

Big Data

Data Science



Data Analytics

Data Science

- **Data science** is the “*systematic study of the organization, properties and analysis of data and its role in inference, including our confidence in the inference*”¹
- Requires skill set spanning mathematics, machine learning, AI, statistics, databases, optimization along with understanding of problem formulation.

1. V. Dhar. Data Science & Prediction. CACM 56(12)

Data architecture

How data needs to be routed and organized to support analysis, visualization & presentation

Data acquisition

How data is collected, represented, transformed and grouped prior to analysis



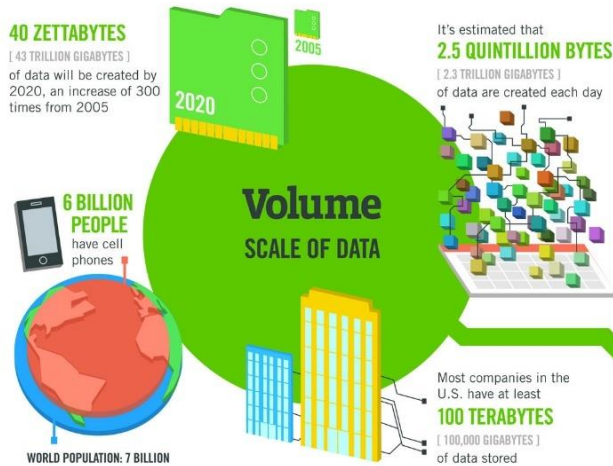
Data archiving

How data needs to be curated and preserved for later use

Data analysis

How data needs to be sampled, processed and visualized

Big Data



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT are shared on Facebook every month



By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month



400 MILLION TWEETS are sent per day by about 200 million monthly active users

Variety DIFFERENT FORMS OF DATA

The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure

Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** - almost 2.5 connections per person on earth



