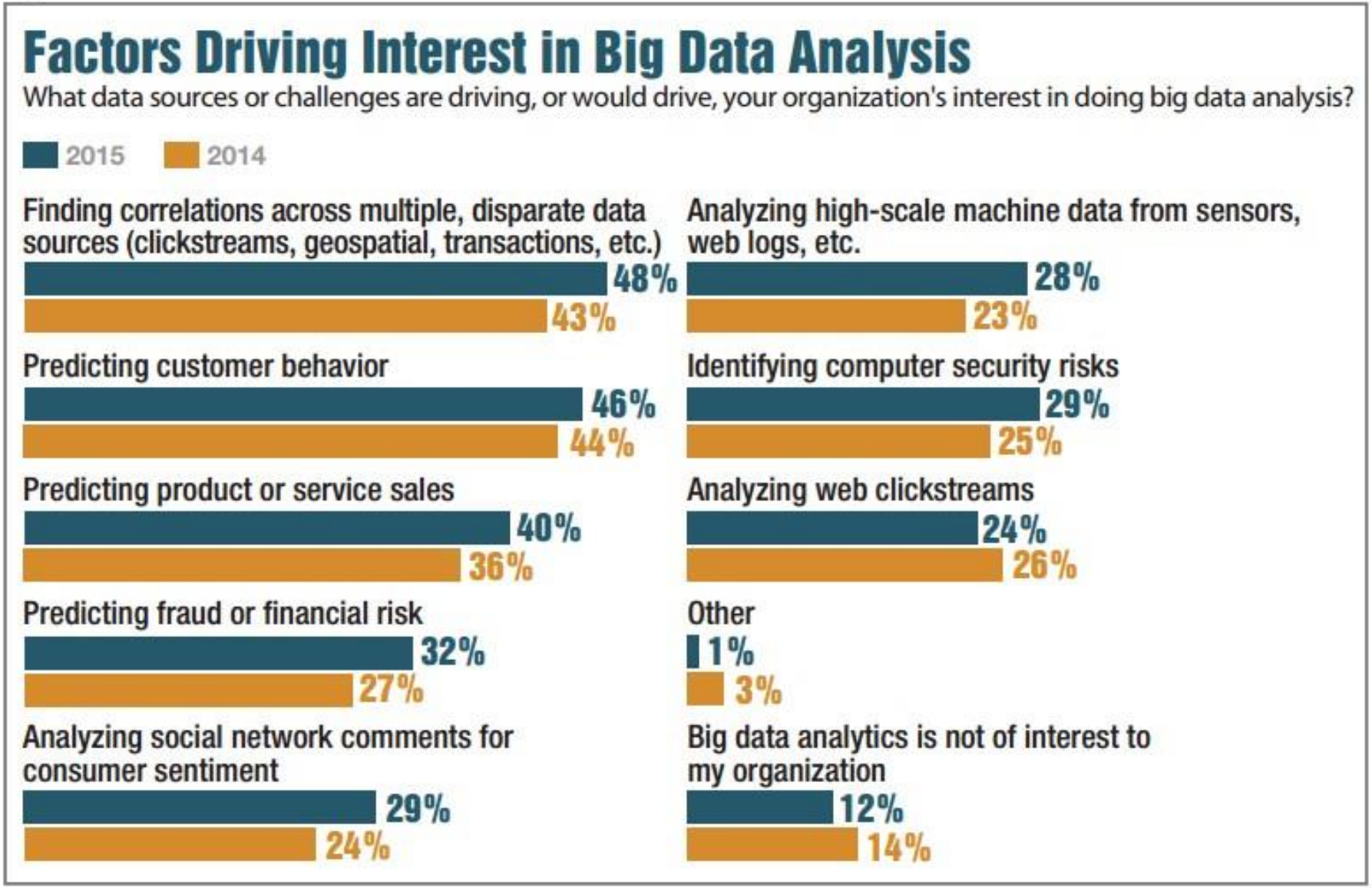
2+ billion
people
on the
Web by
end 2011

76 million smart
meters in 2009…
200M by 2014



With Big Data, We've Moved into a New Era of Analytics




Factors Driving Interest in Big Data Analysis
What data sources or challenges are driving, or would drive, your organization's interest in doing big data analysis?
2015
2014
Finding correlations across multiple, disparate data sources (clickstreams, geospatial, transactions, etc.)
48%
43%
Analyzing high-scale machine data from sensors, web logs, etc.
28%
23%
Predicting customer behavior
46%
44%
Identifying computer security risks
29%
25%
Predicting product or service sales
40%
36%
Analyzing web clickstreams
24%
26%
Predicting fraud or financial risk
32%
27%
Other
1%
3%
Analyzing social network comments for consumer sentiment
29%
24%
Big data analytics is not of interest to my organization
12%
14%
Note: Multiple responses allowed
R8241114/10
Base: 297 respondents in October 2014 and 248 in October 2013 at organizations using or planning to deploy data analytics, BI, or statistical analysis software
Data: InformationWeek Analytics, Business Intelligence, and Information Management Survey of business technology professionals

Big Data and Web 2.0

- O'Reilly Media introduced the term "Big Data" a year after Web 2.0 appeared, as many valuable Big Data situations are indeed related to consumer conduct.
- Web 2.0 provided the impulse to rethink the interaction that was taking place on Internet, and to push it somewhat further.
- In much the same way, the qualification "Big Data" calls attention to the business possibilities of the flood of data on the one hand, and the new technologies, techniques and methods that are directed toward these, on the other.

Business intelligence and analytics (BI&A)

