The Software Alliance

BSA

What's the Big Deal With Data?









Executive Summary

Software innovation continues to spark unprecedented advances that transform the world around us, empower us as individuals, and grow our economies.

Yet the full potential of this digital transformation can only be realized if we tap the potential of the data these innovations have unleashed. We are, in fact, living through a data revolution. Driving this is not only the abundance of data today, but the fundamental technologies that change the way we gather, store, analyze, and transform information.

Today, 90 percent of business leaders cite data as one of the key resources and a fundamental differentiator for businesses, on par with basic resources like land, labor, and capital.

Not long ago, for instance, data collection required observing weather patterns over hundreds of years to discern rainfall patterns. It meant sitting alongside a road to log traffic speed to plan transportation networks. It involved gathering miles of handwritten notes to study how diseases work and could be cured.

Now, data is generated by sensors on millions of devices, machines, vehicles, and even street lamps. While keeping this amount of data was once costly and difficult, storage capacities have grown and costs have plummeted, making stored data a renewable resource. With this ability to reuse and repurpose data, we can continue to analyze and transform it in new ways that produce valuable insights that save time, money, and even lives.

Some of this captured data is personal information, and as such, both cutting-edge security and responsible stewardship models must be used to make sure this information is safe and correctly used. But the vast majority of data comes from the many devices and machines reporting to each other and to those running them. From the assembly line at the manufacturing plant to the passenger jet in flight, millions of bytes of data are generated and then analyzed. Doing so helps improve performance and boost productivity in ways once unimaginable.

While data is everywhere and its ubiquity and utility are improving our lives in so many ways, many do not understand what it is, where it comes from, how it can be used, and its inherently massive potential.

This paper outlines just a few concrete examples of how data innovation is driving extraordinary progress on some of the world's toughest challenges. It describes how fundamental changes in how data is gathered, stored, analyzed, and transformed place us at the brink of all that is possible in our 21st-century digital economy — and beyond. It also addresses some of the myths that have become prevalent as people continue to work toward fully understanding the expanding data innovation industry. Finally, the paper offers a glossary of terms defining the language of data innovation to serve as a guide for those new to understanding the data economy.

The opportunity that data innovation presents the world is virtually unparalleled. Innovative software tools already are revolutionizing our lives in amazing ways; now, these tools are helping people unlock the answers hidden within a growing abundance of data resources. These transformative new tools are translating data into new products, new solutions, and new innovations that stand to change our lives. From an economic perspective, making better use of data could lead to a "data dividend" of $1.6 trillion in the next four years alone. Economists estimate data-enabled efficiency gains could add almost $15 trillion to global GDP by 2030.

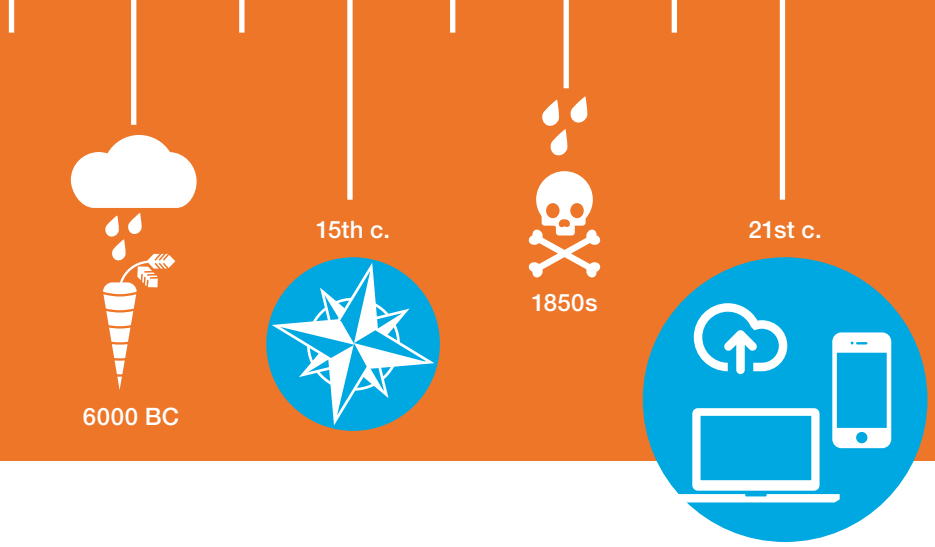
If we make smart choices today, this emerging "data-centric" economy can become a powerful generator of new jobs and industries, new breakthroughs, and new cures — and will fuel economic growth for decades to come.

DEFINING 'DATA INNOVATION'

A good deal of ink has been spilled on the "Four Vs" of data innovation: volume, the amount of data; velocity, the speed at which it is created; variety, the types of data involved; and, veracity, its accuracy. Yet less time has been spent discussing how little value there is in raw data – and the game-changing opportunity we all share to truly maximize its use.

As this paper examines, data must be gathered, stored, analyzed, and transformed to provide benefits ranging from practical to lifesaving. These processes are at the heart of data innovation – the derivation of immense value from the vast amounts of otherwise unproductive information.

Contents



Introduction

Throughout human history, the mileposts of civilization have been punctuated by advances in our ability to observe and gather information. Our ancestors developed tools to measure distance, weight, volume, temperature, time, and location — each improving over time, and each critical to the movement from hunter-gatherers, to farmers, and to city dwellers.

As early as 6000 BC, we used data about crop yields and fallow farming to boost farm outputs and feed more people. In the 15th century, we used data from the skies to navigate our world and open the high seas to global trade. In the 1850s, we used data to link cholera outbreaks to bad water and saved lives.

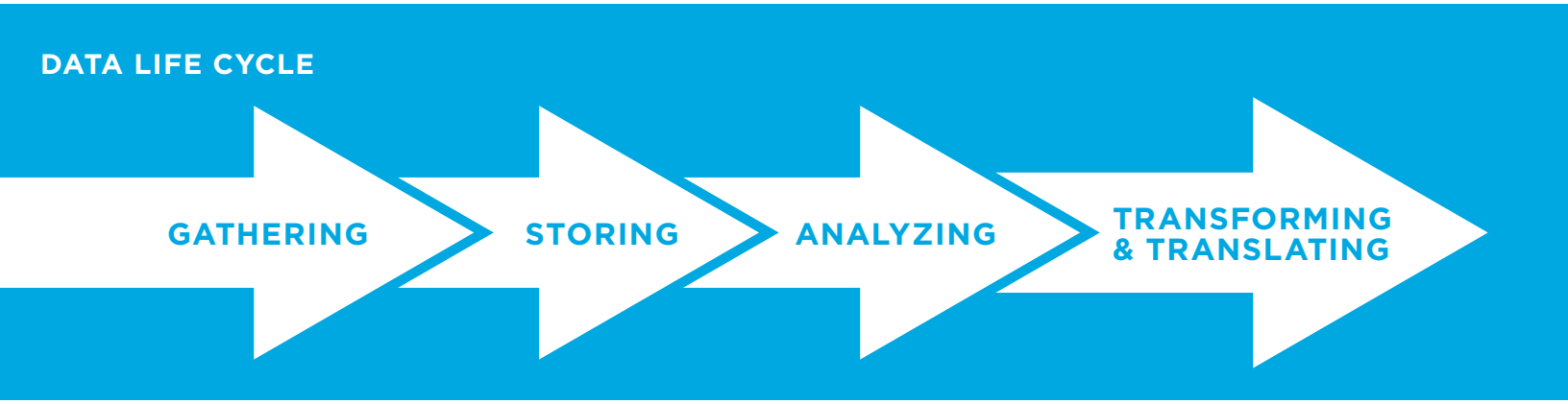
Throughout modern history, even limited amounts of data have provided us with key insights for unexpected solutions to some of our greatest challenges. Whether recorded on a stela, a papyrus scroll, in an illuminated volume, or in a printed book, data — and its increasing prevalence and prominence — has been a key driver of economic and human progress.

In the 21st century, we are undergoing a rapid acceleration of this process. As data becomes more abundant and the cost of data storage plummets, new technologies are equipping data scientists with cutting-edge tools that unlock valuable insights from vast amounts of data. As those technologies that process data become more transformative, their impacts become more profound and opportunities even more pervasive.

We are heading toward a world of almost boundless information and nearly limitless possibilities. Consider how data is being used to make predictions that enhance our daily lives. Predictive data helps us know in advance whether to bring an umbrella to work or take the bus. Traffic data is used to synchronize traffic lights, predict train arrival times, and help us find the fastest route to get to a child's rehearsal on time. Wearable devices help us track our personal fitness so we can make smarter choices to live longer, healthier lives, and scientists are analyzing terabytes of genetic information to find new cures and develop more effective, personalized treatments.

Data Making a Difference

+ **Barcelona** is harnessing data to build a smarter city, giving it the ability to examine the traffic patterns of tourists, see where to put more public bike stations, and identify which corners of the city need more ATMs.
+ In the **United Arab Emirates**, new data tools are being used to design the world's first positive-energy building that actually produces more energy than it consumes.
+ In **Kenya**, mobile data is being used to identify malaria infection patterns and identify hotspots that guide government eradication efforts.
+ Farmers from **Iowa** to **India** are using data from seeds, satellites, sensors, and tractors to make better decisions about what to grow, when to plant, how to track food freshness from farm to fork, and how to adapt to changing climates.



When buying a car, instead of mere access to a car's sticker price, data provides us insight into a car's fuel economy, maintenance, insurance, and safety records to help us make more informed choices. And your car itself is now, in effect, a supercomputer on wheels. It has a processor that is interacting with sensors that analyze performance so drivers can be informed when to get an oil change, shift to an electric motor, or if there is a child playing in the driveway as the car backs up.

Already, this growing abundance of data helps put power in our hands by putting much-needed information at our fingertips.

But what exactly is "data"? Who or what is generating it? What is its potential to improve our lives? How must it be used for maximum benefit? And how do we make sure it is used in a way that is consistent with our values and concerns?

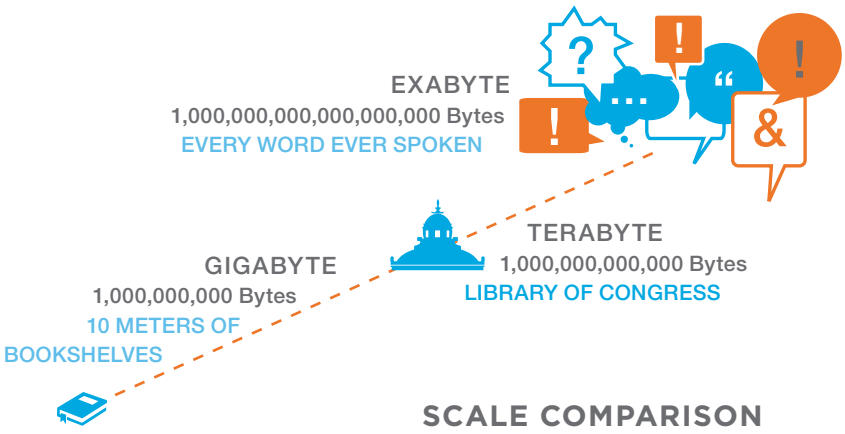
These are important questions in that as data transitions from a once-scarce resource to an increasingly abundant, valuable, and renewable resource, it is becoming a primary source of economic and societal benefits. Historically, it has been access to resources like land, labor, and capital that provided the economic differentiator between those who succeeded and those who failed. Today, 90 percent of business leaders cite data as one of the key resources and a fundamental differentiator for businesses, on par with basic resources like land, labor, and capital.[1]

One example: economists conservatively estimate that if harnessing data more effectively achieved small gains making industries just 1 percent more efficient, that would add nearly $15 trillion to global GDP by 2030.[2] The "next big thing" may come from the billions of small things connected to the Internet producing better data about the world around us to enable even more powerful data-driven solutions.[3] Already we are finding answers to questions we didn't even know we had.

This huge shift is underway. Almost everything we do generates data, and entirely new streams of data are being created every day. In fact, 90 percent of the world's data today has been created in the last two years alone, and we are now doubling the rate data is produced every two years. Most of this data being generated is not personal data. It's an important distinction because while it is imperative that we protect privacy, more often than not the data that is helping improve our lives was generated by a sensor attached to a machine.

Our challenge is to harness data and put it to work, using our ingenuity to make sense of the valuable learnings locked within it. It is this ability to process data and transform observations into insights, and insights into answers, which enables us to achieve meaningful solutions to today's significant challenges.

Businesses and governments must now work actively to crank up the innovation engine.



