+ t a b l e a u ®
S O F T W A R E

What's the Big Deal About Big Data?

January 2013

A New Global Currency

Every year a gaggle of the world's best and brightest minds descend on Davos, Switzerland, to discuss many of the world's most important new developments. In 2012 for the first time, data made the cut as a critical new form of economic currency that, in the eyes of these global thought leaders, is as potent a force as gold, oil, or money itself.[1]

"It's a revolution," says Gary King, director of Harvard's Institute for Quantitative Social Science. "The march of quantification, made possible by enormous new sources of data, will sweep through academia, business, and government. There is no area that is going to be untouched."[2]

While with any new concept there is a risk of hyperbole, big data is driving significant opportunities across industries and the public sector. And as more and more data-generating resources emerge and find their way into the hands of hundreds of millions of new global users, data volumes will continue to increase by orders of magnitude.

"Data, data, and more data … data is everywhere, and it's important," says Ravi Kalakota, CEO of E-Business Strategies. "By 2015 nearly 3 billion people will be online, pushing the data created and shared to nearly 8 zettabytes."[3] For comparison purposes, that is the informational equivalent of 1.8 million new Libraries of Congress being created every year.

Gartner defines big data as "high volume, velocity, and variety information assets that demand cost-effective, innovative forms of information process for enhanced insight and decision making." Obviously, the definition of big data is relative to an organization's capacity to generate, manage, and make sense of it. Smaller companies may consider many terabytes of data "big" while their larger counterparts don't assign it that value until it pushes into the petabytes or more.

Not surprisingly, this rapid onset of the "industrial revolution of data"[4] has taken many organizations by surprise and led to more than a few cases of what some industry sources call "data analysis paralysis."[5] However, this wealth of new data offers exciting new opportunities for spearheading innovation and growth.

Indeed, key determinants to corporate success will be the willingness of organizations not merely to adapt to these changes, but to eagerly embrace this data windfall and adopt new technology and analytical best practices capable of making the most of it.

The Data on Data

Seeds of the Data Explosion

At some point in time most children become fascinated with big numbers and, when their budding vocabularies are no match for all those zeroes, they'll cook up definitions of their own (think gajillion, bazillion, and so on). We say this because at some point a child's imagination may be enlisted to coin a term capable of defining the staggering volume of data created by organizations, the Internet, and its digital peripherals.

And there is every indication that we've only just begun; that the mountains of data already generated by IT system logs, forms, multimedia files, email, social media feeds, Web analytics, metadata, mobile devices, and countless other applications are but the proverbial tip of the iceberg; that, in the words of Google Chairman Eric Schmidt, when it comes to data "people [and organizations] aren't nearly ready for the technology revolution that's going to happen to them."[6]

Consider that just two years ago EMC and IDC jointly predicted an astounding 45-fold annual data growth rate through 2020, yet today experts suggest that figure may seriously underestimate the potential.[7] This year alone IDC predicts that 1.2 zettabytes of raw (or, in analytical parlance, "unstructured") data will be generated—that's 1,200,000,000,000,000,000,000 to the awe-struck child within.

The simple truth is that as more and more of the connective tissue that defines, feeds, and grows the digital Web "comes online" (e.g., content channels, user platforms, data applications) and billions more people

embrace it, nobody really knows just how much data all of that gadgetry in the hands of all those individuals will produce.

The Rise of Unstructured Data

During the early years of the digital age, data wasn't nearly so thorny an issue for organizations. Most companies and their IT departments implemented relational data management systems that, by definition, had their structural intelligence already baked in. This, in turn, enabled them to use basic desktop spreadsheets systems for analysis.

But as the technology pendulum swings squarely into the hands of individuals; as a seemingly endless cavalcade of new data sources emerge; and as data volumes explode, today's organizations are finding themselves overwhelmed by data that falls "outside of the structured rows and columns of databases that enterprises have traditionally looked to as their primary sources of information," says Ananth Krishnan, CTO of $10.2 billion Tata Consultancy Services.[8]

Even for technically sophisticated organizations, the task of preparing for, managing, and analyzing disparate data types can be enormously challenging. Add in massive amounts of the stuff and an organization can quickly reach a kind of information overload. Worse still,

many companies play it safe and focus only on the data and analytical practices with which they already are comfortable.

"The problem with this reasoning," says Thomas Davenport, a senior advisor to Deloitte Analytics, "is that the advance of big data shows no signs of slowing. If companies sit out this trend's early days … they risk falling behind as competitors and channel partners gain nearly unassailable advantages."[9]

What's in it for Business?

Algorithms Over Instinct

Throughout history businesses largely depended on the veteran wisdom of their leaders for making important tactical and strategic decisions. The fates of these organizations often hinged on the outcomes of these educated, yet nevertheless subjective choices. Today, however, the rapid ascendance of data and its ability to deliver critical, objective insights into enterprise operations, consumer behavior, and more, is changing all that.