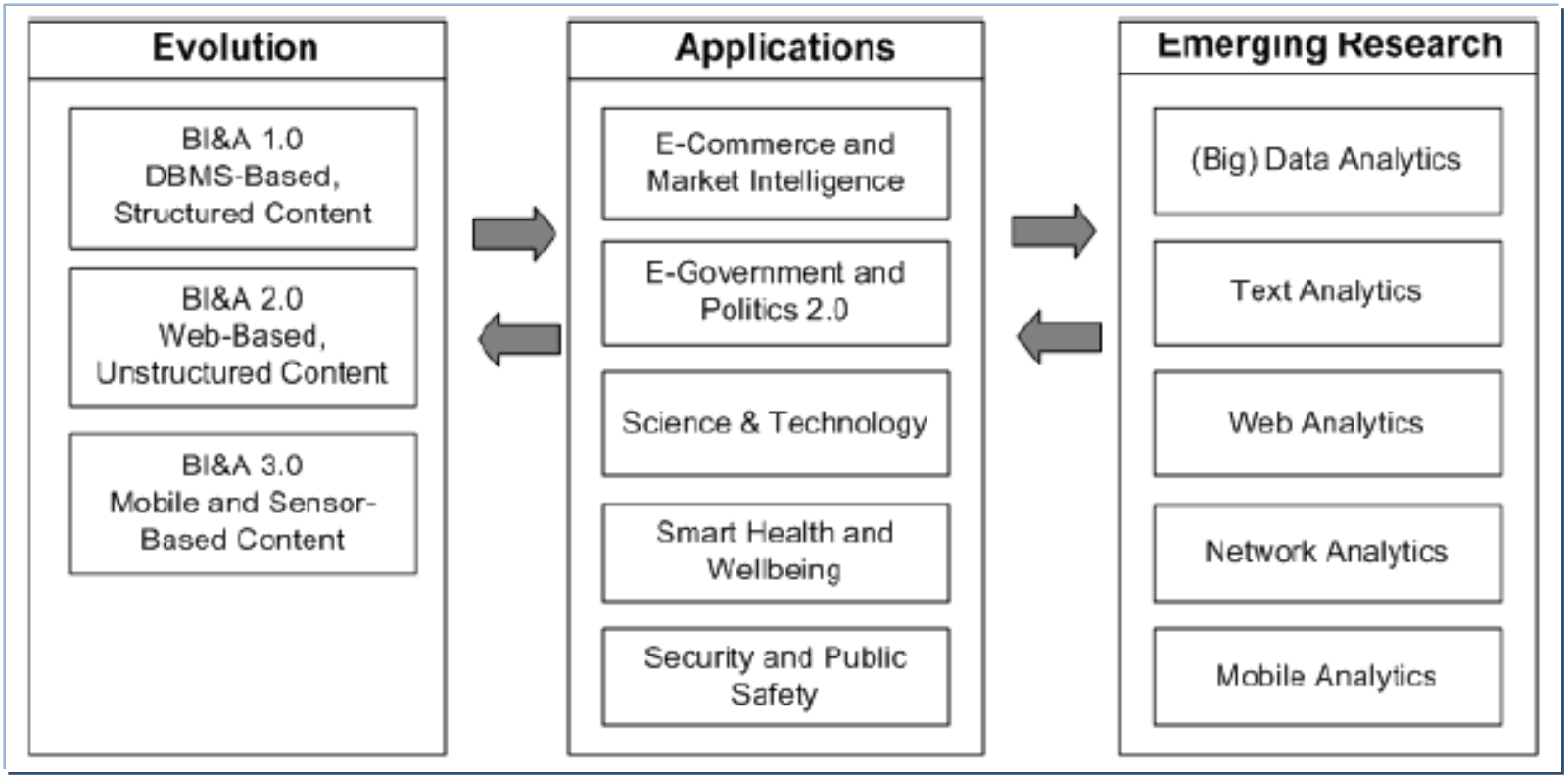
Business intelligence and analytics (BI&A) and the related field of big data analytics have become increasingly important in both the academic and the business communities over the past two decades.

Big impact of Big Data

- **Business intelligence and analytics (BI&A)**

	Key Characteristics
BI&A 1.0	DBMS-based, structured content • RDBMS & data warehousing • ETL & OLAP • Dashboards & scorecards • Data mining & statistical analysis
BI&A 2.0	Web-based, unstructured content • Information retrieval and extraction • Opinion mining • Question answering • Web analytics and web intelligence • Social media analytics • Social network analysis • Spatial-temporal analysis
BI&A 3.0	Mobile and sensor-based content • Location-aware analysis • Person-centered analysis • Context-relevant analysis • Mobile visualization & HCI

Big impact of Big Data



Big data is also transforming the retail market.

- Wal-Mart's transactional databases →more than 2.5 petabytes of data consisting of customer behaviors and preferences, network and device activity, and market trends data.
- Wal-Mart's inventory management system, called Retail Link, pioneered the age of big data by enabling suppliers to see the exact number of their products on every shelf of every store at each precise moment in time.
- **Amazon's** "Customers Who Bought This Also Bought" feature, prompting users to consider buying additional items selected by a collaborative filtering tool.



- ✓ One of the largest online retailers, and cloud and digital content providers is Amazon.
- ✓ Amazon was originally founded as an online bookstore in 1994, but has evolved into a complete retail platform/store.
- ✓ One of Amazon's key success factors was the early use/adaption of machine learning and big data for targeted marketing, with as much as 30 % of the sales driven by automatic customer recommendations.

Big data in retail - Tesco



Major use of BI and loyalty program

Outsourced to Dunnhumby for analysis
80% of customers are Clubcard members
Constantly updating data – every transaction improves data quality
Targeted promotions both direct and sponsored by suppliers (revenue stream)
But Tesco is a financial services company as well...



- Founded in 1995, eBay Inc. is an American multinational consumer-to-consumer Internet corporation with presence in over thirty countries.

- eBay is now a multibillion dollar business., with a business model of simply providing a network that connects buyers and sellers of goods and where anyone can trade.

Value Chain Optimisation

Big data capabilities enable companies such as

- ✓ FedEx ,
- ✓ Wallenius Wilhelmsen (car carrier ship owner,),
- ✓ Maersk “Daily Maersk conveyor belt” (container ship owner),

guaranteeing daily shipments between major Asian and European ports), to streamline their supply chain and improve efficiency. The optimisation may include predictions using weather data, tracking data, traffic data, and other data that may influence the movement of cargo and parcels.

business models with examples.

BUSINESS MODEL	IMPACT FROM BIG DATA EXAMPLES	EXAMPLES
Value Chain	■Big data will enable further optimisation of operating value chains. ■ Big data is also used to move towards becoming a value network.	FedEx, Walmart

business models with examples.

Value Shop	■Mode of operation (e.g. automation of knowledge services) ■Types of problems that customers can be assisted in solving ■Optimisation of efficiency and scalability of services through efficient re-use of data and automated services	McKinsey, Accenture, DNV GL, Taleris, Sekal, ESRG

business models with examples.

BUSINESS MODEL	IMPACT FROM BIG DATA EXAMPLES	EXAMPLES
Value Network	■Increase the network value of the network ■Capitalize on the data generated by the network	Google, eBay, FedEx, Carfax, Twitter, Vestas Online, Maersk line, Shipping KPI, Apple, Amazon, Marinexplore



- Accenture developed an Asset Analytics solution for Canadian utility Hydro One, based on Hydro One's existing ERP, GIS, data historian, and other legacy asset systems.
- It helps them with improving CAPEX/OPEX investment decisions, provides a common understanding of asset risk, supports regulatory compliance, and facilitates communications..

The OstiaEdge® Monitoring Suite

- OstiaEdge® enables improved use of onboard data to help implement a Condition-Based Maintenance (CBM strategy).
- This creates value for the operators, managers, and owners in a variety of ways;
- Prioritizes maintenance spend across a fleet of assets to maximise return on investment.



Lloyd's List Intelligence is founded on the heritage of Lloyd's List plus a global network of specialist sources of business-critical maritime data.

It brings together the expertise of a global staff of maritime analysts and journalists with an extensive system of shore based and satellite intelligence gathering to create a complete information support service for the maritime industry.

"

WALLENIUS WILHELMSEN LOGISTICS

- "Real-time analysis helps make contingency plans for different situations – for example, a sudden rise in oil prices, or a closure of a major port.
- The solution enables planning for best- and worst-case scenarios as well as the standard forecast, enabling the company to respond more effectively to normal demand fluctuations as well as emerging situations, such as the2011 Japanese tsunami.

Wallenius Wilhelmsen Logistics are saving in the order of 100 million dollars annually from route planning and general reductions in speed.