THE FOUR KEY STEPS FOR

Transforming Data Into Answers

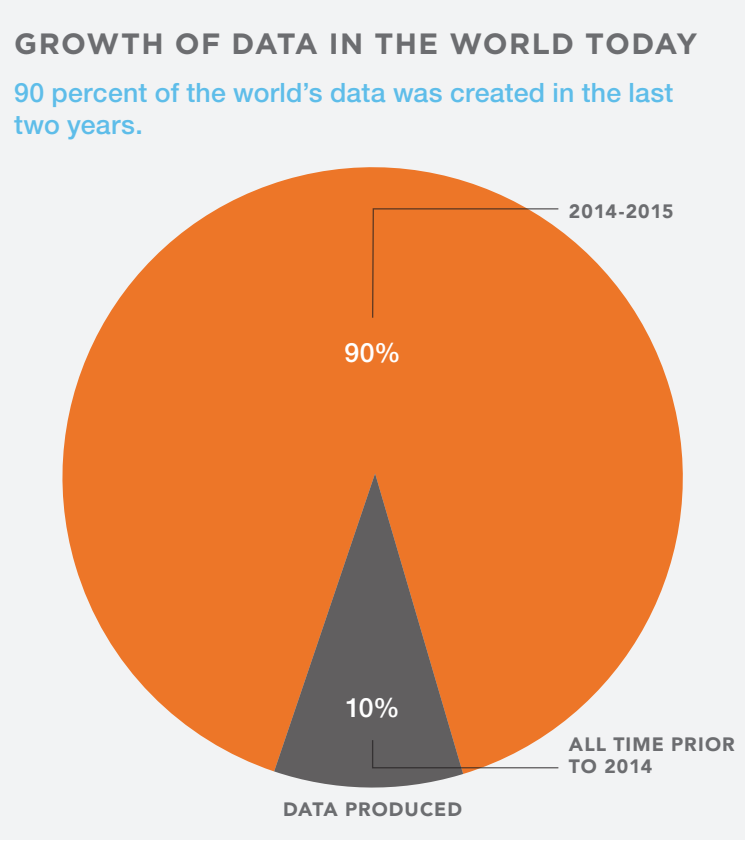
Today's data revolution isn't just being driven by the growing abundance of data; it's being fueled by fundamental technologies that change the way we gather, store, analyze, and transform data. Together, these drivers are enabling us to glean powerful insights from deep within data and thereby unlock new knowledge, discover new connections, and make new predictions.

1

GATHERING DATA

Throughout history we have always gathered data and used it to help advance society. But often, data was too scarce. Today, we are fortunate to be able to harness more data from the world around us — data infused with more meaning, gathered in more useful forms, and producing more deliberate results. Data has gone from a once-scarce resource to an increasingly abundant, vital, and renewable resource.

This has been made possible not by our ability to consolidate and mine personal information; rather, our ability to connect a variety of devices and sensors to the Internet now generates a wealth of new data growing at exponential rates. As a result, data is produced



Source: IBM
http://www.ibm.com/software/data/bigdata/what-is-big-data.html

everywhere — by roadside sensors used to measure traffic flow; by the digital music and movies we create; by the satellites that circle overhead; by the sensors and systems that control our factories and financial markets; and by the tools we use to digitally design the next new thing. This information is growing faster, flowing farther, and increasing in significance.

Already an estimated 2.5 quintillion bytes of data are generated every day.[4] Our analog-era minds have trouble even contemplating the enormity of this data. For context, last year the world created enough digital data

Most data isn't personally identifiable. Soon, connected devices around the planet will help us better understand and improve the world around us.

to fill a stack of DVDs that would stretch from Earth to the moon and back.[5] And the pace at which we are creating data is accelerating, too. The volume of business data worldwide, across all companies, is now doubling every 1.2 years.[6] Where does it all come from? Below are just a few examples of sources, among many:

+ Digital information in hospitals, largely from clinical imaging, is expected to climb to 665 terabytes a day by 2015 — helping find cures and save lives.[7]
+ Modern transcontinental airlines are so packed with connected sensors on their engines, flaps, and landing gear that they can generate half a terabyte of data per flight to improve flight performance,[8] cut turbulence, improve safety, and identify possible engine defects 2,000 times faster than before.[9] Multiply that by the more than 25,000 flights flown each day and you get a sense of the vast amounts of helpful data now being generated just from commercial jets.
+ Weather satellites, weather observatories, radar, and other sensors capture more than 2.25 billion weather data points 15 times per hour — collecting 20 terabytes per day — making more accurate weather predictions around the globe possible.[10]
+ Financial exchanges generate four to five terabytes of data a day used for real-time analytics and spotting problematic trading activity, while helping grow businesses and a more prosperous economy.[11]
+ Telematic sensors in tens of thousands of delivery vehicles track engine performance, improve routing, and anticipate problems in advance. Vehicle sensor data combined with mapping data analytics has enabled companies to save millions of gallons of fuel and reduce emissions by the equivalent of taking thousands of cars off the road for a year.[12]
+ The Large Hadron Collider at CERN, the European Organization for Nuclear Research, generates 40 terabytes of data every second of every experiment, providing new insights into the deepest secrets of how the universe works.[13] Likewise, Chile's Large Synoptic Survey Telescope generates 30 terabytes of data about our universe looking at the sky every night.[14]
+ The sequencing of a single DNA genome can generate 200 gigabytes of data. As the cost of DNA sequencing plummets, scientists are building massive databases filled with hundreds of thousands of these sequences in order to find the differences and similarities that correlate to medical breakthroughs and save lives.[15]

It's not just the amount of data that is exponentially increasing, but also the ways in which it is produced. As the number of devices that connect the Internet to the world around us increases, creating an "Internet of Things," a multitude of sensors are creating entirely new forms of data each day. The next big thing may be based on many small things as an estimated 50 billion devices packed with powerful sensors are projected to be connected to the Internet by 2020.[16]

These devices will create data by doing things like measuring soil moisture, engine performance, energy system efficiency, and the location of asthma attacks. As humans, we use just five senses to understand the world around us. Soon, connected devices around the planet will sense a whole range of features about the physical world in order to help us better understand and improve the world around us — and in so doing, produce exabytes of new, beneficial data.

Because the cost of data storage keeps plummeting and the amount of data keeps growing, the uses of data keep expanding.

2

STORING DATA

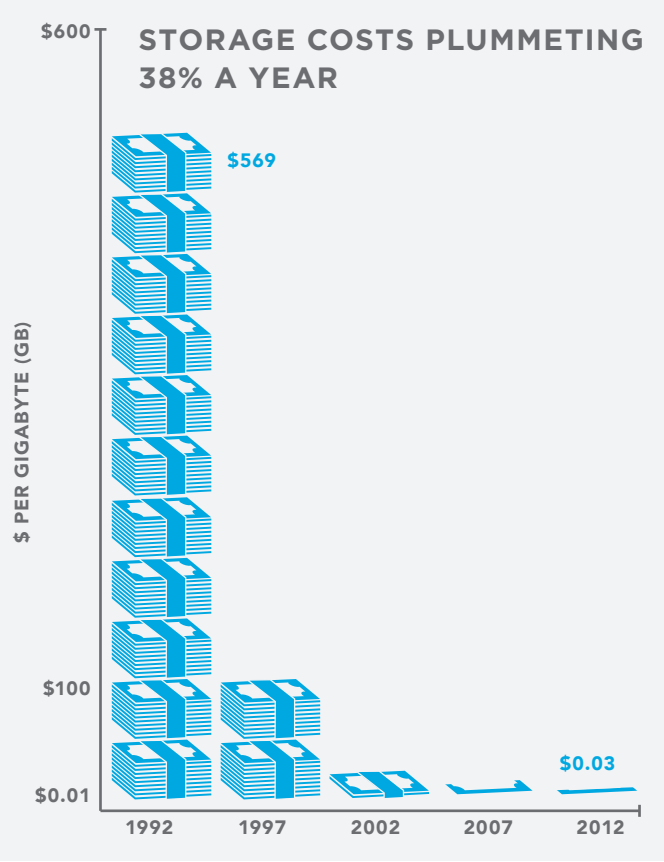
The plummeting cost of storage is enabling data-driven innovation. In 1980, a gigabyte of data storage was scarce to come by, cost hundreds of thousands of dollars, and required a full-time person to manage.[17] Today, a gigabyte of storage costs just pennies, is managed easily, and can be accessed anytime, anywhere.[18] Since the 1980s, the price of storage has dropped by more than a factor of 10 million.[19] To put that in context, if gasoline prices had fallen by the same amount, you could drive a car around the world nearly 10,000 times for what you paid for a gallon of gasoline in 1980.[20]

Because the cost of storage keeps falling, we have been able to store ever-increasing amounts of data. In 1994 only 3 percent of the world's data was stored digitally.[21] By 2007, 94 percent was stored digitally.[22]

Because the cost of data storage keeps plummeting and the amount of data keeps growing, the uses of data keep expanding. Cloud technologies — allowing data to be stored remotely and accessed by an array of devices — have dramatically driven down the price of data storage so that data no longer has to be deleted after its initial use to make room for new data.[23]

As a result, unlike other resources, data is not used up when it has been used once. Data is a renewable resource that can be combined with other data sets and used multiple times to produce answers to questions that were unforeseen when the data was originally created. For example, weather data isn't just used to predict if we need to carry an umbrella; it can also help predict crop yields.

Cloud technologies give users better, more reliable, more affordable, and more flexible access to their data — relieving the pressure that vast amounts of data can place on in-house IT infrastructure. By fundamentally transforming the way data storage is bought, sold, and delivered — and by making data available virtually anywhere at any time — cloud technologies are emerging as one of the most transformative technologies of the decade, and one of the greatest enablers of data-driven solutions.



Source: Hagel III, John et al. *From Exponential Technologies to Exponential Innovation.* Deloitte University Press, 2013. Print. 2013 Shift Index Series.

REAL-TIME DATA: REDUCING TRAFFIC

3

ANALYZING DATA

Data is only valuable when it is understandable; otherwise, it's just a jumble of random observations. Making sense of the insights contained within data can only be achieved by combining human ingenuity with innovative software.

Despite an increasingly autonomous world, it still takes personal curiosity, human skills, and intensive work to unlock answers from within data.

First, raw data needs to be cleaned up to be made useful. By one estimate, data scientists can spend from 50 percent to 80 percent of their time preparing unruly digital data before it can be explored for useful nuggets.[24]

Second, it takes human creativity to ask the right questions and then find answers by sorting through and recognizing bad data and interpreting the results in meaningful ways. The data scientist role has been described as part analyst, part artist, and part storyteller.[25] Each individual piece of data is like a pixel on a screen. Alone, it provides only a small amount of information. But when combined with enough pixels in the right order, a data scientist can paint a picture worth a thousand words and derive new and sometimes unexpected meaning from the data.

By sifting through data, analytic tools can help cut through data clutter to help users discover new patterns and trends, find unexpected insight from seemingly unrelated data, and automatically uncover statistically interesting relationships. Using increasingly rich databases and ever-advancing statistical algorithms, software analytic tools enable us to sift through mountains of data to find the nuggets of information gold.

Fortunately, more powerful processing capabilities in today's computers combined with inventive software are empowering data scientists with cutting-edge tools to make sense of vast amounts of data and unlock the valuable insights contained within it.

While today's networks are impressive, moving huge amounts of data across networks into one location in order to process it all at once is often economically unaffordable and logistically impossible. Yet some of the more powerful analytics engines today are being made possible and affordable through massive parallel-distributed cloud computing platforms. These platforms allow users to run world-class data analysis tools across data stored in multiple locations at the same time.

What does this data analysis enable us to do? Predicting the future used to seem far-fetched, but now seems inevitable. Today, thanks to modern data analysis, we make reliable predictions all the time. Weather forecasts have become more reliable even as far as 10 days out. Fleet managers can predict which engines need fixing before the car breaks down. When data from the present can be compared to that of the past, it often can be used to help predict the future.

Economists are finding ways to better forecast markets, employment, and inflation. For too long, government economic data has forced decision-makers to look in the rear-view mirror. Government economic statistics, like GDP growth, have always looked months behind to tell us, after a long delay, how economies performed in

IBM and the city of Stockholm have partnered to install 1600 GPS systems in taxis. The data from the GPS devices is crunched using IBM streaming software and used to give insights on traffic flow, travel times, and optimal commuting routes.