1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions



Poor data quality costs the US economy around

\$3.1 TRILLION A YEAR



27% OF RESPONDENTS

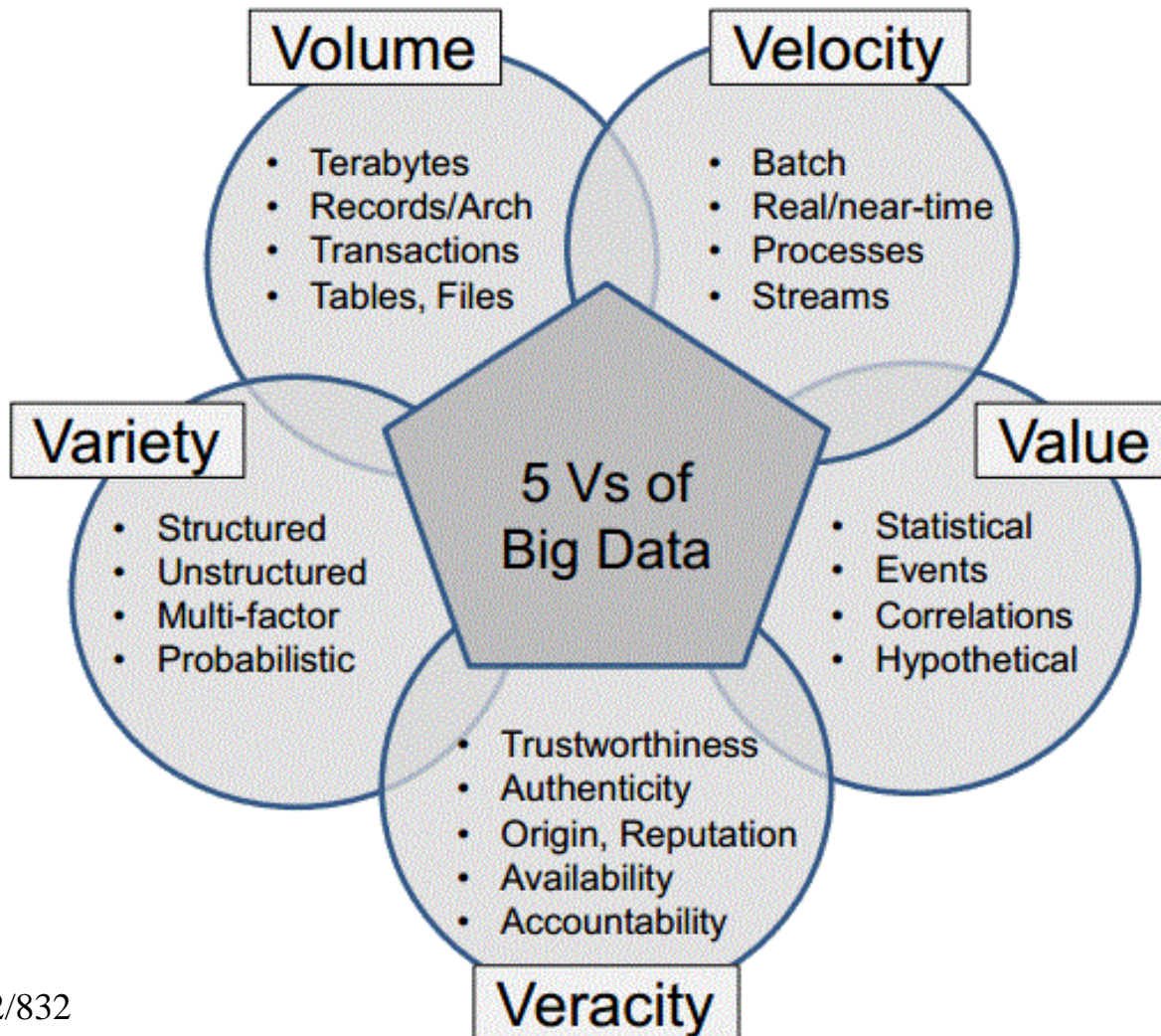
Veracity UNCERTAINTY OF DATA

in one survey were unsure of how much of their data was inaccurate

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPEEC, QAS



Big Data – One More V



What Is Changing In The Realm Of Big Data?



Competitive advantage

Data is emerging as the world's newest resource for competitive advantage.



Decision making

Decision making is moving from the elite few to the empowered many.

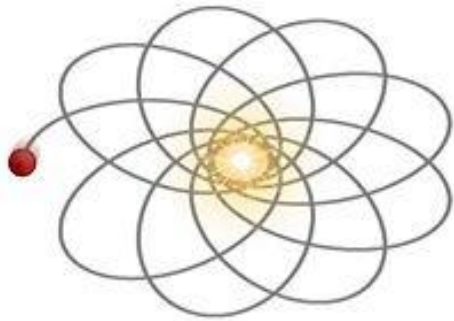


Value of data

As the value of data continues to grow, current systems won't keep pace.

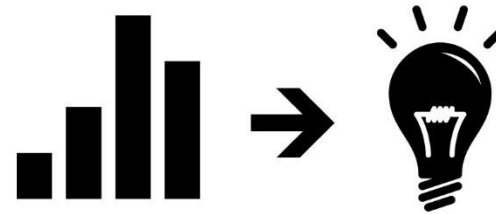
What is Analytics?

Mathematical or Scientific methods that highlight data for insight

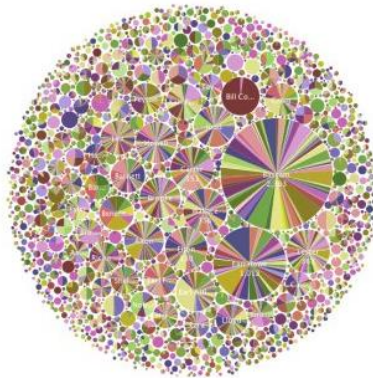


$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = r.$$

Insight = Competitive Advantage
Used to inform actions and decisions



Data is becoming the world's new natural resource



With analytics, insights are created to augment the gut feelings and intuition for decisions



Why Big Data and Analytics?

“The most competitive organizations are going to make sense of what they are observing fast enough to do something about it while they are still observing it.”

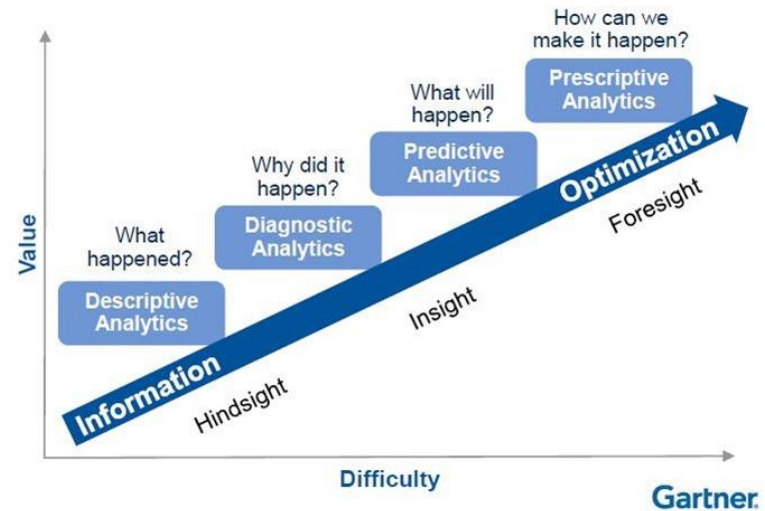
*Jeff Jonas,
IBM Fellow and Chief Scientist,
Context Computing, IBM Corporation*

“... analytics is no longer an emerging field; today’s businesses will thrive only if they master the application of analytics to all forms of data. Whether your office is a scientific lab, a manufacturing company, an emergency room, a government agency, or a professional sports stadium...”