- Carfax is a US company and system that allows car owners to collect and store the complete history of a car in a central database.
- Whenever the car visits a workshop, the work done and the condition of the car is uploaded to the system. For the car owner, being part of network increases the second-hand value of the car.
- The system has simplified buying used cars and, as the service is available for all brands, circumvention of the brand/vendor barrier adds extra value.



Asian Telco reduces billing costs and improves customer satisfaction.

Capabilities:

Stream Computing
Analytic Accelerators

Real-time mediation and **analysis** of
6 Billion CDRs per day

Data processing time reduced from
12 hrs to 1 sec

Hardware cost reduced to 1/8th

Proactively address issues
(e.g. ***dropped calls***) impacting customer ***satisfaction***.

Higher Education's future

- By 2020, higher education will be quite different from the way it is today.
- There will be mass adoption of teleconferencing and distance learning to leverage expert resources.
- Most universities' assessment of learning will take into account more individually-oriented outcomes and capacities that are relevant to subject mastery.
- Requirements for graduation will be significantly shifted to customized outcomes.

SPORTS

Another story, on how the use of data literally changed the game, is told in the book and movie

" Moneyball" is the story about how available data and analytics were applied in a new way that challenged, and entirely altered, the traditional way of valuing baseball players.

This has completely transformed the way in which players in the US National baseball league are valued, and similar methods are also being applied in other sports.

Potential Value & Benefits of Big data Analytics

***Big data analytics* can accumulate**

- ✓ **the** wisdom of crowds
- ✓ reveal patterns
- ✓ yield best practices

BIG DATA "USE CASES" WITHIN BUSINESSES



48% Customer Analytics

21% Operational Analytics

12% Fraud & Compliance

10% New Product & Service Innovation

10% Enterprise Data Warehouse Optimiz

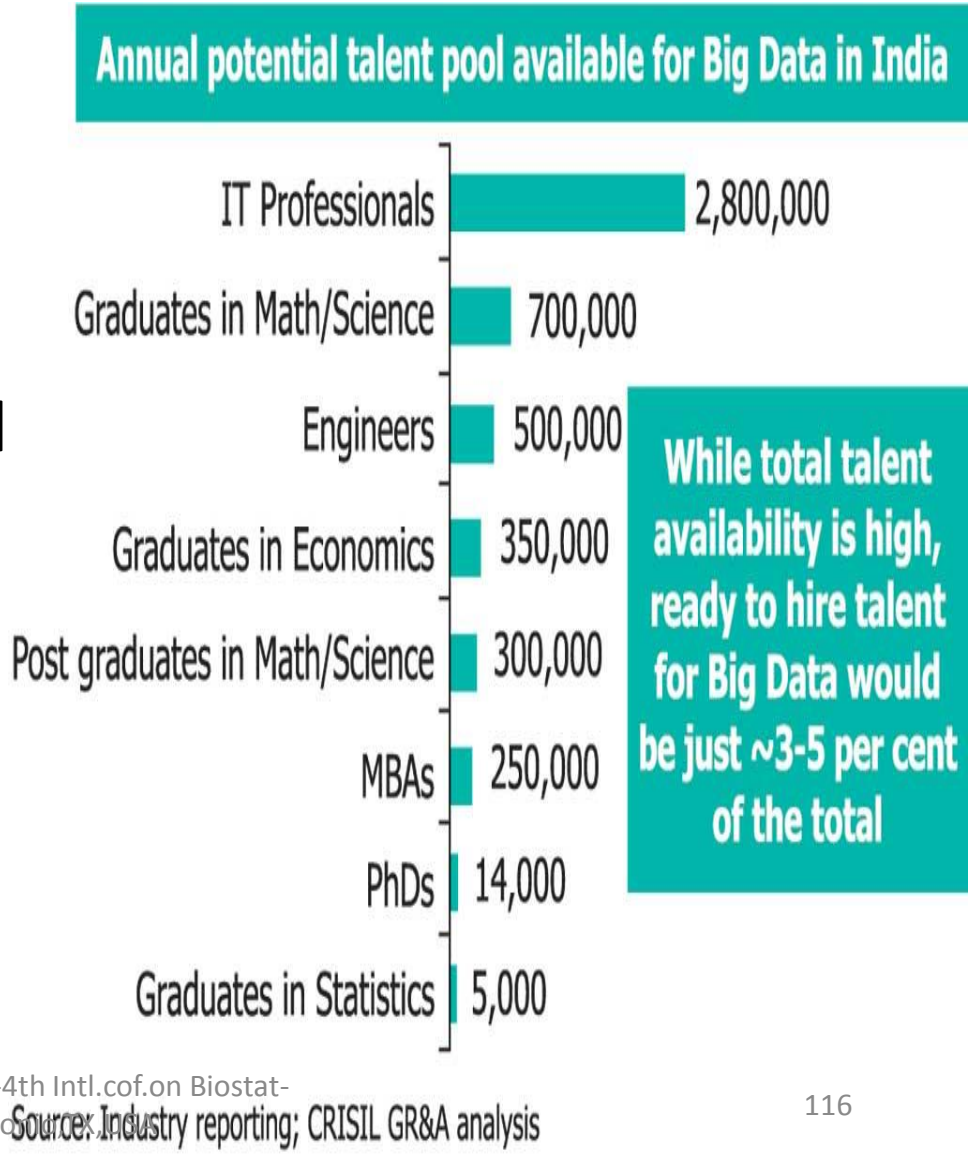
*Adds to 101% due to rounding

How Big data impacts on IT

- ***By 2015 4.4 million IT jobs in Big Data ; 1.9 million is in US itself***
- ***India will require a minimum of 1 lakh data*** *scientists in the next couple of years* ***in addition to data analysts and data managers to support the Big Data space.***
- ***Big data is a troublesome force presenting opportunities with challenges to IT organizations.***

Potential Value of Big Data

- ✓ $300 billion potential annual value to US health care.
- ✓ $600 billion potential annual consumer surplus from using personal location data.
- ✓ 60% potential in retailers' operating margins.



Benefits of Big Data

•Real-time big data isn't just a process for storing petabytes or exabytes of data in a data warehouse, It's about the ability to make better decisions and take meaningful actions at the right time.

•Fast forward to the present and technologies like Hadoop give you the scale and flexibility to store data before you know how you are going to process it.

•Technologies such as MapReduce, Hive and Impala enable you to run queries without changing the data structures underneath.