Decreased the amount of emissions by **10%**

Reduced traffic in the city by **20%**

Reduced average travel times by almost **50%**

Increased the proportion of green, tax-exempt vehicles to **9%**

Source: Bertolucci, Jeff. 'Dublin Points Big Data Tech At Traffic Jams'. InformationWeek 2013. Web.
Nusca, Andrew. 'Stockholm Uses Real-Time GPS Data To Manage Traffic Congestion *ZDNet*, 2010. Web.

the past as the best benchmarks for the future. Now, economists are combining a variety of real-time data, like new job postings and industry orders, and comparing them to historical data in order to paint a more accurate picture of today's dynamics, and formulate better policies to ensure healthy economies.

The rise of real-time data analytics also is enabling autonomous decision-making to help us, or machines we run, make decisions far more quickly and with greater precision. Already, major American auto companies are designing new vehicles packed with hundreds of sensors, telematics, and real-time connectivity to enable such advances as autonomous parking. Automakers also are advancing the real-time analytics tools that enable autonomous crash avoidance and self-driving cars. Such advances may one day save lives by reacting to situations faster and more reliably than humans can.

With an exponentially growing amount of real-time data about the world around us, those who are able to make sense of what they learn as fast as they learn it will be able to maximize the impact of data analysis tools. The power of today's best tools lies in their ability to make new correlations and find unexpected answers buried deep within data — even when people don't know the right questions to ask. Around the globe, analytics tools are finding impactful correlations and producing unexpected results. For example:

More powerful processing capabilities in today's computers combined with inventive software are empowering data scientists with cutting-edge tools to make sense of vast amounts of data, and unlock the valuable insights contained within it.

+ By tracking and correlating more than 1,000 data points a second, Canadian researchers shocked doctors by showing that prematurely born infants with unusually stable vital signs correlated with serious fevers the next day — enabling doctors to take preventive actions.[26]
+ Two decades of past newspaper stories are being used to predict when and where cholera outbreaks will occur in places like Angola.[27]
+ Police departments modified an algorithm originally designed to predict earthquakes, and now are using it to predict within 500 feet where crimes are likely to occur. Burglaries have been reduced by 33 percent and violent crimes by 21 percent in areas where the software is being used.[28]
+ Using data analytics and marine sensors that monitor waves, currents, and other data, researchers are using data analytics to predict tsunamis and other natural disasters as well as their impact.[29]
+ Data from doctors' visits and prescription information revealed how patients with autoimmune diseases are at greater risk of epilepsy.[30]
+ Credit score data is being used to predict which patients will need "friendly" reminders to take their prescription medicines.[31]
+ Using a decade of flight history data correlated with weather patterns, air travelers can now figure out which flights are likeliest to be on time.[32]

Because data surrounds us, so do opportunities. When innovators act responsibly and creatively, data innovation can deliver answers to both everyday problems and some of the world's biggest challenges.

4

TRANSFORMING AND TRANSLATING DATA

Powerful new software tools are equipping us with the ability to use data sets to make better decisions, based on facts and not "gut" or intuition.

In particular, a new set of tools is helping give data purpose, by transforming it in ways that can help us extrapolate, focus, visualize, reflect, refine, model, and predict. These tools include machine-learning technologies that understand data to help us better respond to it; modeling and simulations technologies that can test scenarios and transform data into real-world solutions; and tools that recognize and translate sound, images, or video into new more meaningful forms.

Transforming data in these ways leads to better plans, superior designs, and smarter decisions. For example, doctors practicing medicine today are bombarded with new research that makes it almost impossible to keep abreast of the latest developments, let alone interpret real-time patient data.[33] As a result, hospitals are turning to clinical decision support systems. Essentially, these are software systems that analyze data from disparate sources to help make faster and more reliable diagnoses in a complex data environment — proving to be beneficial in more than 70 percent of cases.[34]

Other software tools are helping translate data into more meaningful forms. Real-time processing of audio, image, and video data is leading to life-changing breakthroughs. To illustrate: as more data is gathered about how people speak, speech recognition technology has continuously improved. This has enabled breakthroughs like real-time, two-way language translation of voice conversations across continents — potentially opening new opportunities for global commerce and trade. Similarly, with an estimated 360 million people suffering from hearing loss, researchers in China have turned to pattern recognition and real-time processing of data from a 3D Kinect sensor to develop a system that understands the gestures of sign language and converts them in real time to spoken and written language — and vice versa.[35]

Faster image processing also is having a profound impact in areas such as cancer detection, cognitive computing, neurobiology, and robotics. For example, due to their unpredictable appearance and shape, brain tumors can be especially hard to identify within medical images. With the help of cloud computing and advanced image analysis algorithms, teams of scientists are now competing to find the best software algorithms for more accurately and quickly identifying brain tumors.[36]

Moving from 2D to 3D imagery for mammography is improving breast cancer detection rates. Three-dimensional mammography uses software to combine many x-rays at different angles to create a three-dimensional image that can increase the detection rate for breast cancer and decrease nerve-racking false alarms.[37]

Indeed, the ability to use data to create both visualizations and simulations is making it easier to comprehend and use it. We now model and simulate complex systems and test designs with greater accuracy and more speed, without actually building them. For example in the 1980s, Boeing tested 77 prototypes of its 767 using physical wind tunnels. By 2005, Boeing ran

only 11 physical tests for its 787, testing prototypes using virtual wind tunnels and supercomputing to save time, save energy, save money, and save lives.[38]

Virtual wind tunnels are one example of tools that crunch massive amounts of data to make 3D computational fluid dynamics easier to use and faster to implement. These tools enable us to better model heat flow, fluid flow, air flow, and process flow for better performance. They are being used to model where pollutants may travel in groundwater, how to boost the performance of wind turbines, and how to design better buildings that can withstand the worst that Mother Nature can throw at them.

In short, these tools transform data into solutions.

Taken together, the powers to gather, store, analyze, and transform data are converging to unlock new opportunities for better solutions.

In practice, though, each of these four drivers is often performed by disparate persons, acting on diverse data sets, stored in distributed locations. However, that is part of the power of this data revolution. Different, previously unrelated sets of data can be combined and analyzed even when stored in different locations, augmented even when the data is unstructured, and acted upon even when parties are uncovering fundamental answers to questions that the producers of the data didn't even know to ask. One may never know in advance exactly the power of the information that may later be extracted from a particular data set — in part because it may later become valuable to a combination of seemingly unrelated data sets and because the algorithm to exploit the insight has yet to be invented.

Because data surrounds us, so do opportunities. When innovators act responsibly and creatively, data innovation can deliver answers to both everyday problems and some of the world's biggest challenges. In order to maximize the data-enabled opportunity in ways that can accelerate new waves of productivity, economic growth and individual benefits, we must crank up the innovation engine once again and pave the way for the leading-edge solutions and empowering impacts it is poised to deliver.

Manufacturing companies that take full advantage of their data could

save $371 billion

over four years.

A Data-Driven Economy

Data is now emerging as one of the most dynamic new forces of economic gains. Data's economic impacts already are rippling through many sectors of the economy, in high-tech and low-tech industries alike. In the next four years alone, making better use of data could lead to a $1.6 trillion "data dividend" worldwide.[39]

Data innovation has proven its ability to boost productivity. Companies that already use data-directed decision-making report a 5 percent to 6 percent boost in productivity.[40] If by harnessing data more effectively we can achieve even small gains across a broad range of industries to make them just 1 percent more efficient, economists estimate it can add about $15 trillion to global GDP by 2030. That's the equivalent of adding another U.S. economy. A 1 percent productivity increase may seem small, but as General Electric's CEO Jeff Immelt puts it, "tell an oil guy you can use software to save him one percent on something, and that guy will be your friend for life."[41]

Every **DATA-RELATED** job in the U.S. creates **3 more jobs** indirectly.

61% of US execs say **DATA ANALYTICS** is important to their companies' **hiring plans.**

Data Innovation Is a Job Creator.

Data innovation isn't just about boosting economic growth. It's also about fueling a powerful new job creation engine and potent job force multiplier. Data innovation is already creating thousands of new, well-paying jobs — from data analysts, to software developers, to the people who run the massive data warehouses that make data innovation possible. Further, 61 percent of senior executives in the US and 58 percent in Europe say data analytics is important to their companies' plans to hire more employees.[42] For every data-related IT job created, another three jobs are estimated to be created for people outside of IT — creating millions more jobs throughout the economy.[43] The end result will be broad, new economic benefits from another new technology-driven jobs boom.

Smart buildings alone could save businesses

$25 billion

a year in energy costs.

A 1% data-driven productivity improvement in aviation could save

$30 billion

in fuel savings worldwide over 15 years.

Data Is Making Businesses More Agile, Responsive, and Competitive.

It used to be that the massive, IT-driven productivity increases that grew our economy and lifted our standards of living were limited to only a few sectors of the economy — like the technology sector itself. However, as more sectors generate more forms of data, data now is poised to deliver gains in sectors that have traditionally lagged — from agriculture to health, from transportation to education, and from energy to finance. In fact, 79 percent of senior executives recently surveyed in the United States and 80 percent in Europe say data analytics is important to their companies' plans to better serve customer needs.[44] In addition, 70 percent in the United States and 72 percent in Europe say data analytics is important to their companies' plans for creating new products or services.[45]

HEALTH CARE

Data Innovation Is Helping People Live Longer, Healthier Lives Through Better Insights.

Health care is one of the most data-rich environments today — generating hundreds of terabytes of data per hospital every day.[46] When use of this data is maximized, it can be the prescription for better care and faster cures. Data today is giving doctors unprecedented insight into their patients' health and enabling better decision making. With the help of software analytics, doctors can leverage the outcomes of every previous patient to inform treatment options for every future patient. Yet too often health data is underused. If the health care sector were to use data more effectively to drive efficiency and quality, it is estimated the sector could save more than $300 billion every year — reducing expenditures by a whopping 8 percent.[47] The biggest impacts aren't just measured in dollars saved, but lives saved. As just one example, researchers developed a machine learning algorithm that can predict cardiac arrest four hours in advance and is accurate 66 percent of the time by combing real-time data with a patient's medical history.[48]

TRANSPORTATION

Data Innovation Is Saving Time, Saving Money, Saving Fuel, and Saving Lives.

Data has placed us on course for a smarter transportation future. Throughout our airways, railways, and roadways, real-time data analytics is leading to smarter transportation systems that are improving our ability to safely and efficiently move goods and people. New cars today can generate up to 25 gigabytes of data per hour, and contain over 10 million lines of software code just to process the data.[49] This data is being used to power new safety and crash avoidance systems that could have as big an effect on safety as seatbelts have had — reducing injuries and fatalities by as much as 50 percent.[50] At a time of an ever-increasing number of vehicles on the road, data can also be put to use in innovative ways to reduce congestion and route traffic more efficiently — saving millions of lost hours, thousands of gallons of gas, and tons of greenhouse emissions.

Data in aviation is taking off, too. Today a modern airline can generate up to half a terabyte of data per flight from sensors throughout the plane; this data is used to improve flight performance, cut turbulence, improve safety, and identify engine defects 2,000 times faster than before.[51] Aviation data is also helping improve flight path planning, and letting crews know that a part needs replacing before it fails. These gains add up. A 1 percent data-driven productivity improvement in aviation could save $30 billion in fuel savings worldwide.[52]

ENERGY/ENVIRONMENT

Data Innovation Is Reducing Energy Consumption and Improving our Environment.

Data is fueling huge energy savings. Better data use can drive billions of dollars in energy savings in a range of industries, while also improving the environment.

If the health care sector were to use data more effectively to drive efficiency and quality, it is estimated the sector could save more than **$300 billion** every year — reducing expenditures by a whopping **8 percent.**

Data-driven insights enable smarter energy use in the electric grid, in buildings, in our homes, in our factories, on the farm, and throughout our environment. By using data design tools, for example, buildings can now be designed in ways that use less energy, and can be infused with intelligent systems that combine reams of sensor data with analytics and actuators to make operations more efficient. Taken together, it is estimated that smart buildings alone could save businesses $25 billion a year in energy costs.[53]

MANUFACTURING

Data Innovation is Improving the Way Products are Designed, Built, and Distributed.

The manufacturing sector stores more data than any other sector. As a result, manufacturers have a lot to gain from better use of data to boost efficiency, drive quality, and improve the way products are designed, built, and distributed. By one estimate, better use of data in manufacturing can yield up to a 50 percent decrease in product development time and assembly costs.[54] In fact, IDC (International Data Corporation) estimates that manufacturing companies that take full advantage of their data are poised to achieve a $371 billion data dividend over four years.[55] Using real-time data, companies can also better track and manage global supply chains, and reduce product defects.