*Brenda Dietrich,
IBM Fellow and Vice President,
Emerging Technologies, IBM Watson*

Big Data and Analytics

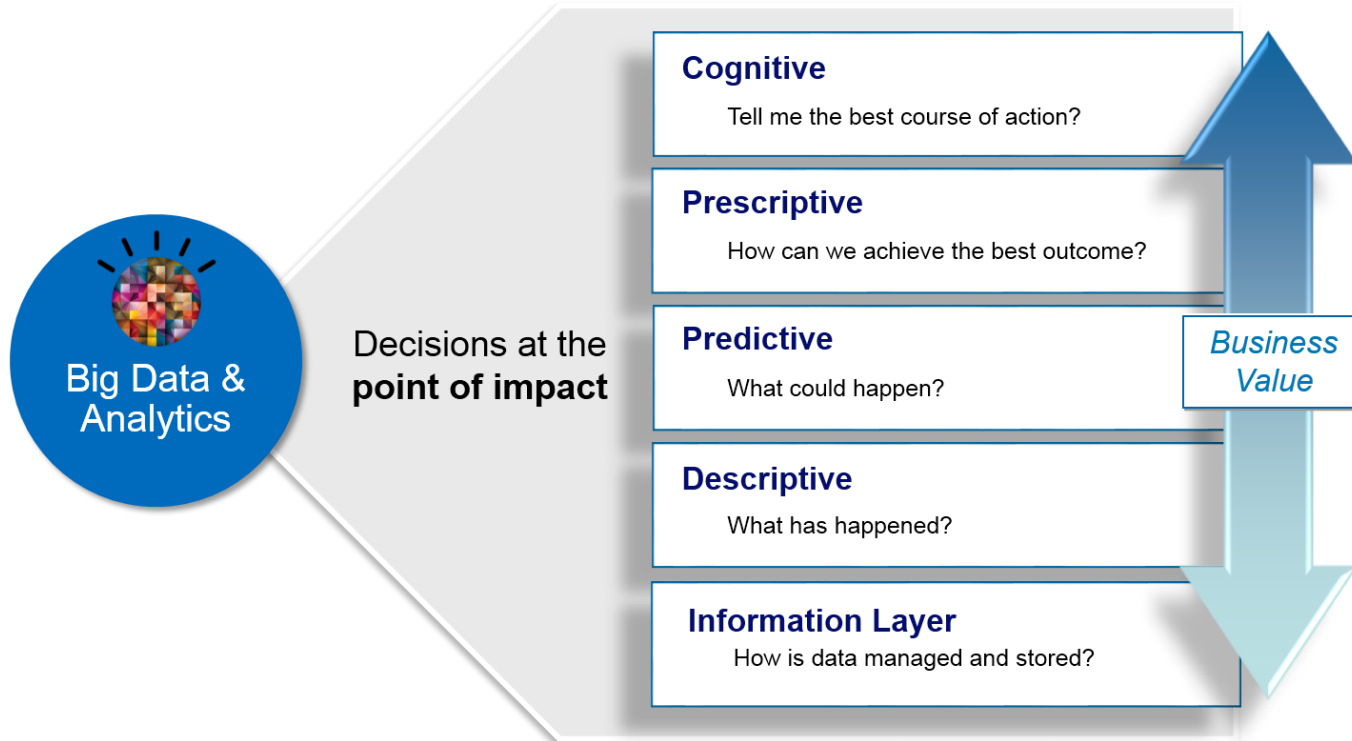
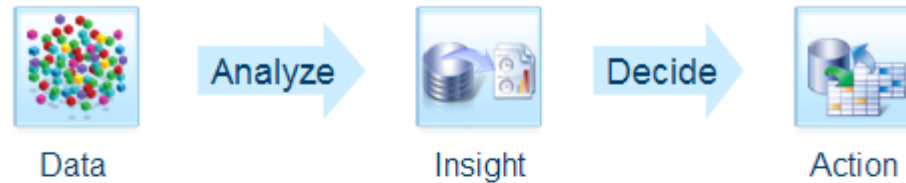
- **Analytics is a progression of capabilities**
 - start with the well-known methods of business intelligence
 - extend through more complex methods involving mathematical modeling and computation