Benefits of Big Data

- Research finds that organizations are using big data to target customer-centric outcomes, tap into internal data and build a better information ecosystem.
- Big Data is already an important part of the $64 billion database and data analytics market
- It offers commercial opportunities of a comparable scale

➢ ***to enterprise software in the late 1980s***

➢ ***the Internet boom of the 1990s***

➢ ***the social media explosion of today.***

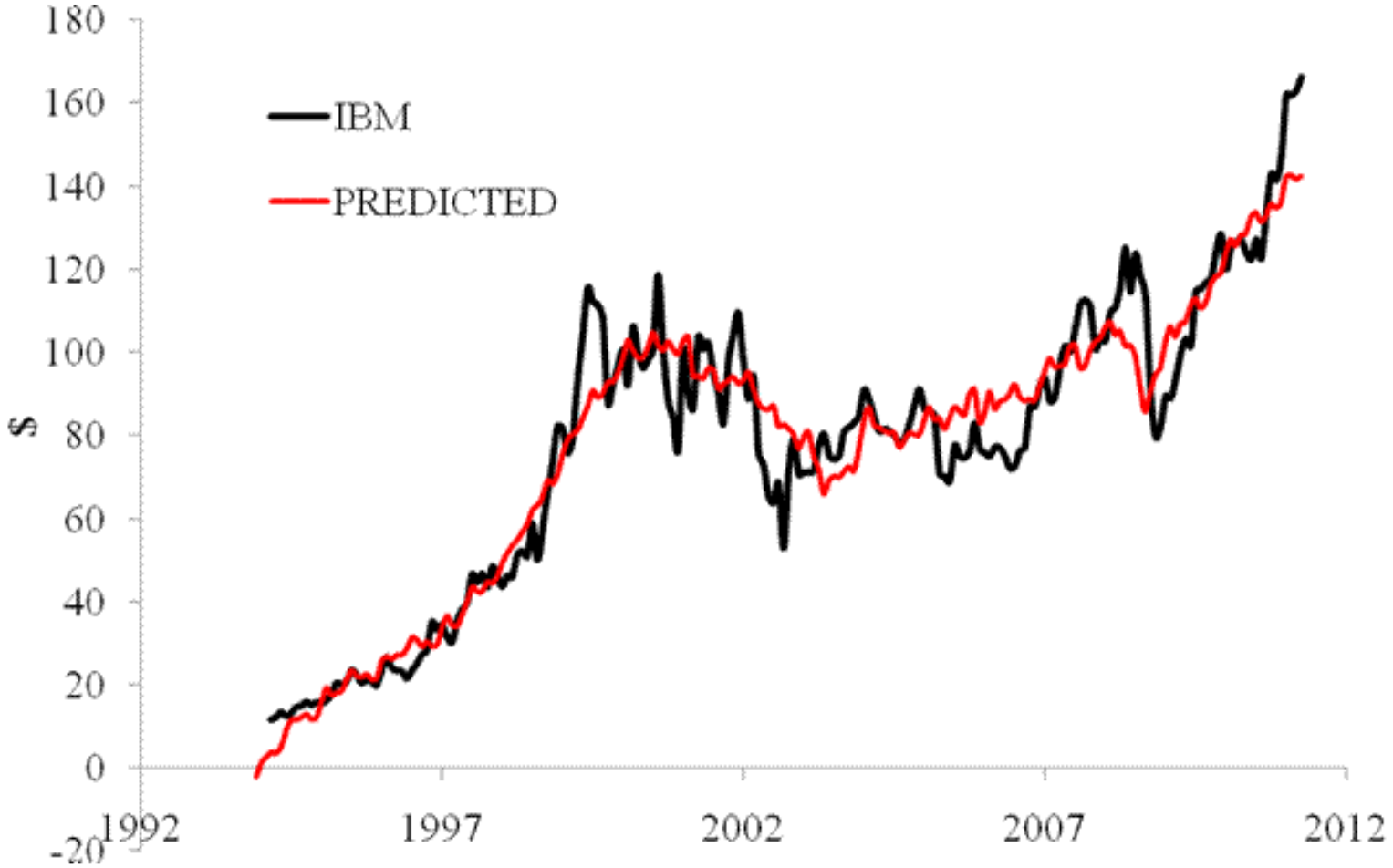
Impact of Big data on Finance

Big Data can be seen in the finance and business where enormous amount of stock exchange,banking, online and onsite purchasing data flows through computerized systems every day and are then captured and stored for inventory monitoring, customer behavior and market behavior.

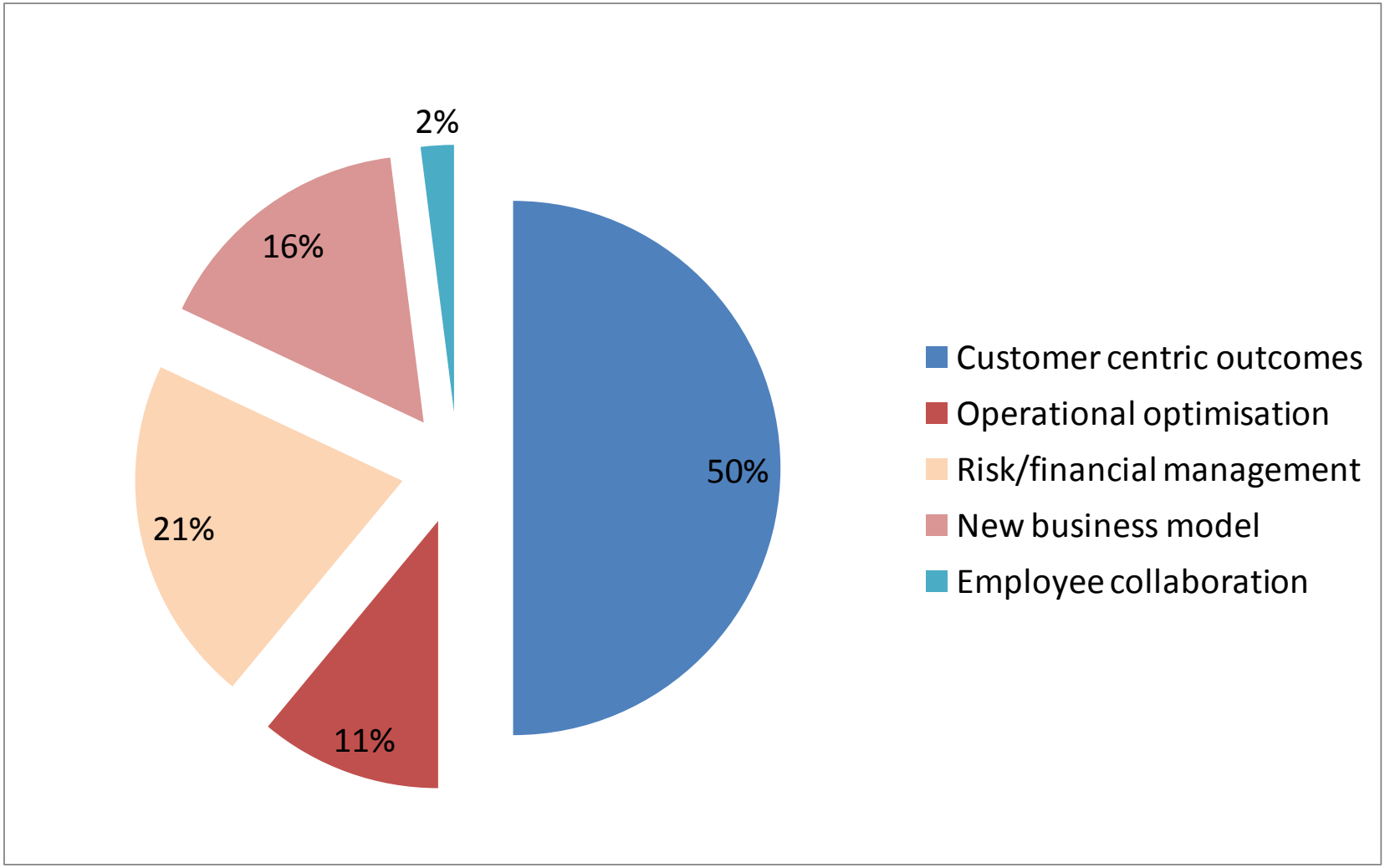
So who uses Big Data?