Data is also helping improve designs. Data-enabled digital designs are enhancing the iterative design process — enabling designers to test countless ideas and tweak them in the virtual environment until finally reaching that optimal product design. Toyota, Fiat, and Nissan have all cut new-model development time by 30 percent to 50 percent through the collaborative use of data and modeling techniques.[56]

FINANCIAL

Data Innovation is Boosting Efficiency, Improving Compliance, and Cutting Fraud.

In the financial services sector, tapping into growing amounts of data can improve operational efficiency, boost compliance, and identify fraud. In one survey, 71 percent of banking and financial markets firms reported that the use of information and analytics is creating a competitive advantage for their organizations.[57] Data investments can pay huge dividends. Data analytics helped one credit card company identify $2 billion in avoidable annual credit card fraud.[58]

AGRICULTURE

Data Innovation is Producing More and Better Food Using Fewer Resources.

Data is cropping up everywhere on farms — helping produce more delicious and nutritious foods for a growing number of people. Thanks to a concept known as precision agriculture, farmers around the globe are able to use data from seeds, satellites, sensors, and tractors to make better decisions that increase yields, decrease costs, and feed more people. With a quarter of the world's population involved in agriculture and food production, there are big gains to be achieved from greater use of data. For example, by harnessing data analytics tools, farmers are able to reduce input costs, pesticide use, and chemical use, while improving yields by five or 10 bushels an acre.[59] For dairy farmers, a Croatian startup has pioneered a cloud-based software analytics platform that provides real-time data on things like the impact of feed quality on production and the conception rates for each animal, which in turn has increased farmer productivity and efficiency by as much as 50 percent.[60]

DATA IMPROVES LIVES WORLDWIDE

By using more than

1,000 data points

per second, Canadian scientists found that prematurely born infants with unusually stable vital signs had a higher risk of developing fevers — allowing doctors to take action early — and save lives.

CANADA

Separating Data Myths from Facts

There are a number of myths surrounding recent data innovations and the data economy. These include myths about:

+ Personal information and data protection,
+ The economic impact of the data economy,
+ Data reliability,
+ 21st-century data innovations,
+ Global benefits of data innovation, and
+ Governments' role in data regulation.

Personal Information and Data Protection

MYTH

Data innovation only benefits IT companies — not individuals.

REALITY

Data innovation empowers consumers to make better decisions and enables merchants to customize goods and services to serve us better. It can lead to major advancements through personalized learning and personalized medicine. It can lead to improved consumer experiences through personalized entertainment. It can help us transition from an economy driven by mass-production to one enabled by mass-customization. There is indeed a set of companies at the heart of developing and delivering the innovative new technologies upon which this data revolution relies. However, if we make smart choices today, this emerging data-centric economy can create new jobs and industries, and put the world in consumers' hands.

MYTH

All data is personal data.

REALITY

Some data may be personal information (e.g., data we generate on our mobile devices or that we create by using social networks). Most data, however, is not personal.

IN INDIA, INTERNET KIOSKS ARE GIVING MORE THAN

4 million farmers

access to crop prices, weather, and other information in

local languages.

INDIA

The vast amount of data being created every day includes information like satellite weather monitoring, jet engine performance, computer-generated stock market trades, and sensors unrelated to individuals. Even when data does pertain to an individual, it is often not accessed by another human and likely is de-identified — essentially stored and used without information that reveals the identity of the individual involved.

MYTH
Companies are not concerned about protecting personal data.

REALITY
When personal data is generated, they need to be protected appropriately. In order to expand data opportunities, public trust and confidence in data should be high. Companies and organizations that use data should practice good data-stewardship. These practices might be standardized through an industry-led effort to create voluntary guidelines for responsible data use. Many leaders in the field already are stepping forward to make it clear to consumers how their data is being collected and if it is shared. Many companies follow best practices that require them to anonymize personal information whenever practical.

MYTH
Data innovation will cause me to lose all privacy.

REALITY
The success of the data economy depends upon consumer trust. Individuals must feel that their personal information is secure. Leading software developers already build in privacy protections to their systems from the beginning, called "privacy by design." In addition, developers often use anonymization, de-identification, and encryption tools so that they can further minimize the impact of any data breach. When data is aggregated so individual users aren't specifically identified, data can generally still be analyzed for patterns of behavior without violating a user's trust or privacy. Furthermore, enforceable privacy policies can take into account the context and relative risks involved in any exposure or misuse of data, with the most sensitive data (like financial or health care data) getting the highest level of privacy protection. This means that data like weather data or business analytics that does not involve personal information doesn't require the same level of protection as patient-specific health care data.

MYTH
You can never fully de-identify data. De-identification of data is ineffective.

REALITY
De-identification of data is a process used to prevent a person's identity from being connected with information. Once data is de-identified, it can be analyzed without connection to an individual. Experts have developed techniques that allow data to be de-identified in ways that can maximize both privacy and data quality.[61] According to experts, if de-identification is done properly, the risk of re-identifying individuals from anonymized data is less than 1 percent in most cases.[62]

MYTH
Companies that use data can't be trusted.

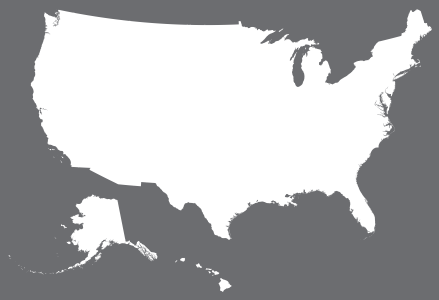
REALITY
Industry is listening to and heeding privacy concerns. Today, there are signs of vibrant competition among leading companies in a race toward better privacy protections. For example, the companies responsible for the operating systems that run a combined 96.4 percent of smartphones worldwide have both recently announced enhancements to their privacy settings. They are giving users additional controls and moving

IN THE U.S., MAJOR AUTO COMPANIES
are designing new vehicles packed with

hundreds of sensors

and analytics to enable such advances as autonomous parking and crash avoidance.

UNITED STATES



to encrypt data to protect personal privacy.[63] Leading software companies now appeal directly to consumers by touting their commitment to protecting privacy.[64] Companies even are changing their services and policies to make them more privacy-friendly by, for example, by not scanning customer communications in order to target advertising.[65]

MYTH
Individuals have no control over their data.

REALITY
While it may sometimes seem like we have no control over our data, there are a variety of tools of which consumers may not yet be aware to help them better control their data. For example, some leading software companies have chosen to enable "Do Not Track" features by default in their web browsers, so the websites you visit and the third-party advertising firm they use automatically receive a "Do Not Track" request. It informs websites that you do not want to be tracked, which may help protect you against forms of tracking on the web. In addition, some data-brokers have created web pages where consumers can see the kinds of information that have been collected about them, opt-out of further data-gathering, and correct information if it's wrong.[66] Collectively, these tools enable consumers to better control how their information is collected and used, or to opt out of certain marketing uses.

The Economic Impact of the Data Economy

MYTH
Data innovation won't create new jobs and may even take jobs away.

REALITY
Data innovation can be a powerful driver of economic growth. In fact, 61 percent of recently surveyed senior executives in the United States and 58 percent in Europe say data analytics are important to their companies' plans to hire more employees.[67] Yet data innovation is not just about creating jobs within the IT sector. Every data-related role will create employment for an estimated three people outside of IT, creating even more jobs throughout the economy.

While there will be some job changes as data innovation helps find new ways to do old things, this shift is likely to resemble the proliferation of the Internet where an estimated 2.6 jobs have been created for every job disrupted and lost.[68] One of the biggest areas of potential job growth is data analytics. Making sense of the insights contained within data can often only be achieved through human ingenuity. Today, there is a global shortage of the skilled analysts and data managers who can help make sense of data. According to McKinsey, the United States alone faces a shortage of 140,000 to 190,000 people with the deep analytical talent needed to unlock the hidden potential in data, and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of the data economy.[69]

The 2014 mean base salary for a staff data scientist is $120,000 and $160,000 for a manager.[70] Many have suggested that to capture the full potential from data-driven innovation, companies and policymakers around the world must take steps to overcome a looming talent shortage.

IN BRAZIL, one of the largest

soy producers

is investing in software and the power of data analytics to increase the efficiency of damage control techniques, reduce costs, and boost crop techniques.

BRAZIL

MYTH

Data analytics is about getting human judgment out of the process.

REALITY

While some questions can be answered with data (for example, is the population of my town increasing or decreasing?), many of the most insightful answers are not as clean cut. You may not always know how the various data elements relate to one another. And because you may not know in advance the right question to ask, data analysis is often an iterative process of asking successive questions to ultimately find the answer. For these reasons, we can never do away with human judgment and input to reconcile differences and to sort through potential inconsistencies.

Data alone is not a panacea, and cannot work miracles. In fact on its own, data often has little value. It's often messy, not inherently organized, or neatly structured. The hard work comes from making sense of it and finding the relevance within it. Whether or not data can solve problems depends upon the effective execution of a smart data strategy that can lead to faster and better solutions. It also depends on asking precisely the right questions of it. But if we harness data in the right ways, we can help unlock answers to some of society's most pressing challenges, help stoke the innovation bonfire, and fuel a powerful new round of IT driven jobs and economic growth.

MYTH

Data innovation is only for big companies — not small businesses.

REALITY

With data becoming more ubiquitous, storage costs falling, and analytics tools becoming more powerful and more affordable, now even the smallest companies can take advantage of advanced data analytics — tools that were once only available to the biggest of businesses. For example, the Trends feature in Intuit's QuickBooks Online allows small businesses to benefit from the collective wisdom of fellow Intuit users — allowing small businesses to see how their income and expenses compare in order to highlight opportunities. It enables them to make smarter decisions about how they operate. While the use of business intelligence and analytics solutions is not widespread among small and medium-size enterprises, adoption is expected to grow quickly.[71] A recent study found that data analytics are important to 60 percent of small companies.[72] That includes 57 percent of US companies with 50 or fewer employees and 62 percent of same-sized European companies, according to their senior decision-makers. In medium-sized companies (those with 51 to 500 employees), 87 percent of US executives and 79 percent of European executives say data analytics are important.

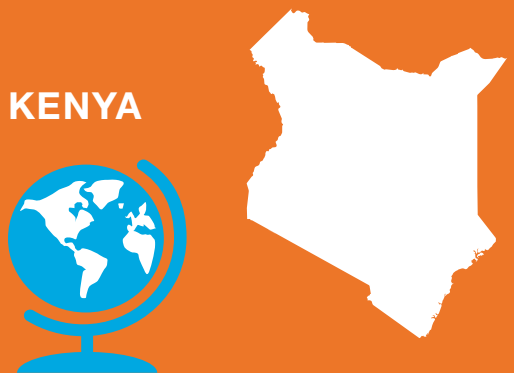
SCIENTISTS STUDYING

patterns of malaria

recently used mobile phone data in Kenya to pinpoint hotspots where disease transmission was taking place, thus guiding government

eradication efforts.

KENYA



MYTH
Data only benefits the IT sector — not other sectors of the economy.

REALITY
Data is now at the heart of a major technological transition that promises to transform and improve almost every sector of the economy. In fact, many believe that data innovation has the ability to improve productivity throughout economic sectors. Although companies that use data-directed decision-making report a 5 percent to 6 percent boost in productivity, if data innovation were able to achieve just a 1 percent efficiency improvement, the impacts throughout all sectors of the economy would be substantial — creating savings in energy and fuel, producing better health outcomes at lower costs, and increasing the performance and life of physical assets.[73] In the commercial aviation industry, for example, GE predicts that as we are able to capture real-time data to improve engine efficiency and route travelers more effectively, just a 1 percent boost in fuel savings would yield savings of $30 billion over 15 years.[74]

MYTH
Correlation always implies causality.

REALITY
This search for the "unknown unknowns" is one of the major insights that can come from data. However, sometimes correlations just don't make sense, and correlations don't always mean causation. For example, while the rate of violent crimes and murder has been found to jump when ice cream sales do, it is highly unlikely that buying ice cream turns people into killers.[75] Still, finding causal relationships in data is among the most valuable insights that we can discover within data. And many interesting correlations and causalities are being discovered every day at increasing rates, and with greater importance. The ability to sort through the difference between mere correlation and causality is one of the reasons why taking advantage of data requires qualified data scientists that understand how to separate mere correlation from causality and eliminate results that fail the implausibility test.

MYTH
Data innovation requires a big budget.