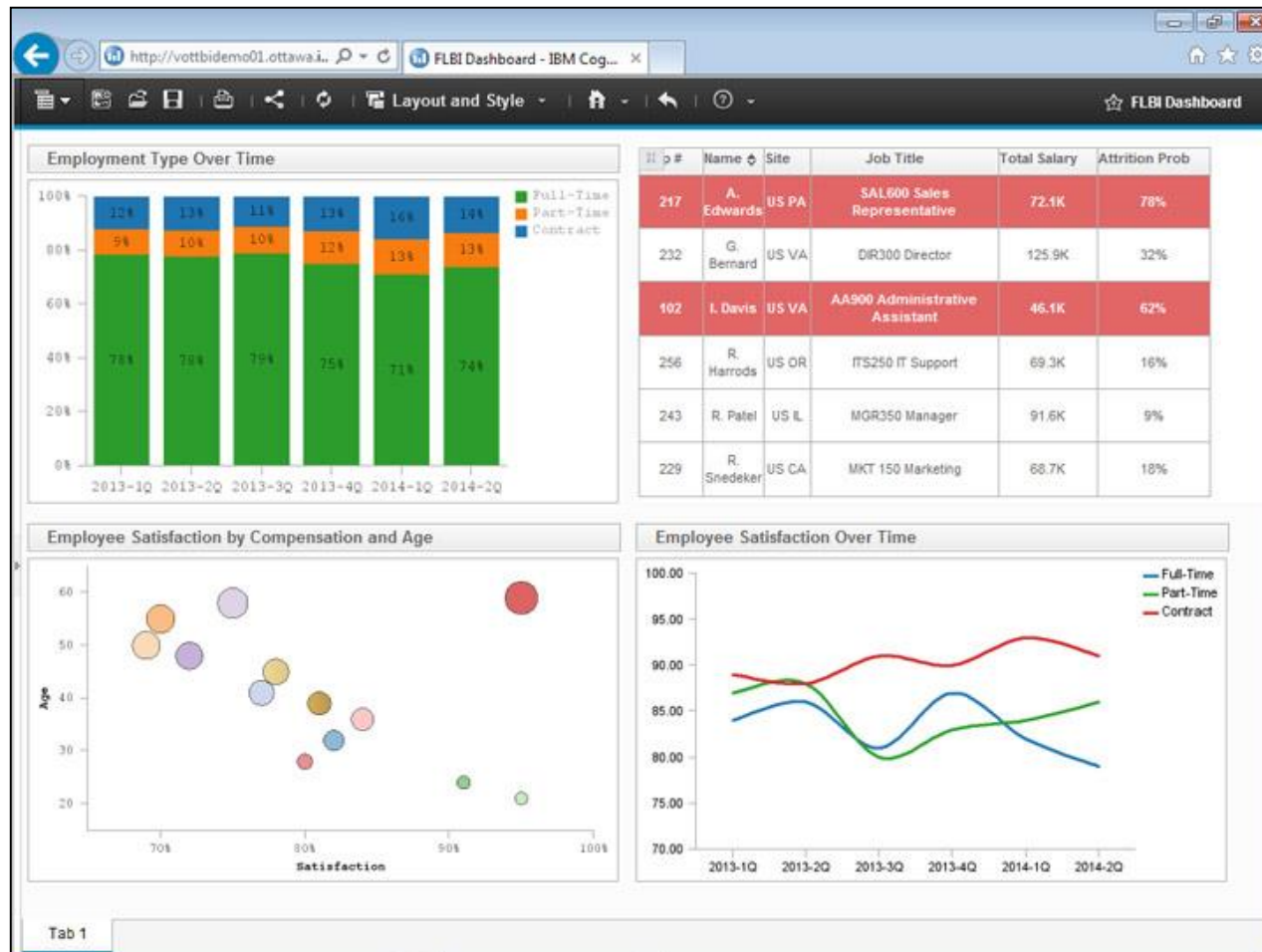
- **Reporting is the most widely used analytic capability**
 - gather data from multiple sources and create standard summarizations of the data
 - Visualizations are created to bring the data to life and make it easy to interpret.



Big Data and Analytics (cont)



Descriptive Analytics – What Has Happened?



Predictive Analytics – What Will/Could Happen?

Helping to predict and stop the spread of infectious disease

Scientists from IBM are collaborating with John Hopkins University and University of California, San Francisco to combat illness and infectious diseases in real-time with smarter data tools for public health by applying the latest analytic models, computing technology and mathematical skills on an open-source framework.



The rise of **global transportation, trade and climate change** allows insects to easily carry disease organisms like **malaria and dengue fever** across borders, infecting animals as well as humans.

Malaria is responsible for nearly

1 million annual deaths¹



Dengue fever has spread to over

100 countries, including the United States

Finding and implementing new, innovative methods of predicting outbreaks is key to **saving lives**.



Data from the World Health Organization showing general sensitivity to outbreaks

Addition of population models (including humans and mosquitos)

Integrating climate data for temperature and precipitation can more accurately pinpoint potential outbreaks

These model results make it possible to **predict where disease incidence** is most likely to increase or decrease.

Public health organizations can then be informed, and specific **intervention strategies** can be developed in locations where they can have the greatest impact.