"Organizational judgment is in the midst of a fundamental change," says Erik Brynjolfsson, director of MIT Sloan's Center for Digital Business. Businesses are being forced to transition "from a reliance on a leader's 'gut instinct' to increasingly data-based analytics."[10]
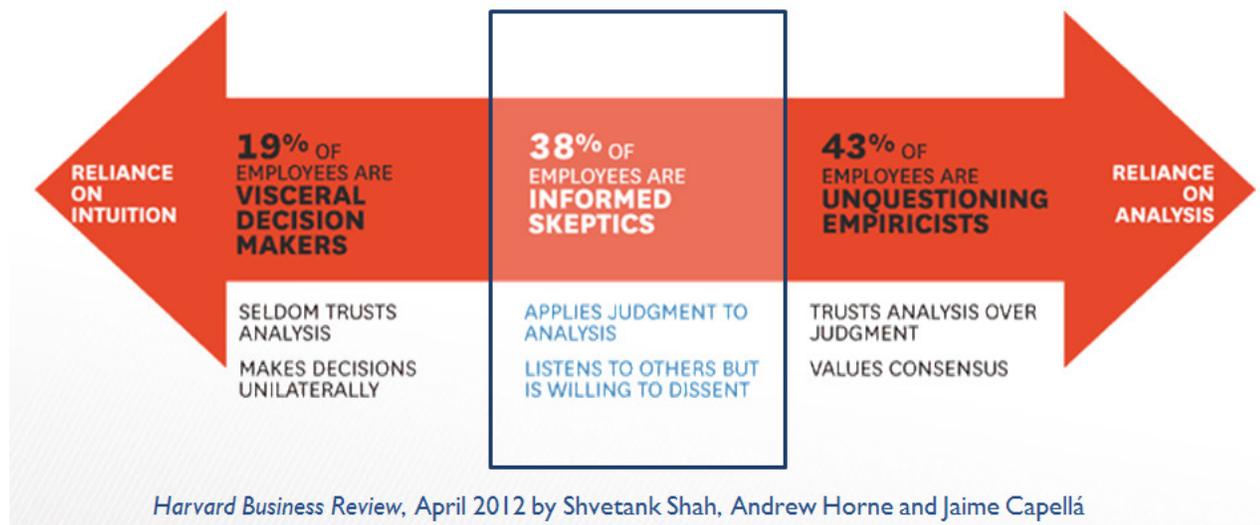

RELIANCE ON INTUITION — 19% OF EMPLOYEES ARE **VISCERAL DECISION MAKERS** — SELDOM TRUSTS ANALYSIS / MAKES DECISIONS UNILATERALLY

38% OF EMPLOYEES ARE **INFORMED SKEPTICS** — APPLIES JUDGMENT TO ANALYSIS / LISTENS TO OTHERS BUT IS WILLING TO DISSENT

43% OF EMPLOYEES ARE **UNQUESTIONING EMPIRICISTS** — TRUSTS ANALYSIS OVER JUDGMENT / VALUES CONSENSUS — RELIANCE ON ANALYSIS

Harvard Business Review, April 2012 by Shvetank Shah, Andrew Horne and Jaime Capellá

Figure 1: Companies Need "Informed Skeptics" Who Combine Data and Business Experience

" Organizational judgment is in the midst of a fundamental change—from a reliance on a leader's 'gut instinct' to increasingly data-based analytics. "

– Erik Bryjolfsson, Director, MIT Sloan Center for Digital Business

In the Harvard Business Review, three executives from the Corporate Executive Board advocate a middle ground of "informed skeptics" who take the best of intuition and analysis. Gut instinct alone is imperfect, as is accepting data without asking intelligent questions (e.g., is our data accurate and complete, or is this data showing just one piece of a broader puzzle?).[11]

Consider the example of Xerox hiring practices for its call centers. As in other companies, Xerox management instinctively believed that work experience mattered most in determining the best candidates to hire. But new data analysis suggested otherwise. The company opted to go with the data and saw its new employee attrition rate fall by 20 percent—no small feat for a company with nearly 50,000 call center employees who cost upward of $5,000 each to train. "Some of the assumptions we had weren't valid," acknowledged Connie Harvey, Xerox's COO of Commercial Services.[12]

In a recent study, MIT's Brynjolfsson found that U.S. companies using data-guided management processes enjoyed productivity gains upward of 6 percent—significant figures in a tight economy where margins are often razor-thin. And in one international study of more than 600 C-level executives, big data was said to have improved business performance by an astounding 26 percent "and that impact will grow to 41 percent over the next three years."[13]

What Xerox and these other companies are discovering is that hidden within their data are the insights to reduced costs, improved performance, better customer care, and all-around smarter business decisions.