REALITY
Taking advantage of data doesn't necessarily require big budgets. It often only takes minor investments, and those investments can lead to impressive returns. For example, organizations that take a more holistic approach to their data are projected to realize some of the biggest returns — achieving a massive "data dividend" of roughly $1.6 trillion in additional revenue, lower costs and improved productivity over just the next four years.[76] In fact, using small data sets, and data that may already be in hand, can often be a good starting point. Examining a week's worth of financial transactions data for fraud may illuminate trends more easily than examining five years' worth of historical data. In one analysis, companies that effectively employed data economy analytics were 26 percent more profitable than their industry competitors, generated 9 percent more revenue through their employees and physical assets, and enjoyed 12 percent higher market valuation ratios.[77]

In the United Arab Emirates, new data tools are being used to design the world's first

positive-energy building

that produces more energy than it consumes.

U.A.E.

Data Reliability

MYTH
Insights from data are always accurate on their own.

REALITY
Insights are only accurate if the underlying data is accurate, the tools are smart, and a smart data scientist is engaged in developing the model to avoid skewed results. For example, Boston's innovative Street Bump app used smartphones that incorporate accelerometers and GPS location to locate potholes from a large number of users each time someone hit a bump in the road. However, if smartphone users are more likely to be young and affluent, then the data from the app may not be accurately capturing pothole data from all demographic areas of the city — and city response to potholes could risk becoming skewed. That is why data scientists often need to be involved to ensure statistical relevance, and avoid bias in data outcomes.

MYTH
Gut decisions are usually the right decisions.

REALITY
From the time we are born, we learn to trust our intuition in decision-making. In fact, an estimated 19 percent of global managers describe themselves as "visceral decision makers" that rely almost exclusively on gut instinct and intuition.[78] Yet today, data can help us make better informed decisions, with greater precision, faster speed, and bigger impact. According to one survey of IT decision-makers across a range of industries, 59 percent report that improving the quality of decision-making is the primary goal driving investments in data technologies.[79] It doesn't just benefit the workplace; we also make instinctive decisions throughout our daily lives. In some cases, we can benefit greatly from tools that can help us make faster and better decisions. For example, human error is the culprit in an estimated 93 percent of automobile crashes, but new vehicle sensors that collect huge amounts of data about their surroundings combined with real-time analysis and automated decision-making (even without becoming fully autonomous) could reduce car related injuries and fatalities by as much as 50 percent.[80]

MYTH
Bigger data is always better data.

REALITY
Data sets don't always have to be bigger to be better. In fact, size may be one of the least important factors. For many problems and questions, even small amounts of data, when analyzed with the right tools can lead to keen insights. Developing these insights often depends more upon the quality of the underlying data, and the quality of the tools used to act upon it. Yet there is sometimes a belief that with more data come more truth, and the larger the data set the more objective it is. Sometimes it is having access to simple data that can have the most immediate impact; for example, knowing when a local store closes, how many miles a car has gone since its last oil change, or how much a phone bill costs this month. Depending on the question being asked, the data contained on web pages, in Excel spreadsheets, or CRM databases can be small, and just as powerful at providing answers as big data. What matters most is creating robust data, securely storing the data, having access to the data, and being able to process the data — whatever its size — so that it can be utilized when and where it is needed to solve problems.

Weather data

isn't just used to predict if we need to carry an umbrella; it can also be used to help predict crop yields, the likelihood of a flu outbreak, and help estimate how much desalination capacity needs to be built in Morocco.

MOROCCO

MYTH
Data that is unstructured is unusable.

REALITY
Unstructured data can be useful. In fact, some of the most powerful data doesn't always fit into nicely structured tables of columns and rows. It can be unstructured like text documents or x-ray images. The key is transforming data into its most useful forms. If the average Fortune 1000 company could increase the usability of its data by just 10 percent, the company could expect an increase of over $2 billion in revenue.[81]

MYTH
Data should only be used for the original purpose for which it was collected.

REALITY
Cloud technologies have dramatically driven down the price of data storage so that data no longer has to be deleted after its initial use to make room for new data. As a result, unlike other resources, data is not used up once it's used once. Data is a renewable resource that can often be reused, combined with other data sets, and used multiple times to produce answers to questions often unforeseen when the data was originally created. For example, weather data isn't just used to predict if we need to carry an umbrella; it can also be used to help predict crop yields, the likelihood of a flu outbreak, and help estimate how much desalination capacity needs to be built in Morocco.

21st Century Data Innovations

MYTH
The most profound benefits from data are far off in the future.

REALITY
While data innovation will continue to generate benefits decades into the future, its powerful results can already be seen almost everywhere around us today. Already, terabytes of daily data are being used to provide you with more accurate weather predictions; data from millions of GPS signals are used to predict commute times and speed you to your destination; millions of health data points are being used to identify potential causes of disease; and terabytes of financial data are being used to help prevent credit card fraud.

In fact, when the Economist Intelligence Unit asked survey respondents to describe the impact data has had on their organizations over the past five years, nearly 10 percent said it had completely changed the way they do business.[82] 46 percent of respondents said it had become an important factor that drives business decisions. New research suggests that in just the next four years alone, making better use of data could lead to a $1.6 trillion "data dividend" worldwide as data helps unlock new opportunities and solutions.[83] In a separate survey looking at the current year, 33 percent of senior executives in the United States and 24 percent in Europe said they expected 10 percent or more of their companies' growth to be related to data analytics.[84] Looking ahead five years, 58 percent of those US senior executives and 43 percent in Europe made the same prediction.

In the aftermath of the 2004 South Asian tsunami, Indonesian fisherman were given cell phones. Their incomes went up by

30 percent

because for the first time in their lives they had data on the actual market price of fish.

INDONESIA

MYTH
Data is overhyped.

REALITY
Using data to solve human problems is hardly new. We've been using data for better decision-making since the beginning of modern civilization — leading to fallow farming techniques that feed more people, navigation techniques that have fueled global trade, and health insights that have avoided millions of cholera deaths. However, in the past, data was a scarce resource that was costly to store and difficult to manipulate. What is different today is that data has become more abundant, storage costs have plummeted, and the tools to manipulate it have become more powerful. As a result, as we face a new set of emergent challenges, powerful new data analytics technologies can help sort through growing volumes of data to help us discover powerful insights and unexpected solutions to some of our most pressing challenges.

MYTH
The era of IT-driven economic growth is over — data innovation can't boost productivity.

REALITY
IT innovation and its ability to grow economies, create jobs, and raise standards of living around the globe has been rooted in its proven ability to boost productivity — for example, increasing productivity by as much as 1 percent to 2 percent in the 1990s.[85] A prominent economist at Northwestern University, argues that the greatest gains from IT innovation are behind us.[86] However, the productivity era is alive and well. In fact, companies that use data-directed decision-making report a 5 percent to 6 percent boost in productivity.[87] Even if this unfolding data opportunity only boosts productivity in the U.S., for example, by 1.5 percent, over a 20-year period it would save enough money to raise average national incomes by as much as 30 percent.[88] If by harnessing data more effectively we can achieve small gains across a broad range of industries to make them just 1 percent more efficient, economists estimate it can add about $15 trillion to global GDP by 2030, that's the equivalent of adding another U.S. economy to the global economy.[89]

Barcelona is harnessing data to build a

smarter city,

improve government services,
and provide more

sustainable transportation

solutions.

SPAIN

The Global Benefits of Data Innovation

MYTH

Only the United States is benefitting from data today.

REALITY

Around the world, data is already being put to use to solve important challenges. For example:

+ By tracking more than 1,000 data points a second, Canadian scientists found that prematurely born infants with unusually stable vital signs had a higher risk of developing fevers — allowing doctors to take action early — and save lives.
+ In Brazil, one of the largest soy producers is investing in software and the power of data analytics to increase the efficiency of damage control techniques, reduce costs, and boost crop productivity.
+ Barcelona is harnessing data to build a smarter city, improve government services and provide more sustainable transportation solutions.[90]
+ In the United Arab Emirates, new data tools are being used to design the world's first positive-energy building that produces more energy than it consumes.[91]
+ Farmers in India are using data from seeds, satellites, sensors, and tractors to make better decisions about what to grow, how they grow it, how to track food freshness from farm to fork, and how to adapt to changing climates.[92]

MYTH

Developing countries aren't yet ready to take advantage of data analysis.

REALITY

The data revolution and the benefits it creates are a global phenomenon. Some of data's most important benefits and biggest opportunities lie in the developing world where technology has often lagged. According to IDC, emerging markets in the digital universe will surpass mature markets by 2017 — growing from 36 percent to 62 percent of the expanding digital universe between 2012 and 2020.[93] Likewise, a survey of NGOs in the developing world found that over 90 percent believed that data analytics would be the most important tool to deliver better insight for helping their end beneficiaries.[94]

Examples of data innovation in the developing world abound:

+ Scientists studying patterns of malaria infection recently used mobile phone data in Kenya to pinpoint hotspots where disease transmission was taking place, thus guiding government eradication efforts.[95]
+ In the aftermath of the 2004 South Asian tsunami, Indonesian fisherman were given cell phones. Their incomes went up by 30 percent because for the first time in their lives they had data on the market price of fish.[96]
+ In Peru, historical sites are under threat from development. Using aerial technologies and powerful software that stitches together imagery, Peru created detailed three-dimensional data point clouds to map, monitor, and safeguard its endangered treasures.[97]
+ To preserve its cultural heritage, Vietnam is using 3D scanners and has set a goal of digitizing 40,000 historical artifacts over the next five years.[98]
+ In India, Internet kiosks are giving more than 4 million farmers access to crop prices, weather, and other information in local languages. They use advanced analytics and mobile technologies to track data from individual farms and, after analysis, can offer the farmers supplies based on their needs, such as fertilizer and seeds.[99]

TO PRESERVE ITS CULTURAL HERITAGE,
Vietnam is using 3D scanners to digitize

40,000
historical artifacts
over the next five years.

VIETNAM

Governments' Role in Data Regulation

MYTH
Data will be used as a tool for exclusion — with the ability to reinforce disadvantages faced by low-income and disadvantaged communities.

REALITY
Resolving any worries that data is being used in ways that are unfair to some people or classes of people is a challenge that should be a top priority for industry and government alike. At the same time, there are opportunities that shouldn't be overlooked to use data to fight unfair discrimination and empower groups. In fact, when used responsibly, data can be a powerful new tool used to help uncover existing discrimination that impairs access to jobs, finance, education, and opportunity.[100]

MYTH
Governments have no role to play.

REALITY
With the massive growth of remote computing power, storage, analytics, and software services, a range of new policy issues has emerged. Most treaties, laws and regulations did not foresee these possibilities when they were written. This ambiguity can be problematic for governments and the growth of the data economy. Longstanding policy questions like the balance of privacy and security, the free flow of trade, basic research, and workforce development must be considered anew as the growth of these technologies has exploded.

There are several concrete steps that policymakers can take to help capture the data revolution's full potential and accelerate the new waves of productivity, economic growth, and consumer benefits that lie just over the horizon. They can invest in advancing the data solutions that can address some of society's greatest challenges — improving how we learn, produce food, live our lives, save energy, travel from place to place, and grow economies. They can establish clear rules for international law enforcement cooperation. They can foster the marketplace freedoms that affect the technology industries own ability to invest, innovate, trade, and grow. They can help foster the free flow of data unencumbered by physical boundaries. They can help ensure a talented workforce and overcome the looming data scientist's talent gap. And they can help stoke the innovation bonfire to speed the benefits from data to both businesses and consumers. With pragmatic policies that tap human talent and tenacity; harness innovation and investment; and expand capacity and capabilities; policymakers can help maximize the benefits that this data centered economy can deliver.

IN PERU, HISTORICAL SITES ARE UNDER THREAT FROM DEVELOPMENT.
Using aerial technologies and powerful software that stitches together imagery, Peru created

detailed three-dimensional data point clouds

to map, monitor, and safeguard its endangered treasures.

PERU

MYTH
Data localization helps protect privacy and improve security.

REALITY
Some may believe that data localization — requiring data to be stored within the confines of a certain country's border — can improve privacy and security. However, today's technology benefits are being enabled by the global force that is the Internet, and fueled by data that crisscrosses the globe between disparate data centers. Cross-border Internet traffic has increased by over 50 percent since 2005.[101] Enabling data to flow freely across borders is allowing even the smallest companies and entrepreneurs to become corner stores for the entire planet as they begin selling and sourcing products, services, and ideas across borders. Yet governments around the globe are often considering policies that would restrict the free flow of data, or require that data servers be located within their jurisdictions as a condition of serving the market.[102] These restrictions undermine the enormous efficiencies of scale and economic benefits that can come from data innovation, and the ability to combine different data sets in different locations to discover beneficial insights from the growing abundance of data. It can also undermine security by preventing valuable data from being backed up in multiple locations to protect it in the event of a natural disaster or technical failure. To achieve the benefits that data can deliver, every country's laws don't need to be identical, but they do need to be compatible. Enabling the free flow of data across borders is one of the fundamental tenets for enabling data-driven benefits.