Internet	Google, Amazon, eBay, AOL – everything
Mobile Gaming	On line betting, multi user games
Marketing	Social Network Analysis, Digital Advertising
Telcos	Call Routing Management, Subscriber Churn
Retail	Tesco – customer segmentation Netflix – everything
Utilities	Smart Grid for meters
Banking	Fraud detection, Customer segmentation
Insurance	Telemetrics on cars

Trend of Big Data use in finance?



So how is Big Data used in finance?



So who uses big data in finance?

- Monitor customer journeys:
- website clicks
- transaction records
- bankers' notes
- voice recordings
- Make pre-emptive competitive credit offers
- Selective BankAmeriDeals cash back offers



So who uses big data in finance?

- Insurance uses:
- Fraud Detection & Analysis
- Personalised Pricing
- Customer Sentiment Analysis
- Catastrophic Planning
- Call Detail Record
- Loyalty Management
- Social Media Analytics
- Advertising and Campaign Management

Machine-learning model

An experiment by the Jumiya project in conjunction with Filene Research Institute estimates that credit unions using a machine-learning model could:

- ✓ Improve on traditional credit scoring by 10%
- ✓ Approve 1% more loans
- ✓ Reduce the cost of delinquencies by 15%
- ✓ Have the biggest benefits for members with credit scores between 450 and 600

Big data in insurance – the single view and the spy in the car





Further Areas of Research

Big Data is of central importance are Energy Sector, seastronomy, oceanography etc. among many others.

Impact of Big Data on Energy sector

- Grid analytics

■Outage management;■Voltage optimisation;■Asset Management

- Consumer analytics

■Revenue protection;■Load forecasting ■Detailed customer segmentation■ Energy efficiency, advice and assessment

- WIND Industry
- There are already many untapped sources for insights and performance enhancements in the data available in the wind industry today.

Benefits-----

- The smart grid is designed to allow electricity service providers, users, and other third parties to monitor and control electricity use.

- Benefits: who are able to reduce energy consumption by learning which devices and appliances consume the most energy, or which times of the day put the highest or lowest overall demand on the grid.

GE´s “Grid IQ Solutions as a Service”

General Electric is taking the smart grid platform to the next level.

***GE’s smart grid-as-a-service pitch* can be illustrated as follows: the city utility of Norcross, Georgia, USA. wants a smart grid, but doesn’t have the means to buy one.**

GE’s “Grid IQ Solutions as a Service” will build the grid for a monthly fee, and then run it from a giant cloud-computing platform in Atlanta.



GE imagination at work

Offshore Oil and Gas

- In the oil and gas industries, companies have grappled with big data challenges for quite some time, for example to process seismic data.
- The new trend is that data and analytics are playing larger roles in other parts of the operations such as integration of offshore operations across offshore assets and system barriers, improving efficiency of drilling operations and optimizing production processes etc.
- These uses of data pave the way for further improving reliability and managing risk across the whole enterprise, including the risk of environmental impact as exploration and production move into more sensitive areas.

ConocoPhillips forecasts ice floe movement in or near real-time

- In order to help improve productivity at its Arctic facilities, ConocoPhillips wanted to monitor and forecast ice floe movement in or near real-time.
- Big data technologies helped the company to collect and manage thousands of data points per second from multiple sources.
- Through predictive analytics, ConocoPhillips can visualize in real-time the position of ice floe near its Arctic facilities.

Big Data for Global Development---

- Global Pulse-UN Project
- Ban Ki Moon, the Secretary-General of the United Nations, presented the so-called "Global Pulse project" at the end of 2011.
- Global Pulse is a U N initiative, to leverage innovations in digital data, rapid data collection and analysis to help decision-makers gain a real-time understanding of how crises impact vulnerable populations

Big Data for Global Development & *Well-being*

- Global Pulse, a new initiative by the United Nations, wants to leverage Big Data for global development.
- The group will conduct so-called sentiment analysis of messages in social networks and text messages — using natural-language deciphering software — to help predict job losses, spending reductions or disease outbreaks in a given region.
- The goal is to use digital early-warning signals to guide assistance programs in advance to, for example, prevent a region from slipping back into poverty.

Global Pulse

The project is being run as a cooperative endeavor with five main objectives:

1. *A Global Snapshot of Well-being through Mobile Phones*
2. *Real-Time E-Pricing of Bread*
3. *Tracking the Food Crisis via Online News*
4. *Unemployment through the Lens of Social Media*
5. *Twitter and Perceptions of Crisis-Related Stress*

What kind of “big data”

- Automatically collected
- Byproduct of another activity, digital crumbs, "passively" generated
- Digitally generated through transactions online ("crumbs"), active/passive sensor monitoring/recording
- Velocity/volume… (variety)
- Geographically or temporally trackable – e.g. mobile phone location data or call duration time.
- Potentially continuously analysed - in "real time" or not for "reality mining" (UN Global Pulse (2012) Big data for development: challenges & opportunities, p.18):
  - “Continuous data analysis over streaming data” (e.g., online prices, GPS & optimal routing)
  - “Online digestion of semi-structured data and unstructured ones” (e.g., news, reviews, blogs, tweets)
  - “Real-time correlation of streaming data (fast stream) with slowly accessible historical data repositories.”

Big Data: UN Global Pulse taxonomy*

1. Data Exhaust: digital services create networked sensors of human behavior.

- Passively collected transactional data from people's use of digital services
  - Mobile phones: Call Detail Records (CDR) from mobile phones - i.e. log of calls for billing purpose with basic metadata
  - Purchases (in-store and online credit cards) and financial transfers
  - Web searches, and search engines trends and analytics --" Google flu"-style
  - Geolocation and all kind of individual / personal / local sensors on computers, phone, watch, bracelet, necklace, etc + motion/sound/photo/video capturing / processing, etc
- Operational metrics and other real-time data collected by UN agencies, NGOs and other aid organisations to monitor their projects and programmes: e.g. stock levels, school attendance, IDP & refugee registration, etc.

2. Online Information – web usage and content as a sensor of human intent, sentiments, perceptions, and want.

- Web content such as news media, news articles obituaries, e-commerce, job postings, bibliographic databases, online full-text libraries
- Social media interactions (e.g. blogs, Twitter) and social media bulk contents
- Web scrapping from open public online contents (web sites, Instagram, …, text/photo/audio/video processing and pattern recognition, feature extraction, etc.)

3. Physical Sensors – focuses on remote sensing of changes in human activity.

- Remote sensing, weather data + astronomical + earth science data: land use, urban development and topographic changes, etc
- Scanned or image/audio/video recording/transmission/processing + new personal sensors (watch, bracelets, phones, etc.) + home sensors, environmental sensors for pollution, etc.

4. Citizen Reporting or Crowd-sourced Data – Information actively produced or submitted by citizens through mobile phone-based surveys, hotlines, user-generated maps, etc; While not passively produced, this is a key information source for verification and feedback

5. [UNPD: Simulated probabilistic data and agent-based simulations] – including probabilistic estimations and/or projections with thousands of trajectories, parameters, and multidimensional data arrays (e.g., indicator, location, time, age, sex, etc.)