432/832



Tackling Rugby Injuries with Analytics

1 in 4 rugby players will be injured during a season

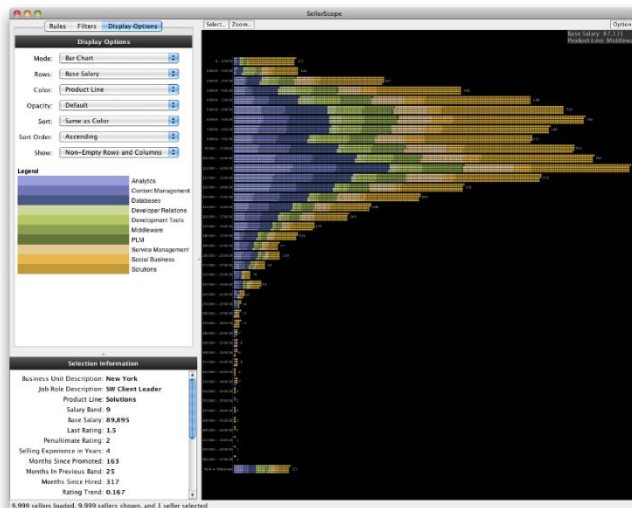
Harassing injuries cost a rugby team an average of **14** playing days

Researchers are using math equations to predict sports injuries

Organizations that apply predictive analytics are **2.2x** more likely to outperform peers

1. "Rugby: A New Era of Analytics" - IBM Business Analytics Journal
 2. "Predicting Sports Injuries with Analytics" - IBM Business Analytics Journal
 3. "Rugby: A New Era of Analytics" - IBM Business Analytics Journal
 4. "Outperforming with Analytics" - IBM Business Analytics Journal

Prescriptive Analytics – What Actions can be Taken?



Engaging with Customers as Individuals is a Growing Challenge

4 in 10 smartphone users search for an item in a store

Mobile commerce poised to reach \$200 billion by 2015

Mobile device use set to exceed desktop internet usage 2014

Online sales via mobile devices up 300% over 2010

57% of CEOs looking to use social media platforms to connect with individuals within the next three to five years (IBM CEO study)

In social media alone every 60 seconds

600 new blog posts are published

34x 1000 tweets are sent

IBM

Social Media Analytics

IBM Register to download the 2012 CEO Study

Do you think like an outperformer?

Looking ahead, what will be your most critical source of insights?

Financials Customers

73% of CEOs told us they identify customer insights as the most critical investment area.

Face-to-face interactions and social media
Traditional media and call centers

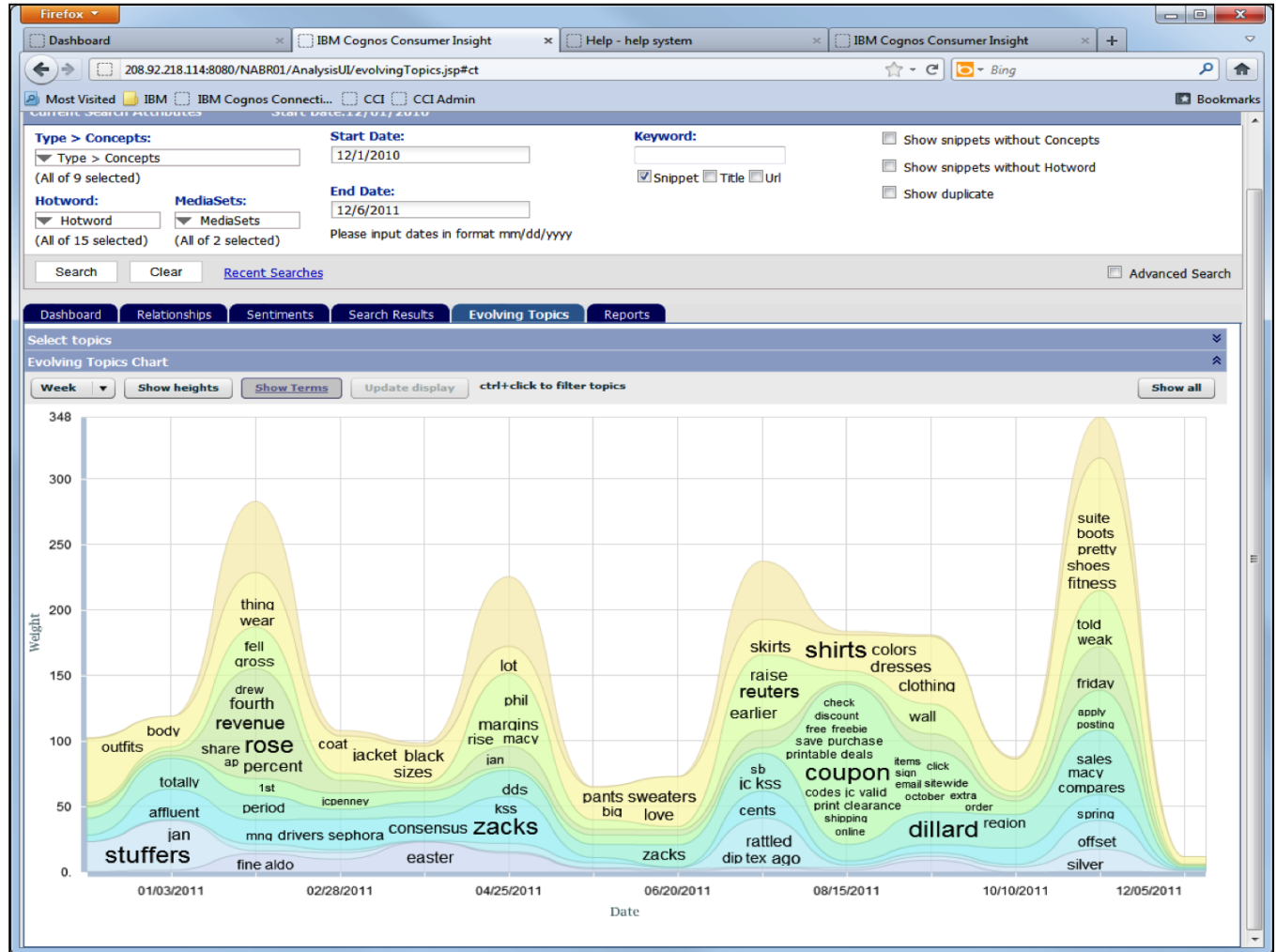
256% increase in social media engagement