MYTH
The only way data can be protected is if governments step in to require that it be protected.

REALITY
Existing government privacy rules can be combined with rigorous, innovative privacy advances, and voluntary industry best practices to ensure data is secure and personal information is protected. By contrast, government mandates that attempt to approach privacy and security by requiring that data be stored locally could inhibit innovation and limit the kinds of societal benefits that data innovation can deliver.

DIGITAL DISCOURSE

Understanding the Language of Data

Abundant data
Once scarce, today the abundance of data has been made possible by a growing ability to gather meaningful forms of digital data in entirely new ways, combined with the plummeting costs of storing data, and new ways to create value from it.

Adaptive intelligence
Adaptive intelligence is computer intelligence that doesn't just involve the statistical processing of data, but combines it with data containing specific domain intelligence. By combining models of intelligent behavior with expert knowledge, systems can better learn from examples and adapt to novel situations.

Algorithm
An algorithm is a step-by-step procedure or series of computer instructions that uses math to analyze data in order to solve problems. Algorithms are used in almost every software program.

Ambient intelligence
Ambient intelligence represents the vision that technology will become increasingly invisible and unseen as growing amounts of ubiquitous and low-cost computing are embedded in the world around us. Ambient intelligence would be available whenever we need it as literally everything becomes connected, intelligent, and responsive.

Analytics
Analytics is the simultaneous use of statistics and software-based algorithms to discover meaningful insights, patterns, and connections from within data.

Anomaly detection
Anomaly detection is the identification of data items in a data set that do not match an expected pattern. Anomalies are also called outliers, exceptions, or contaminants in data and can often provide critical and useful information.

Anonymization
Anonymizing data involves removing all personally identifiable information that could lead to the identity of a person.

Bad data
Bad data is data that is missing or incorrect. It can be as simple as an incorrect street address, but bad data costs Fortune 1000 companies billions of dollars every year.

Big data
Big data is an umbrella term that often refers to the process of applying computer analytics to massive quantities of data in order to discover new insights and improve decision-making. It often describes data sets that are so large in volume, so diverse in variety, and moving with such velocity that it is difficult to process using traditional data processing tools.

Brontobyte
A brontobyte is an unofficial measurement term for an extraordinarily enormous amount of data. A brontobyte is generally considered to be the equivalent of 1,000 yottabytes and is represented by a 1 followed by 27 zeros.

Business intelligence (BI)
Business intelligence refers to the set of technologies and applications that transform raw data into operational insights that can improve business performance and decision-making.

Clustering analysis
Clustering analysis is the process of identifying data objects that are similar to each other and clustering them together in order to better understand the differences as well as the similarities between data.

Cloud
The cloud is a broad term that refers to any application, service, or data that is hosted remotely. In general, it is made possible by large groups of remote servers that are networked together to enable ubiquitous, on-demand network access to computing or storage resources.

Cognitive computing
Cognitive computing is the process of combining large amounts of information with machine learning techniques, pattern recognition technologies and sometimes natural language processing to mimic the way the human brain works. These systems are often able to learn and interact with people by combining information sources with context and insight.

Computer-generated data
Computer-generated data refers to data that is automatically generated by a computer without human intervention — like a computer log file, satellite telemetry data, or sensor data from an industrial machine.

Dark data
Dark data consists of unstructured and untapped data that is being stored, has not been analyzed or processed, and is believed to be neglected or underutilized in some way.

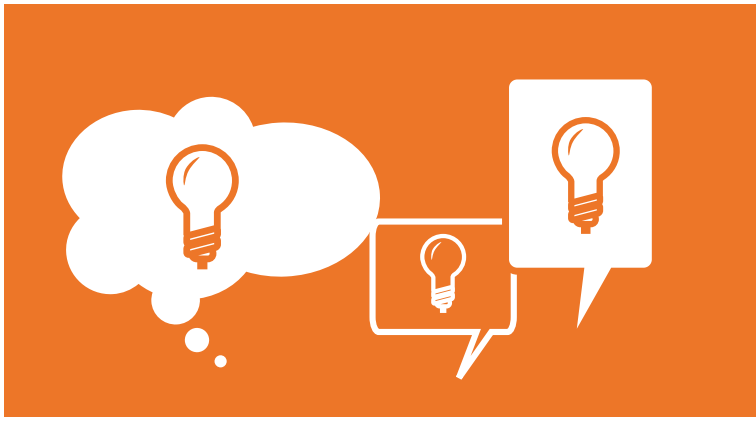
Data
Data is information in a raw and unorganized form that can be digitally manipulated to represent conditions, objects, or ideas. Common types of data include sales figures, marketing research results, readings from weather sensors, or a list of cities and their populations. We now generate an estimated 2.5 quintillion bytes of data each day.

Data aggregation
Data aggregation is the act of gathering data from multiple sources for the purpose of providing a higher order analysis.

Data aggregation tools
Data aggregation tools transform scattered data from multiple sources into a single new set of data.

Data analytics
Data analytics is the application of software as a way of transforming and modeling data in order to derive useful information, insights or meaning from data. It is often used to uncover hidden patterns or unknown correlations, and aid in decision-making.

Data analyst
A data analyst is someone responsible for preparing, cleaning, and processing data.



Data architecture and design
Data architecture is generally performed in the planning phase of a new system to design and structure how data will be processed, stored, used, and accessed. By defining at the start how specific data will be related to each other and put into motion, it is possible to design how the data will flow and control the flow of data to ensure it is protected throughout the system.

Database
A database is a large structured set of organized digital data designed so that the data within it can be rapidly searched, accessed, and updated.

Data center
A data center is a physical facility that houses a large number of networked servers and data storage repositories typically used for remote storage and processing of large amounts of remotely accessible data. There are an estimated half a million data centers worldwide, many of which make up the cloud.

Data cleansing/cleaning
Data cleansing is the process of reviewing and revising raw data to find and delete duplicates, correct errors, add missing data, remove corrupt data, and provide more consistency.

Data-directed decision making
Companies that use data-directed decision making gather, process, and analyze data to support crucial decisions. Research by Eric Brynjolfsson, an economist at the Sloan School of Management at the Massachusetts Institute of Technology, shows that companies that use data-directed decision-making enjoy a 5 percent to 6 percent boost in productivity.

Data mining
Data mining is the process of using powerful computer algorithms to find patterns or knowledge from within large data sets.

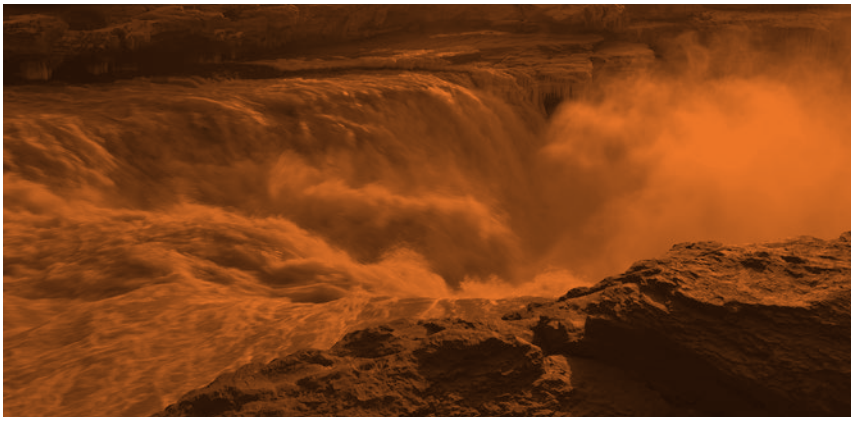
Data quality
Data quality is a metric used to define the value of data to the user. It refers to the reliability, efficiency, and worthiness of the data for decision making, planning, or operations.

Data science
Data science is a discipline that incorporates statistics, data visualization, computer programming, data mining, machine learning, and database engineering in order to extract meaningful insights that can solve complex problems.

Data scientist
A data scientist is someone who is able to combine human insights, mathematical know-how, and technological tools to make sense out of data, for example by developing and deploying computer algorithms.

Data security
Data security is the practice of protecting data from destruction, misuse or unauthorized access. Appropriate data security measures can help prevent data breaches, ensure data integrity, and protect privacy. It often involves a combined focus on people, processes, and technology.

Data set
A data set is a collection of related sets of information, typically separate elements, in a tabular form that can be manipulated as a unit.



Data source
A data source is the primary location where data comes from, for example, from a database, spreadsheet, or a data stream.

Data visualization
Data visualization involves creating visual representation of data in order to derive meaning or communicate information more effectively.

Data virtualization
Data virtualization is the process for retrieving and manipulating different data sources without having to know the technical details about where it is located or how it is formatted.

De-identification
De-identification of data is the process of stripping out information that links a person to a particular piece of information.

Disruptive shifts
Disruptive shifts are the big and fundamental changes in society and businesses, often enabled by transformative new technologies that set up a whole new context for how we work, live, play, and create value. Data innovation is often described as a technology that enables disruptive shifts.

Exabyte
An exabyte is an enormous unit of data storage — a 1 followed by 18 zeros. To put it in context, today we create one exabyte of new information on a daily basis.

Hadoop
Hadoop is an open source software framework that was built to enable the processing and storage of huge amounts of data across distributed file systems.

Internet of Things
The Internet of Things describes a world where ordinary devices are made much smarter, and connected to the Internet to extend the smart revolution from the palm of our hands to the world around us. Because everything that can be connected, will be connected, some have more aptly described it as the Internet of Everything. By one estimate, we have only connected about 1 percent of the things in the world that can be connected. By 2020, an estimated 50 billion devices will be connected to the Internet.

Legacy system
A legacy system is any computer, application, or technology that is outdated or obsolete, but continues to be used because it performs a needed function adequately.

Machine learning
Machine learning is the use of algorithms to allow a computer to analyze data for the purpose of “learning” from experience the actions to take when a specific pattern or event occurs.

Metadata
Metadata is the data about data. It can include basic summary information about the data like the author of the data, the date it was created, the file-size, and date last modified.

Outlier detection
An outlier is a piece of data that deviates significantly from the general average within a larger data set. It is numerically distant from the rest of the data and therefore, the outlier indicates that something is going on and generally therefore requires additional analysis. *(See also Anomaly detection.)*

Pattern recognition
Pattern recognition is the process of looking for and identifying patterns within data. It can be simple, like identifying a repeating set of sequences within a DNA sequence, it can be finding a pattern in the way two data sets interact to discover whether there is a pattern connecting one event to another, or with the help of machine learning it can be looking for more complex patterns like finding numerical characters in a picture.

Petabyte
A petabyte is an enormous measure of storage capacity that is represented by a 1 followed by 15 zeros, or a million gigabytes. A petabyte is roughly four times the amount of data contained in the Library of Congress.

Predictive analytics
Predictive analytics involves using software algorithms on one or more data sets to predict trends or future events. When data from the present can be compared to the past, it can often be used to help predict the future.

Predictive modeling
Predictive modeling is the process of developing a model that will most likely predict a trend, future behavior, or outcome — often by comparing events from today to events from the past.

Real-time data
Real-time data is data that is acted upon as it is created. It is often created, processed, stored, and analyzed within milliseconds. Real-time data can include everything from stock market prices to the speed of a wheel as used in a car's anti-lock brake system.

Recommendation engine
A recommendation engine is a computer algorithm that makes recommendations, suggestions, or that can personalize something for you based upon a variety of data patterns often derived through machine learning techniques.

Regression analysis
Regression analysis is a statistical process for using data to estimate the relationship between two or more variables.

Risk analysis
Risk analysis is the use of software data analytics tools to identify the likely risk of a project, action, or decision. New data tools can help identify possible risks up front, better model an array of scenarios to help reduce the risk facing organizations, and monitor systems to identify problems if things begin to head off course.

Root-cause analysis
Root-cause analysis is a method of problem solving that is focused on looking at the relationship between cause and effect to identify the root cause of a fault or problem. The cause is a root cause if once it is removed from a sequence of events, it prevents the undesirable event from repeating.

Semi-structured data
Semi-structured data is not structured by a formal data model, like those used in databases, but provides other means of describing the data and hierarchies. Semi-structured data often uses tags or other data markers in what is sometimes knows as self-describing structure.

Small data
Small data is about harnessing even small amounts of data, like that contained in a customer survey, to achieve actionable results. It generally refers to data sizes small enough that a human could comprehend and analyze it.

Structured data
Structured data is highly organized and generally organized into rows and columns making it easy to search and manipulate.