Concludes Davenport: "If your organization stores multiple petabytes of data, if the information most critical to your business resides in forms other than rows and columns of numbers, or if answering your biggest question would involve a 'mashup' of several analytical efforts, you've got a big data opportunity."[14]

Smarter Answers Leading to Smarter Questions

To make the very most of this data, however, companies also must understand the nature of the data. In other words, it isn't just that today's companies are awash in data, it's that they increasingly don't know what they're awash in.

Today's data analytics isn't just about finding needles in haystacks, says Haydn Shaughnessy, a social innovation expert with the Next Generation Enterprise. "Enterprises actually don't know what is in the stack, so they don't know if they are looking for needles or something else."[15]

By capturing, contextualizing, comparing, and analyzing all of this new data, unexpected patterns, trends, and possibilities emerge. "Pulling data together from different domains—such as retail location and spending—will, I think, lead to innovation," says Ovum Research Director Denise Montgomery, who these days tells her banking clients that geographical data is more important than traditional credit card and transactional data for mining new business opportunities.[16]

Similarly, an ambitious new U.N. program known as Global Pulse (shown below) has found innovative new ways of using mobile phone data to "provide critical information relevant to areas like urban planning, crisis management, and global health";[17] and American epidemiologists have happily discovered that by monitoring Google search queries they can more easily predict and pinpoint influenza outbreaks.

Abhi Mehta, the founder and CEO of Tresata and a pioneer in the big data industry foresees an era of "data production lines" capable of processing data the way manufacturing once processed raw materials, the difference being that the output will deliver actionable insights into things like customer motivations leading to better business decisions.

The point is that there are virtually unlimited ways for companies to learn—and grow—from today's data revolution. "Data is the new soil—a kind of ubiquitous resource that can be used to develop new innovations and insights," says David McCandless, author of "Information is Beautiful."[18]

Or as Farnam Jahanian, director of the National Science Foundation's Computer and Information Science directorate sees it: "Data, in my view, is a transformative new currency for science, engineering, education, commerce, and government. Foundational research in data management and data analytics promise breakthrough discoveries and innovations across all disciplines."[19]

3 Steps to Big Data Success

For many companies, big data brings both challenges and opportunities. To be sure, all of these organizations are hungry for the analytical gems that will help them cut costs, improve service, and give them a competitive edge. Yet most of these organizations are still accustomed—and dependent on—the same modes of data analysis or IT-driven protocols that increasingly are not up to the task, including:

- Spreadsheet-driven modes of analysis. While programs like Excel can be extremely useful, they also are limited in their ability to accommodate large volumes of data or disparate data types or easily share results across an organization.[20]

- Data exists in silos, with expensive complex manual processes for data integration, versus the ability to easily connect and blend data from a variety of data sources and data types.

- IT spends an inordinate amount of time and resources to build custom reporting, while business users wait in a queue to get questions answered or detail queries in advance versus the ability to explore data for new, unexpected insights.

The simple truth is that the volume, velocity, and variety of data is forcing companies to recognize that that data

"must now be analyzed using new methodologies foreign to their IT departments."[21]

"Those reports could be built and altered in seconds or minutes, not days or weeks, and they don't require specialized programmers," says Tableau CEO and co-founder, Christian Chabot.[22]

How, then, does a modern organization go about leveraging the enormously important commodity known as big data within time and budget constraints? Herewith, three essential ingredients to any successful big data analytics implementation plan.

1 Corporate Adoption

It is imperative that organizations "establish a culture that values and rewards the use of big data" in their decision making processes.[23] Such a commitment is necessary if big data's true potential is to be realized.

Furthermore, they must be flexible and open-minded enough to welcome new insights that inevitably will be borne out of the data, to "let the dataset change the mindset."[24]

2 Encourage "T-shaped Skills"

Rather than relegating analytics to some backroom office, companies should empower staff across the organization to adopt analytics into their daily efforts. Dr. DJ Patil, data scientist in residence at venture capital firm Greylock Partners, uses the popular "Star Trek" series as an analogy. Rather than keeping your analytics people tucked away in the engine room a la Scotty and only turning to them when something is broken, look for more Spock-types who are capable of using data-driven decision making skills on the bridge and in real time.

Hire, mentor, and empower staff with "T-shaped skills" that combine deep business and technical expertise (the T's vertical line) together with the ability to collaborate with colleagues in other departments and functions (the T's horizontal line). For example, a major retailer holds a training session every six months to educate every full-time employee on how to understand

" When you come across visualizations of the data it's almost a relief—like coming across a clearing in the jungle. "

– David McCandless, Author, Information Is Beautiful

a financial statement, compare data across stores, or use data about buyer personas to know best how to target each person who walks into a store. In other words, that retailer is operationalizing big data.