Outperforming CEOs distinguish their organizations in three ways:

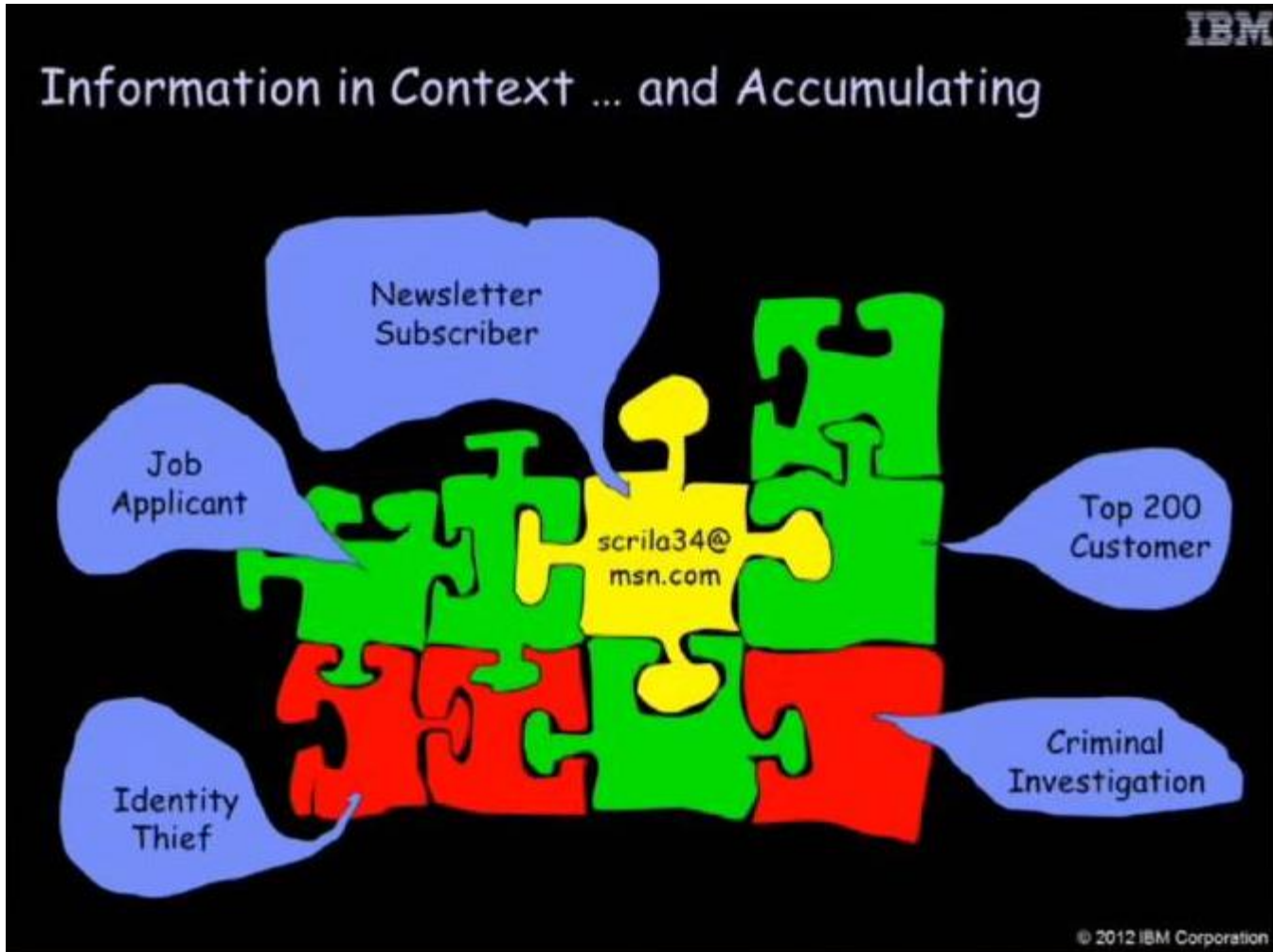
- Ensuring increased access to data (108% more for outperformers)
- Successfully deriving insights from data (108% more for outperformers)
- Translating these insights into action (84% more for outperformers)