Terabyte
A terabyte is a measure of data that is represented by a 1 followed by 12 zeros. Terabyte hard drives can now be commonly found in home and work computers, or accessed via the cloud. To put it in context, a terabyte can store about 300 hours of high-definition video.

Text analytics
Text analytics is the use of statistical, linguistic, and machine learning techniques on text-based data to derive meaning, extract concepts, or unlock insights. Text analytics is generally performed on natural language text like that contained in documents, transcripts, web postings, commentary, or forms. It can be useful for the summarization, discovery, or classification of content.

Transactional data
Transactional data is data that is derived from specific events like financial purchases, invoices, payments, and shipping data. It generally includes a timestamp and supports the daily operations of an organization.

Unstructured data
Unstructured data has no pre-defined structure — for example, notes from a meeting. According to some estimates, unstructured information might account for more than 70 percent to 80 percent of all data in an organization.

Variety
Variety, one of the four Vs defining data innovation, represents the various kinds of data often from different sources that are combined and analyzed to produce insights. The variety of types of data that today are being processed in applications can include textual databases, transaction data, streaming data, images, audio, and video.

Velocity
Velocity, one of the four Vs defining data innovation, is the speed at which the data is created, stored, analyzed, and visualized. For example, large data warehouses may receive billions of rows of new information each day. Time-sensitive data must be used as it is streamed in order to maximize its value.

Veracity
Veracity, one of the four Vs defining data innovation, is used to signify the accuracy, certainty, and precision of the data.

Volume
Volume, one of the four Vs defining data innovation, refers to the amount of data processed — ranging from megabytes to brontobytes.

Yottabytes
A yottabyte is a very large measure of data storage that is represented by 1 followed by 24 zeros. To put it in context, a yottabyte represents that amount of data stored on 250 trillion DVD's.

Zettabytes
A zettabyte is a measure of storage that is represented by 1 followed by 21 zeros. As of 2013, the World Wide Web is estimated to have reached 4 zettabytes. By 2016, more than a zettabyte of data is projected to cross our networks globally on a daily basis.

Endnotes

1 Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making.* Cap Gemini, 2012. Web. Point Of View. https://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making

2 A 1 percent productivity increase may seem small, but as Jeff Immelt, CEO of GE puts it, "tell an oil guy you can use software to save him one percent on something, and that guy will be your friend for life." Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines.* GE, 2012. Web. http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf

3 An estimated 50 billion devices will be connected to the Internet by 2020 enabling the Internet of Everything. Evans, Dave. *The Internet Of Things How The Next Evolution Of The Internet Is Changing Everything.* Cisco Internet Business Solutions Group (IBSG), 2011. Web. White Paper. http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

4 "Bringing big data to the enterprise." IBM. http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

5 Golub, Ben. "Enough Data to Fill a Stack of DVDs to the Moon (and Back)." *ComputerWorld* 2011. Web. http://blogs.computerworld.com/18351/a_stack_of_dvds_to_the_moon_and_back

6 Corry, Will. "BIG Data / The Volume Of Business Data Worldwide, Across All Companies, Doubles Every 1.2 Years, According To Estimates." *The Marketing Blog* 2012. Web. http://www.themarketingblog.co.uk/2012/10/big-data-the-volume-of-business-data-worldwide-across-all-companies-doubles-every-1-2-years-according-to-estimates/

7 "Digital Imaging in the Cloud." *There Magazine* 2012: 16. Web. http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf

8 The 787 uses data sensors to reduce fuel, monitor systems, and even employs accelerometers in the nose of the plane to counteract turbulence. If the sensors register a sudden drop, they immediately tell the wing flaps to adjust (in a matter of nanoseconds) and in so doing, what used to be a 9 feet drop in an older plane can be reduced to just 3 feet in the 787, making for a much smoother flight. Humphries, Matthew. "The Boeing 787 produces over 500GB of Data during Every Flight." Geek.com 2013. Web. Gosling, Kevin. "E-Enabled Capabilities of the 787 Dreamliner." *Aero Quarterly* 2009: 22-24. http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf

9 Jet engine maker GE says the engine data allows it to figure out things like possible defects 2,000 times as fast as it could before. Hardy, Quentin. "What Cars Did for Today's World, Data May Do for Tomorrow's." *New York Times* 2014. B7. Print. http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs&region=Body&_r=1&

10 Henschen, Doug. "Bad Winter Weather Meets Big Data Prediction." *InformationWeek* 2014. Web. http://www.informationweek.com/big-data/software-platforms/bad-winter-weather-meets-big-data-prediction/d/d-id/1114192

11 Groenfeldt, Tom. "At NYSE, The Data Deluge Overwhelms Traditional Databases." *Forbes* 2013. Web. http://www.forbes.com/sites/tomgroenfeldt/2013/02/14/at-nyse-the-data-deluge-overwhelms-traditional-databases

"NYSE Euronext Adapting To Market Changes with Near-Real-Time Insight into Information." IBM Corporation, 2013. Print. NYSE Euronext. http://www.ibmbigdatahub.com/sites/default/files/document/NYSE-Euronext-IMC14787USEN.PDF

12 Mayyasi, Alex. "Why UPS Trucks Don't Turn Left." *Priceonomics.com* 2014. Web. http://priceonomics.com/why-ups-trucks-dont-turn-left/

13 *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper.* Cisco Visual Networking Index, 2015. Print. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

14 "LSST General Public FAQs." LSST.org 2015. http://www.lsst.org/lsst/faq

15 Quazi, Ed. "Next Generation Sequencing and the Big Data Bottleneck." *Arkivum* 2014. Web. http://arkivum.com/next-generation-sequencing-and-the-big-data-bottleneck/

[16] Valerio, Pablo. "Internet Of Things: 50 Billion Is Only The Beginning." *EE Times* 2014. Web. http://www.eetimes.com/document.asp?doc_id=1321229

[17] In 1980, there was a rule of thumb that one needed a data administrator for 1GB of storage. At that time a GB of disk cost about a million dollars, and so it made sense to have someone optimizing it and monitoring the use of disk space. Gray, Jim, and Prashant Shenoy. "Rules Of Thumb in Data Engineering." Redmond, WA: Microsoft Research Advanced Technology Division, 2009. Print. Technical Report. http://research.microsoft.com/pubs/68636/ms_tr_99_100_rules_of_thumb_in_data_engineering.pdf

[18] Wohlsen, Marcus. "Dropbox Slashes Its Price As The Cost Of A Gigabyte Nears Zero." *Wired* 2014. Web. http://www.wired.com/2014/08/dropboxs-plan-to-stay-relevant/

[19] From more than $200,000 a gigabyte in 1980 (even up to million dollars) to $0.02 per gigabyte in 2013 Meeker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. Presentation. http://cryptome.org/2014/05/internet-trends-2014.pdf

[20] Based on average fuel efficiency of passenger cars in 1980 (24.3 mpg), enabling a person to buy 10 million times more capacity for the same price leads to 10 million gallons of gas, which could fuel 243 million miles of travel. If the circumference of the earth is 24,901 miles, then a person would be able to circle the earth 9,758 times or nearly 10,000 times. U.S. Department of Transportation. Table 4-23: Average Fuel Efficiency Of U.S. Light Duty Vehicles. Washington, DC: Bureau of Transportation Statistics, 2013. Print. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_04_23.html

[21] Savitz, Eric. "Big Data: The Hidden Opportunity." *Forbes* 2012. Web. http://www.forbes.com/sites/ciocentral/2012/05/01/big-data-the-hidden-opportunity/

[22] The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s according to according to research by Martin Hilbert and Priscila López. Hilbert, M., and P. Lopez. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science* 332.6025 (2011): 60-65. Web. http://www.sciencemag.org/content/332/6025/60

[23] Storage costs have now fallen from $569 per gigabyte of storage in 1992 to $0.02 per gigabyte in 2013 — at a rate of about 38 percent annually. Meeker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. Presentation. http://cryptome.org/2014/05/internet-trends-2014.pdf

[24] Lohr, Steve. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle To Insights." *New York Times*. 2014: B4. Print. http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?_r=0

[25] "Data, data everywhere," *The Economist*, Feb. 25, 2010. http://www.economist.com/node/15557443

[26] Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal.* 2013. Print. http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740

[27] Reports of droughts in Angola in 2006 triggered a warning about possible cholera outbreaks in the country, because previous events had taught the system that cholera outbreaks were more likely in years following droughts. The system's warnings were correct between 70 percent and 90 percent of the time. Simonite, Tom. "Software Predicts Tomorrow's News by Analyzing Today's And Yesterday's." *MIT Technology Review* 2013. Print. http://www.technologyreview.com/news/510191/software-predicts-tomorrows-news-by-analyzing-todays-and-yesterdays/

[28] "Ten Big Data Case Studies in a Nutshell." *TechTarget*, 2013. Print. Essential Guide. http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell

[29] "Big Data to Predict Offshore Accidents, Tsunamis and Other Natural Disasters." *Predictive Analytics Today.* 2013 Web. http://www.predictiveanalyticstoday.com/big-data-predict-shore-accidents-tsunamis-natural-disasters/

[30] "New Developments in Big Data Visualization." *USTelecom Media* 2014. Web. http://www.ustelecom.org/blog/new-developments-big-data-visualization#sthash.HefD5H52.dpuf

[31] Quinn, Tom. "New and Unexpected Uses for Scoring Technology." *Credit Score Blog* 2011. Web. http://blog.credit.com/2011/06/new-and-unexpected-uses-for-scoring-technology/

[32] Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal* 2013: Print. http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740

[33] Data overload: Today's experienced clinician needs close to 2 million pieces of information to practice medicine and doctors subscribe to an average of seven journals, representing over 2,500 new articles each year,

making it almost impossible to keep abreast with the latest information about diagnosis, prognosis, therapy and related health issues. "Clinical Decisions Support Systems: The Time Has Come…." Frost & Sullivan, 2009. Print. Market Insight. http://www.frost.com/prod/servlet/cio/181298788

[34] "Clinical Decisions Support Systems: The Time Has Come…." Frost & Sullivan, 2009. Print. Market Insight. http://www.frost.com/prod/servlet/cio/181298788

[35] "Kinect Sign Language Translator Expands Communication Possibilities." *Microsoft Research* 2013. Web. http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx

[36] "Brats 2012 - Multimodal Brain Tumor Segmentation Challenge." *CodaLab*, 2012. Print. https://www.codalab.org/competitions/191

[37] Grady, Denise. "3-D Mammography Test Appears To Improve Breast Cancer Detection Rate." *New York Times* 2014: p. A1 Print. http://www.nytimes.com/2014/06/25/health/breast-cancer-3d-mammography-test-x-ray.html?emc=edit_th_20140625&nl=todaysheadlines&nlid=435891&_r=0

[38] "The game-changing technology that's transforming manufacturing." *Manufacturing Weekly*, January 31, 2014. http://web.archive.org/web/20140131233544/http://www.manufacturingweekly.com/supercomputers/

[39] "The Return on the Data Asset in the Era of Big Data: Capturing the $1.6 Trillion Data Dividend." *Cloud Platform News Bytes Blog* 2015. Web. http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx

[40] Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making.* Cap Gemini, 2012. Web. Point Of View. http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx

[41] A 1 percent productivity increase may seem small, but as Jeff Immelt, CEO of GE puts it, "tell an oil guy you can use software to save him one percent on something, and that guy will be your friend for life." Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines.* GE, 2012. Web. http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf

[42] BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey

[43] Gartner, "Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally To Support Big Data By 2015." 2012. Print. http://www.gartner.com/newsroom/id/2207915

[44] BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey

[45] BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey

[46] According to ESG research, data managed per hospital is expected to increase from 168 terabytes in 2010 to 665 terabytes by 2015. "Digital Imaging in the Cloud." *There Magazine* 2012: 16. Print. http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf

[47] Manyika, James et al. Big Data: *The Next Frontier for Innovation, Competition, and Productivity.* McKinsey Global Institute, 2011. Print. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[48] Researchers trained a machine-learning algorithm on data from 133,000 patients. The model still needs more work to reduce false positives. Rutkin, Aviva. "Machine Predicts Heart Attacks 4 Hours Before Doctors - New Scientist." *New Scientist.* 2014. Web. http://www.newscientist.com/article/mg22329814.400-machine-predicts-heart-attacks-4-hours-before-doctors.html

[49] Ford's modern hybrid Fusion model generates up to 25 GB of data per hour. Hemsoth, Nicole. "How Ford Is Putting Hadoop Pedal To The Metal." *Datanami.* 2013. Web. http://www.datanami.com/2013/03/16/how_ford_is_putting_hadoop_pedal_to_the_metal/