(*) Big data for development: challenges & opportunities, p.16 http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf

Migrations and online search

- **Estimations and predictability of migration flows using Google Trends**:
  - National and sub-regional patterns of in-migration from EU8 countries to UK, and the language of their search. Office of National Statistics from the UK (Williams & Ralphs, 2013)
  - Comparison of the popularity of migration-to-Spain related queries introduced to Google Search in Argentina, Colombia and Peru, to changes in a quantity of residents' registrations in Spain, performed by immigrants proceeding from these countries between the years 2005 and 2010 (Wladyka, 2013)
  - Comparison of global Google search query data to historical official monthly statistics on migration by country (on-going Google, UN Global Pulse and UNFPA Research Project)

Migrations and social media

- Infer **migration trends and compare patterns of internal and international migration in OECD countries using geo-located social media data adjusted for selection bias** (Zagheni et al. 2014): using geo-located posts on Twitter of 15,000 users with an established minimum level of activity and for which they have consistent information over time, distinguishing between residents, who were tweeting from one country, and migrants, who were tweeting from different countries.
- Infer **lifetime migration using aggregated, anonymized data on all Facebook users** who list both their hometown and their current city on their Facebook profile (Facebook Data Science team 2013)
- Analyse transnational networks and diaspora groups or migration-related **public discourse** through social media content (Nedelcu, 2012; Oiarzabal, 2012), political activism of migrants and minority groups (Conversi, 2012; Kissau, 2012), migrants' integration into the host society (Rinnawi, 2012; Unite Europe project) etc.

Big data and fighting criminal migration-related activities

- Human trafficking:
  - How Big Data Battles Human Trafficking: From services for victims to prosecuting offenders, new technologies are being utilized to address exploitation. U.S. News. Jan. 14, 2015
  - Command, Control and Interoperability Center for Advanced Data Analysis at Rutgers University: CCICADA's Proprietary Algorithms Sort through Millions of Bits of Online Data, Sniffing Internet Ads for Clues, May 9, 2014
  - Microsoft Research Faculty 2012 Summit: panel on the Role of Technology in Human Trafficking [slides]
  - USC Center on Communication Leadership & Policy (2011). Human Trafficking Online: The Role of Social Networking Sites and Online Classifieds - http://technologyandtrafficking.usc.edu/report/
- Migrant smuggling:
  - In the context of the European migrant crisis in the Mediterranean, see references to fight migrant smuggling by taking down websites used by smugglers

Big data and financial transfers

- **Financial data** (banks, postal offices, etc.): analysis of remittance flows
- **Credit card transaction** and analysis of residents and foreign visitors in Spain (Sobolevsky et al., 2014)
- **Mobile money transfers**: e.g., M-PESA in Kenya (Hughes and Lonie, 2007) since 2007, now 15 million users and processes 2 million transactions per day in a country of 25 million adults) and now available in 70+ countries, and modalities and determinants of mobile money transfers in the aftermath of natural disasters in Rwanda (Blumenstock et al., 2013)
- Question about cross-border financial flows: how do we know that the financial flows are transmitted by migrants?

Dynamic population mapping using mobile phone data

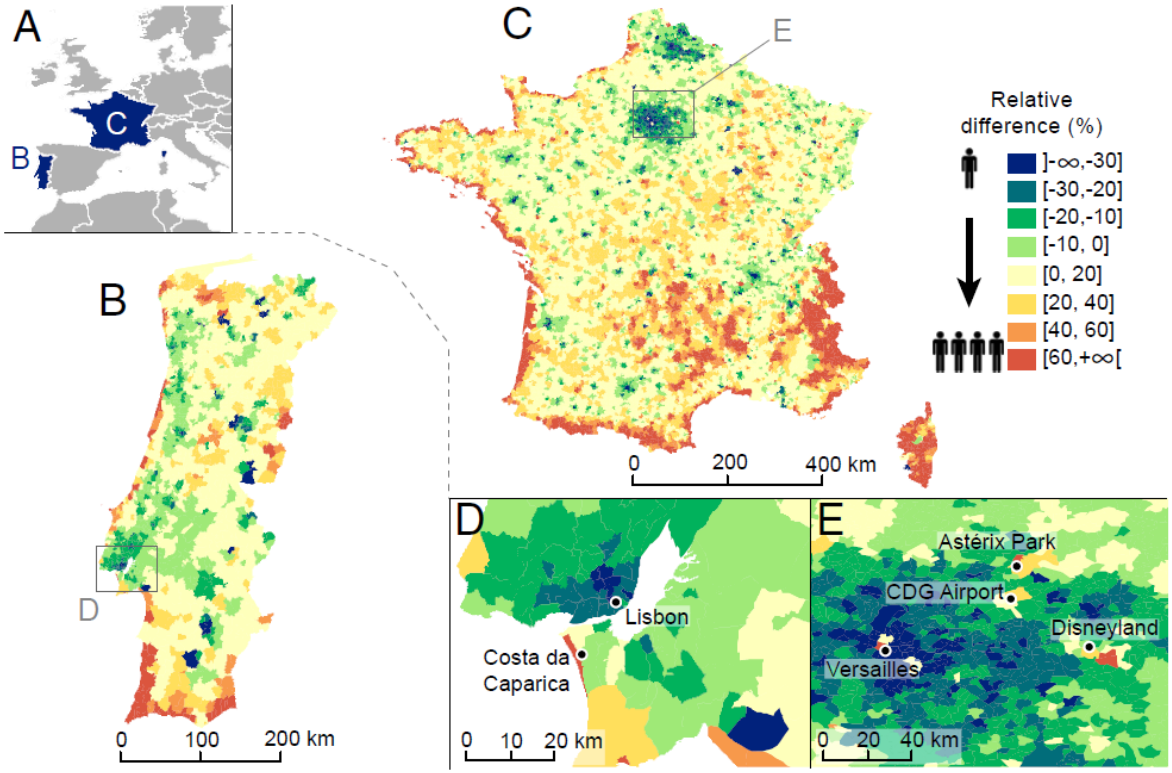


Fig. 3. Seasonal changes in population distribution in Portugal and France. (*A*) Location of Portugal and France in western Europe. (*B–E*) Relative difference in predicted population density between the main holiday period (July and August) and the working period (September to June) by administrative unit level 5 (ADM-5) in (*B*) continental Portugal and (*C*) metropolitan France. (*D*) Close-up around Lisbon with labels showing the city center of Lisbon and the seaside resort Costa da Caparica. (*E*) Close-up around Paris with labels showing the busiest airport in the country (Paris Charles de Gaulle), one of the most visited places in France (Palace of Versailles), and two popular recreation areas (Disneyland and Asterix Park).