432/832

Source: 2012 IBM Global CEO Study. Register to download the 2012 CEO Study



Entity Analytics



Why analytics matter

CIOs rank analytics as the **#1 factor** contributing to an organization's competitiveness.¹

8 out of 10 CEOs expect complexity to increase significantly in the next five years.²



Financial outperformers are **64%** more likely to use analytics to evaluate talent supply and demand on an ongoing basis.³



Enterprises that apply advanced analytics have **33%** more revenue growth and 12X more profit growth.⁴



Organizations that embrace analytics are more than **2X** as likely to outperform their peers.⁵



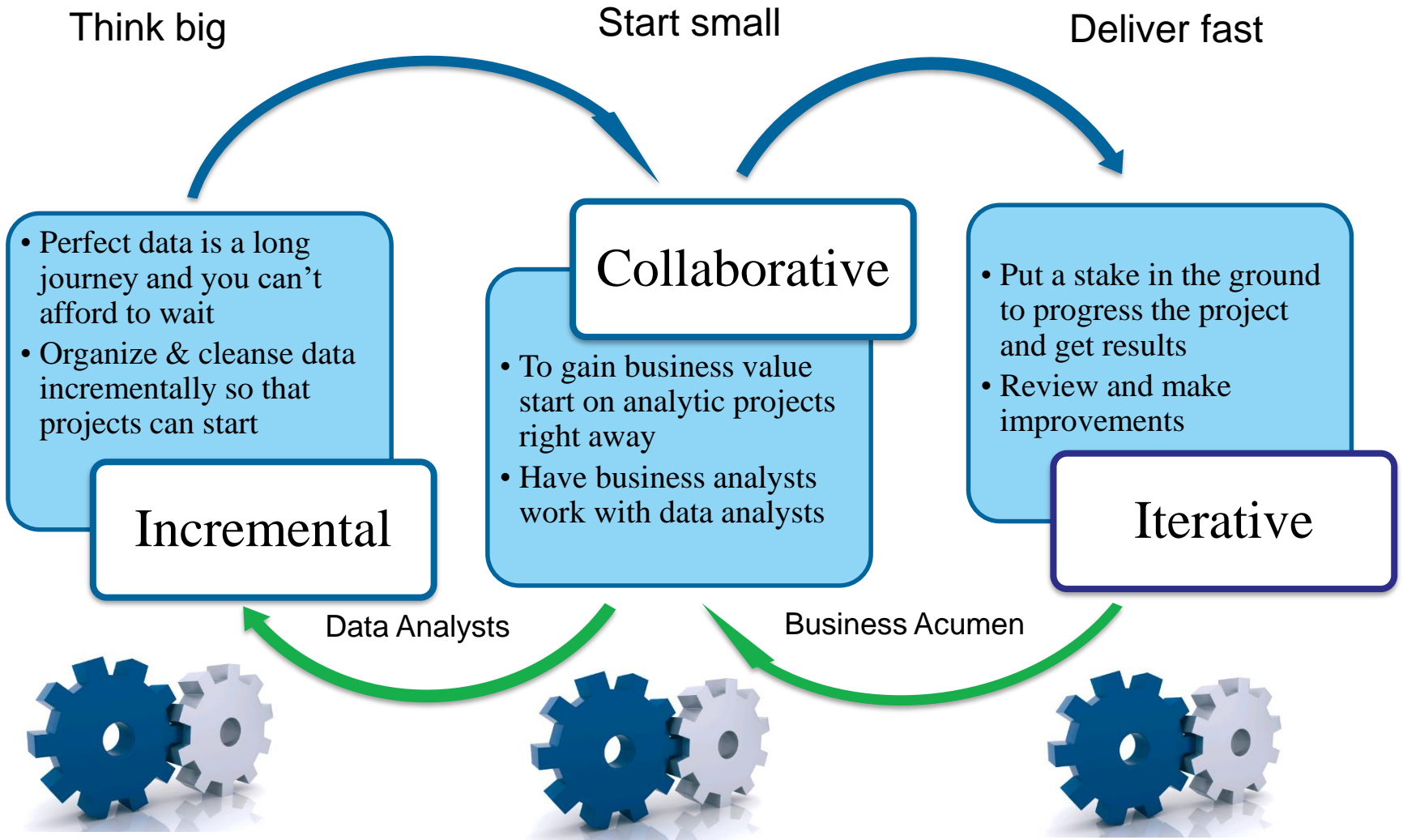
Top-performing enterprises use business analytics **5X** more than lower performers.⁶

Human brains were not built to process the amounts of data that are today being generated through social media, sensors, and more.