The Chevy Volt contains over 10 million lines of software code, and software developer is one of the fastest growing technical professions in Southeast Michigan, a region long known for its manufacturing prowess. Trop, Jaclyn. "Detroit, Embracing New Auto Technologies, Seeks App Builders." *New York Times.* June 30, 2013. http://www.nytimes.com/2013/07/01/technology/detroit-embracing-new-auto-technologies-seeks-app-builders.html

[50] Miller, Claire Cain. "If Robots Drove, How Much Safer Would Roads Be?" *New York Times* 2014: A3. Print. http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&_r=0

[51] The 787 uses data sensors to reduce fuel, monitor systems, and even employs accelerometers in the nose of the plane to counteract turbulence. If the sensors

register a sudden drop, they immediately tell the wing flaps to adjust (in a matter of nanoseconds) and in so doing, what used to be a 9 feet drop in an older plane can be reduced to just 3 feet in the 787, making for a much smoother flight. Gosling, Kevin. "E-Enabled Capabilities of the 787 Dreamliner." *Aero Quarterly* 2009: 22-24. http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf

[52] Jet engine maker GE says the engine data allows it to figure out things like possible defects 2,000 times as fast as it could before. Hardy, Quentin. "What Cars Did for Today's World, Data May Do for Tomorrow's?" *New York Times* 2014: B7. Print. http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs&region=Body&_r=1&

[53] "Energy-Smart Buildings: Demonstrating how information technology can cut energy use and costs of real estate portfolios." Accenture 2011. http://nstore.accenture.com/corporate-marketing/ccr/2010-2011/Accenture-Energy-Smart-Buildings.pdf

[54] The manufacturing sector stored nearly 2 exabytes of new data in 2010 alone. Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* McKinsey Global Institute, 2011. Print. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

[55] Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* McKinsey Global Institute, 2011. Print. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

[56] "'Data Smart' Strategies for Customers Are Yielding 'Early But Impressive Returns.'" *Microsoft Research the Fire Hose* 2014. Web. http://blogs.microsoft.com/firehose/2014/05/22/data-smart-strategies-for-customers-are-yielding-early-but-impressive-returns/

[57] Somers, Dan. "Manufacturing 4.0 – From Industrialization to Data-Driven Product Lifecycle." *Citizentekk*. 2013. Web. http://citizentekk.com/2013/11/05/manufacturing-4-0-industrialisation-data-driven-product-lifecycle/

[58] McKinsey reports that by using these data driven design techniques, Toyota was able to eliminate 80 percent of defects prior to building the first physical prototype. Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* McKinsey Global Institute, 2011. Print. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

[59] Findings of the New Intelligent Enterprise Study. IBM 2010 *New Intelligent Enterprise Global Executive Study.* 2010. Print.

[60] Geron, Tomio. "Cows in the Cloud: The Hot Startup Moving Farmers into the Cloud." *Forbes* 2012. Web; Helmer, Jodi. "Get Ready For Robot Farmers." Yahoo 2014. Web. https://www.yahoo.com/tech/get-ready-for-robot-farmers-100613764059.html

[61] De-Identification of Personally Identifiable Information, National Institute of Science and Technology, DRAFT NISTIR 8053 (April 2015).

[62] Cavoukian, Ph.D., Ann, and El Emam, Ph.D., Khaled, Dispelling the Myths Surrounding De-Identification: Anonymization Remains a Strong Tool for Protecting Privacy, Information and Privacy Commissioner of Ontario, (June 2011); Cavoukian, Ph.D., Ann, and Daniel Castro Castro. *Big Data And Innovation, Setting The Record Straight: De-Identification Does Work.* ITIF, 2014. Print. http://www2.itif.org/2014-big-data-deidentification.pdf

[63] See for example, Microsoft's add on protecting privacy as their priority https://www.youtube.com/watch?feature=player_embedded&v=bt51MWll1oY

[64] Apple, *Government Information Requests*, noting that the company has incorporated state-of-the-art encryption into its iPhone operating system so that "your personal data such as photos, messages (including attachments), email, contacts, call history, iTunes content, notes, and reminders is placed under the protection of your passcode," at www.apple.com/privacy/government-information-requests/

[65] Hachman, Mark, "Microsoft's updated privacy policy makes It clear it's not selling ads against your words," *PCWorld*, June 11, 2014, http://www.pcworld.com/article/2362130/microsofts-updated-privacy-policy-makes-it-clear-its-not-selling-ads-against-your-words.html

Timberg, Craig. "Newest Androids Will Join IPhone In Offering Default Encryption, Blocking Police." *Washington Post* 2014: Print. http://www.washingtonpost.com/blogs/the-switch/wp/2014/09/18/newest-androids-will-join-iphones-in-offering-default-encryption-blocking-police/

66 Data broker Acxiom opens consumer-facing data website, offers opt-out http://cir.ca/news/acxiom-gives-consumers-data-peek

67 BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey

68 McKinsey Global Institute. Internet Matters: *The Net's Sweeping Impact On Growth, Jobs, And Prosperity.* McKinsey & Co., 2011. Print.

69 Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* McKinsey Global Institute, 2011. Print. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

70 According to Salaries of Data Scientists, an April 2014 study from Burtch Works.

71 Bagley, Rebecca. "How The Cloud And Big Data Are Changing Small Business." *Forbes* 2014. Web. http://www.forbes.com/sites/rebeccabagley/2014/07/15/how-the-cloud-and-big-data-are-changing-small-business/

72 BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey

73 Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making.* Cap Gemini, 2012. Web. Point Of View. http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx

74 Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making.* Cap Gemini, 2012. Web. Point Of View. http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx

75 Gerbis, Nicholas. "10 Correlations That Are Not Causations." *HowStuffWorks.* 2015. Web. http://science.howstuffworks.com/innovation/science-questions/10-correlations-that-are-not-causations.htm

76 Vesset, Dan, Henry D. Morris, and John F. Gantz. *Capturing the $1.6 Trillion Data Dividend.* IDC, 2014. Print. IDC White Paper.

77 Westerman, George, Didier Bonnet, and Andrew McAfee, "The Advantages of Digital Maturity." November 2012, MIT Sloan.

78 "Meeting the Big Data Challenge: Don't Be Objective." *Forbes* 2013. Web. http://www.forbes.com/sites/darden/2013/02/01/meeting-the-big-data-challenge-dont-be-objective/

79 IDG Enterprise's 2014 Big Data research. IDG. *CEOs Call for Big Data and IT Continues To Lead Investment Decisions.* 2014. Print. http://www.idgenterprise.com/press/ceos-call-for-big-data-and-it-continues-to-lead-investment-decisions

80 Miller, Claire Cain. "If Robots Drove, How Much Safer Would Roads Be?" *New York Times* 2014: A3. Print. http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&_r=0

81 Clemens, Samuel. "7 Facts about Data Quality [Infographic]." *InsightSquared*. January 3, 2012. Web. http://www.insightsquared.com/2012/01/7-facts-about-data-quality-infographic/

82 Economist Intelligence Unit. *Big Data Harnessing a Game-Changing Asset*. SAS, 2011. Web. http://www.sas.com/resources/asset/SAS_BigData_final.pdf

83 "The Return on the Data Asset in the Era of Big Data: Capturing the $1.6 Trillion Data Dividend." *Cloud Platform News Bytes Blog* 2015. Web. http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx

84 BSA/IPSOS Global Data Analytics Poll, November 2014, www.bsa.org/datasurvey

85 IT investments in the entire US economy, including retail, through the high-growth 1990s added 1 percent to 2 percent to the compound annual growth rate of US productivity. Farrell, Diana et al. *How IT Enables Productivity Growth. San Francisco: McKinsey Global Institute High Tech Practice*, 2002. Print. http://www.mckinsey.com/insights/business_technology/how_it_enables_productivity_growth

86 National Bureau of Economic Research. NBER Working Paper No. 18315*: Is U.S. Economic Growth Over? Faltering Innovation Confronts The Six Headwinds.* 2012. Print. http://www.nber.org/papers/w18315

87 Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. Web. Point Of View. http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx

88 Gertner, Joey. "GE for Making the 'Internet of Things' Real." *Fast Company* 2014. Web. http://www.fastcompany.com/most-innovative-companies/2014/ge

[89] Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines.* GE, 2012. Web. http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf

[90] "City Of Barcelona Realizes Vision of Innovative City Governance with Cloud, Devices, and Apps." Customers.microsoft.com. 2014. Web. https://customers.microsoft.com/Pages/Home.aspx

[91] "Autodesk the Gallery Masdar Headquarters Positive Energy Building." *Autodesk.com*. 2015. Web. http://www.autodesk.com/gallery/exhibits/currently-on-display/adrian-smith-gordon-gill-architecture-masdar-headquarters

[92] Bunge, Jacob. "Big Data Comes To The Farm, Sowing Mistrust." *Wall Street Journal* 2014. Print. http://online.wsj.com/news/articles/SB10001424052702304450904579369283869192124

"Supply Chain Management Solution for Fast Moving Consumer Goods & Food Industries - Farm to Fork Tech Mahindra." *Techmahindra*. 2015. Web. http://www.techmahindra.com/en-US/wwd/solutions/Pages/Enterprises/retail_farm_fork.aspx

[93] Between 2013 and 2020 the division of the digital universe between mature and emerging markets (e.g., China) will switch — from 60 percent accounted for by mature markets to 60 percent of the data in the digital universe coming from emerging markets. EMC Digital Universe. *Executive Summary Data Growth, Business Opportunities, and the IT Imperatives.* IDC, 2014. Print. http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

[94] Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World.* Accenture, 2013. Print. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf

[95] Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World.* Accenture, 2013. Print. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf

[96] Levy, Stephen. "Bill Gates and President Bill Clinton on the NSA, Safe Sex, and American Exceptionalism." *Wired* 2013: Print. http://www.wired.com/2013/11/bill-gates-bill-clinton-wired/2/

Chhachhar, Abdul Razaque, and Siti Zobidah Omar. "Use of Mobile Phone among Fishermen for Marketingand Weather Information." *Archives Des Sciences* 65.8 (2012): 107-119. Print. http://www.academia.edu/4592505/Use_of_Mobile_Phone_among_Fishermen_for_Marketing_and_weather_information

[97] Neuman, William, and Ralph Blumenthal. "New to the Archaeologist's Tool Kit: The Drone." *New York Times* 2014. Print. http://mobile.nytimes.com/2014/08/14/arts/design/drones-are-used-to-patrol-endangered-archaeological-sites.html?_r=1&referrer

[98] "Forty Thousand Relics to Be Digitized In Five Years." *Thanhnien News*. 2010. Web. http://www.thanhniennews.com/entertainment/forty-thousand-relics-to-be-digitized-in-five-years-22816.html

[99] Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World.* Accenture, 2013. Print. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf

[100] Future of Privacy Forum. *Big Data: A Tool for Fighting Discrimination and Empowering Groups.* Future of Privacy Forum and Anti-Defamation League, 2014. Print. http://www.futureofprivacy.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf

[101] Wladawsky-Berger, Irving. "The Changing Nature of Globalization in Our Hyperconnected, Knowledge-Intensive Economy." *Wall Street Journal* 2014. Print. http://blogs.wsj.com/cio/2014/06/20/the-changing-nature-of-globalization-in-our-hyperconnected-knowledge-intensive-economy/?mod=wsj_ciohome_cioreport

[102] For example, Argentina, Australia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Greece, Hong Kong, India, Indonesia, Korea, Mexico, Peru, Russia, Switzerland and Vietnam have adopted or have proposed rules that prohibit or significantly restrict companies from transferring personal information out of their respective domestic territories.

ABOUT BSA | THE SOFTWARE ALLIANCE

BSA | The Software Alliance (www.bsa.org) is the leading advocate for the global software industry before governments and in the international marketplace. Its members are among the world's most innovative companies, creating software solutions that spark the economy and improve modern life.

With headquarters in Washington, DC, and operations in more than 60 countries, BSA pioneers compliance programs that promote legal software use and advocates for public policies that foster technology innovation and drive growth in the digital economy.

www.bsa.org

BSA Worldwide Headquarters

20 F Street, NW
Suite 800
Washington, DC 20001

T: +1.202.872.5500
F: +1.202.872.5501

BSA Asia-Pacific

300 Beach Road
#25-08 The Concourse
Singapore 199555

T: +65.6292.2072
F: +65.6292.6369

BSA Europe, Middle East & Africa

2 Queen Anne's Gate Buildings
Dartmouth Street
London, SW1H 9BP
United Kingdom

T: +44.207.340.6080
F: +44.207.340.6090