Deville et al. (2014). Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45), 15888-15893. doi: 10.1073/pnas.1408439111

Mobility and Call Detail Records (CDR) from mobile phones

- **Track post-disaster displacement**: Haiti (Bengtsson et al., 2011), New Zealand (ACAPS, 2013), Mexico (Moumni, 2013)
- **daily mobility to monitor the diffusion of epAnalyze idemics and effectiveness of various public health measures** to reduce person-to-person contacts in case of pandemic (e.g., swine flu, H1N1, ebola) – (e.g., Frias-Martinez, 2012; Flowminder Foundation West Africa human mobility models)
- **Internal and circular migrations**: Rwanda (Blumenstock, 2012), urban-rural (Eagle et al. 2009; Yadav et al. 2013), impact of socioeconomic status on migration in one Latin American city (Frias-Martinez et al. 2010), predictability of human mobility (Lu et al. 2012; Lu et al. 2013)

Mobile phone usage patterns and type of human activities

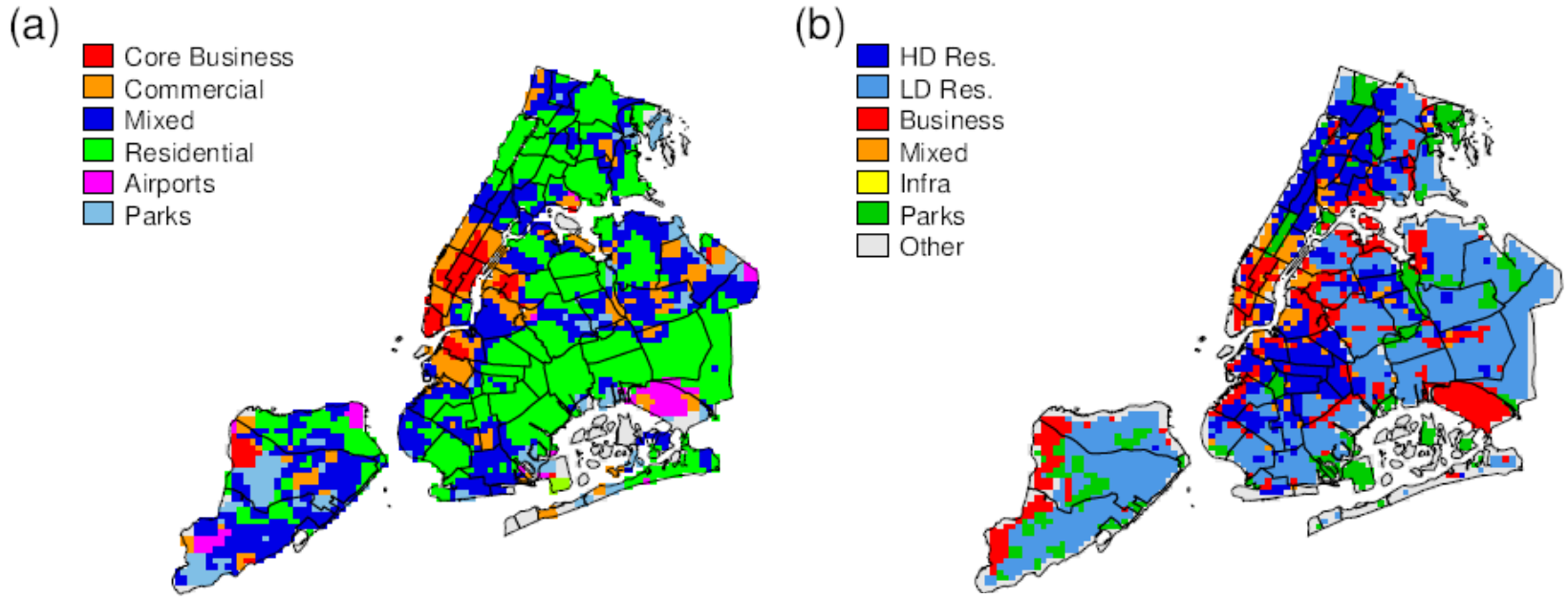


Figure 7: **New York clusters**. **(a)** Spatial projection of $K = 6$ clusters, with their interpretation in legend (see details in main text). **(b)** Actual land use maps as extracted from census data. **(c)** Signatures of the clusters in the different components of activity. **(d)** Deviations of the signatures compared to the whole city signatures displayed on Figure 4. Colors on the signatures plots match those on the cluster map (a), grey areas correspond to zones with no recorded data.

Grauwin, S., Sobolevsky, S., Moritz, S., Gódor, I., & Ratti, C. (2015). Towards a Comparative Science of Cities: Using Mobile Traffic Records in New York, London, and Hong Kong. Computational Approaches for Urban Environments (pp. 363-387): Springer.

Big Data Impacts Society

The more organisations that implement a Big Data strategy, the bigger the impact on the economy will be.

As the McKinsey report famously states:

- Big Data can add $ 300 billion potential annual value to US health care,
- add € 250 billion potential annual value to Europe's Public sector administration
- create a $ 600 billion potential annual consumer surplus.
- Gartner predicts that Big Data will drive up to $ 232 billion IT spending through 2016.
- These are massive gains and investments that of course only occur when all organisations and governments fully start to use Big Data.
- It does show however the impact of Big Data on the global economy, also when only 12% have started implementing a Big Data strategy.

Security/Intelligence Extension: Needs



Security/Intelligence Extension **enhances** traditional security solutions by **analyzing** all types and sources of **under-leveraged data**



Enhanced Intelligence & Surveillance Insight

Analyze data-in-motion & at rest to:
- Find associations
- Uncover patterns and facts
- Maintain currency of information



Real-time Cyber Attack Prediction & Mitigation

Analyze network traffic to:
- Discover new threats early
- Detect known complex threats
- Take action in real-time



Crime prediction & protection

Reduce Customer Churn

Analyze Telco & social data to:
- Gather criminal evidence
- Prevent criminal activities
- Proactively apprehend criminals
- Customer Retention

© 2013 IBM Corporation



Asian Government Agency

National Intelligence Platform

Capabilities:

Stream Computing

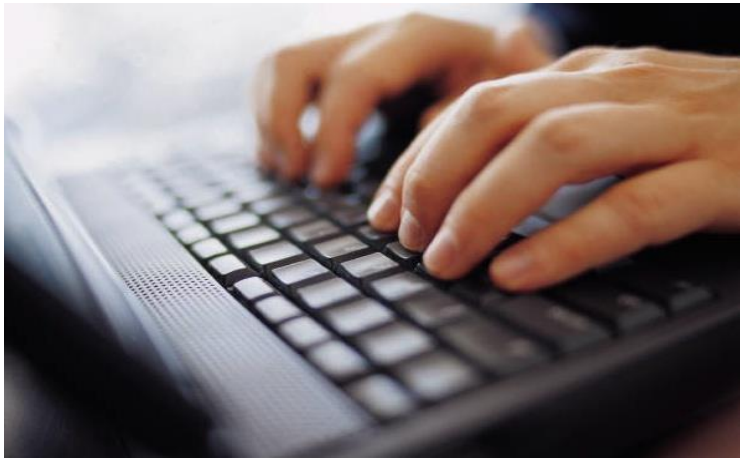
- **Analyze all Internet traffic (social media, email, etc)**
- **Track persons of interest (drug/sex traffickers, terrorists, illegal refugees/immigrants) and civil/border activity**

Potential strength of big data

- Frequent and potential in real time or with short lag
- No cost or low cost
- Often geolocated
- Usually with time stamp
- Potential / optional unique stable ID for matching / linking
- Potentially invaluable insights for longitudinal follow-up (including geolocation)
- Social interactions: ego-centric ties and full network
- Might allow to know more or collect info about life history and vital events
- Any individual attributes linkable?