The Importance of Analytics in Decision Making?

- Decisions are based on a wide variety of factors
 - Includes personal experience, intuition, testing or analytics and data
- Employing analytical decision making:
 - DOES NOT guarantee a positive outcome
 - CAN provide a small edge to those who employ them
- Traditional analytics in business support internal decision-making
 - What should we charge for a product?
 - What promotion will make this customer buy from us?

The Importance of Analytics in Decision Making?



The Importance of Analytics in Decision Making?

- Decision makers evaluate information from two sources
 - Quantitative analysis
 - Non-quantitative sources
 - Intuition – should we trust it?
 - Experience – is it representative of broad situations?
 - Rules of thumb
 - Hearsay
 - Guessing – risky!

What Types of Business Decisions Can Be Made Analytically?

MARKETING



- Pricing
- Locations of stores and branches
- Promotion targeting
- Web site customization
- Advertising placement

Supply Chain

- How much inventory to keep
- Where to place distribution centers and warehouses
- Routing of products or vehicles
- Truck loading

Finance



- Drivers of financial performance
- Performance scorecards
- Forecasts of various types

Human Resources



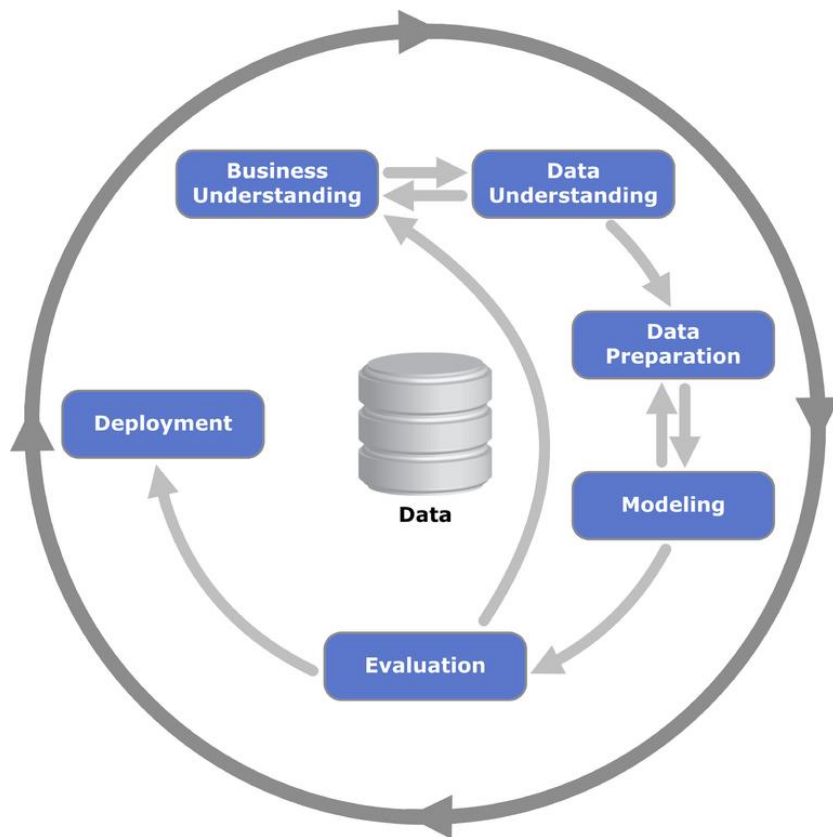
- Which employees to hire
- Which employees are likely to leave
- How much to compensate employees
- What education would benefit an employee the most



- Which product features are most desired by customers
- How effective a particular product is
- Which product design is most appealing

CRISP-DM

(Cross Industry Standard Process for Data Mining)



- **Business Understanding**
 - Understanding the project objectives and developing plan to achieve them
- **Data Understanding**
 - Data collection and activities in order to get familiar with the data,
- **Data Preparation**
 - Activities to construct the final dataset

CRISP-DM

(Cross Industry Standard Process for Data Mining)

- **Modeling**
 - Various modeling techniques are selected and applied
- **Evaluation**
 - Evaluate the model and confirm it properly achieves the business objectives.
- **Deployment**
 - Depending on the requirements, can be as generating a report or implementing a repeatable data scoring or data mining process.