Big data and privacy concerns

The growing globalization of data flows via **big data** increases the **risk** that **people can lose control** of their own data



No More Secrets era

- The question remains how much the Big Data future is influenced by the

  Snowden revelations which have put data protection and privacy at the center of our attention.

There are no secrets.

- In January 2010, Facebook CEO

Mark Zuckerberg stated that the privacy era was completely over.

to counter terror activities

- In March 2012, CIA chief David Petraeus conceded that the relationship between identity and secrecy should be fundamentally discussed as all secret services have the task of being “world-class Big Data swimmers” in order to counter terror activities:

Potential privacy issues, data protection, and privacy

Despite these challenges, the utilisation of big data could also have huge benefits for society at large. By using demographic and mobility data we can obtain important insights into human behaviour, including:

- ✓ traffic patterns,
- ✓ crime trends,
- ✓ disaster responses,
- ✓ early indicators of poverty and social unrest, etc.

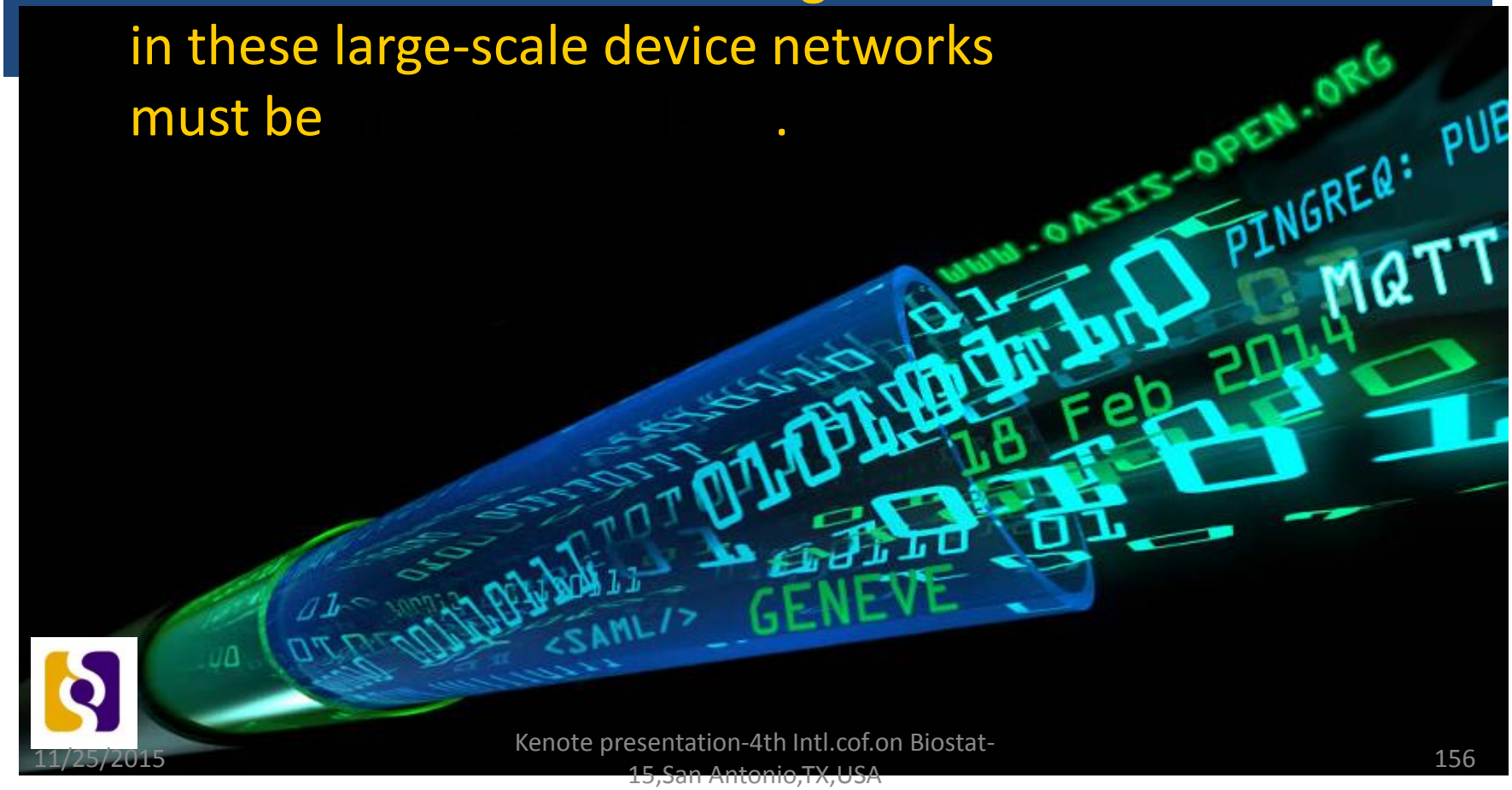
The Challenges

Every one of those sensor and control points is generating data. Often, it's very informative and very private data. Systems are needed to help those devices ***talk to each other***, ***manage all that data***, and ***enforce proper access control***.



Big Data means BIG Challenges

All of the ***messaging***, ***management***, and ***access control*** technologies used in these large-scale device networks must be .



*"The sexiest job for the next ten years will be **statistician**…"*

Hal Varian
Chief Economist, Google
Speaking in 2009



Market Forecast

- **Wikibon projects the Big Data market will top $84B in 2026, attaining a 17% Compound Annual Growth Rate (CAGR) for the forecast period 2011 to 2026.**
- The Big Data market reached $27.36B in 2014, up from $19.6B in 2013.

Opportunity: Expanding Access to Jobs

- **If data are the new oil—**the fuel of the information economy—**the new oil barons will be the data scientists** and knowledge workers, the world will need plenty of them.
- By 2018, the United States is projected to have 190,000 unfilled analytics positions and a shortage of 1.5 million managers and analysts skilled in big data.
- According to the research firm Gartner, by 2015, more than 85 percent of Fortune 500 organizations will be unable to effectively exploit big data.

McKinsey Institute on Big Data Jobs

- By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions.
- IU Data Science Decision Maker Path aimed at 1.5 million jobs. Technical Path covers the 140,000 to 190,000

Big Data and Analytics market

- **IDC predicts the global Big Data and Analytics market will reach $125B in hardware, software and services revenue this year.**
- The professional services-to-technology ratio will also increase 25% this year, due to increasingly complex system configurations and enterprise-wide solutions.
- IDC also predicts the Internet of Things (IoT) market will also attain a 30% CAGR from 2014 to 2019. Source: IDC Predictions 2015: Accelerating Innovation — and Growth — on the 3rd Platform by Frank Gens & IDC Predictions 2015 Team

Concerns/pending issues

- What kind of big data?
- For what purpose?
- Who has access to what kind of information?
- Validation of results
- Issue of comparability of information across space and time
- Transparency, accountability and replication
- Individual rights, privacy and confidentiality

A final thought from W Edwards Dening

"In God we trust…



…All others must bring data."