This is not to say that back-end systems aren't important or won't require upgrades to tackle the volume, variety and velocity of big data, or that you won't have to fix outdated business processes. But a server or process flowchart diagram by itself won't tell you how to outperform your competitors.

3 Visualize

It is impossible to overstate the importance of being able to quickly and easily manipulate disparate data sources and visualize the results in graphically rich, equally diverse ways. "There's something almost magical about visual information—it's effortless," says Information is Beautiful author McCandless. "If you're struggling with dense amounts of data, when you come across visualizations of the data it's almost a relief—like coming across a clearing in the jungle."[25]

Indeed, biological studies have confirmed that the human brain is wired best for visual understanding, enabling viewers to more easily make the necessary connections that are of most importance to an organization. The eye is "extremely good at tasks such as edge detection, shape recognition, and pattern matching," says Noah Illinsky, technical editor of "Beautiful Visualization."[26] It is in the pattern matching that we learn the most. "Typically, the important messages in data are represented in the patterns and pattern violations: trends, gaps, and outliers. This is the interesting stuff. This is meaning. This is what we go to the data hoping to find."

Additional benefits include the ability to pack far more information into a visualization than in raw data, and visual data's ability to more easily and effectively change someone's mind. Furthermore, visualization technology like that offered by Tableau Software enables non-technical users to easily compare data across categories, siloes, data types, and other old-world firewalls, and share results with colleagues. Notes Kiril Evtimov, eBay's analytics program director, "Tableau's capabilities and

ease of use enable eBay's teams to take a collaborative approach to exploring data—and to making results available seamlessly across the business."
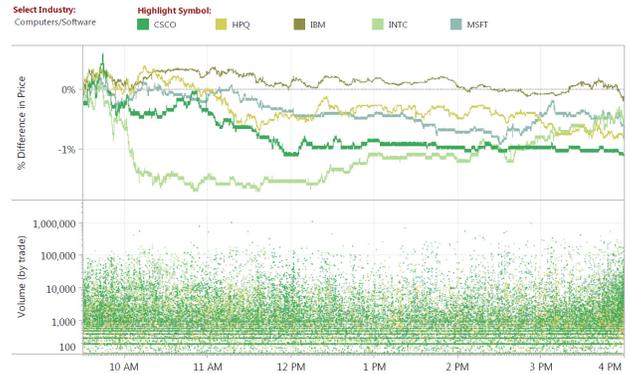

Figure 2: Example of Financial Big Data Showing Stock Trading Pricing and Volumes.

Wrapping Up

Big data. The phrase itself along with the incomprehensibly large numbers that it represents can easily intimidate. If anything, however, we believe big data should elicit thoughts of opportunity and promise.

Passionately embraced, carefully planned for, and appropriately managed through the right visual analytics technology, big data can deliver exciting new glimpses into a company's operations, employees, customers, competitive environment, and much more.

By bringing together new and existing data from multiple disparate sources, the organization can quickly, easily, and clearly identify the kinds of contextually-based patterns, connections, and insights previously invisible to it. This not only allows the company to improve results and cut costs—efforts once predicated on highly subjective, instinct-based decision making practices—it also brings to the surface potential new approaches to virtually any aspect of its operations.

"Big data's potential goes beyond traditional 'rear view' business intelligence," says Howard Elias, president and COO of EMC Information Infrastructure and Cloud Services. Analytics increasingly will reveal patterns that

let organizations make a "quantum leap from incremental improvement to predictive business processes and even entirely new business models."[27]

The good news is that big data analytics is not rocket science and you can get started now. The tools and practices already exist to start delivering the benefits of big data analytics in companies of all sizes and industries.

Sources

1. "Big Data, Big Impact," World Economic Forum, Jan. 2012

2. "The Age of Big Data," NY Times, Feb. 11, 2012

3. "Gartner 2012 Top 10 Strategic Tech Trends," Business Analytics 3.0, Nov. 2011

4. Joe Hellerstein, Chancellor's Professor of Computer Science, University of California at Berkeley

5. "How to Avoid Big Data Analysis Paralysis," Chief Marketer, July 31, 2012

6. "The New Factors of Production and the Rise of Data-Driven Applications," Forbes, Oct. 31, 2011

7. "The Digital Universe Decade - Are You Ready?" – May 2010

8. "The Unstructured Information Most Businesses Miss Out On," MIT Sloan Management Review, April 2011

9. "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review, Oct. 2012

10. "Strength in Numbers: How Does Data-Driven Decision Making Affect Firm Performance?" – MIT, Penn, 2011

11. Harvard Business Review, April 2012, by Shvetank Shah, Andrew Horne and Jaime Capellá

12. "Meet the New Boss: Big Data," The Wall Street Journal, Sept. 20, 2012

13. "Capgemini Reports Shows Rising Impact of Big Data on Decision Making," June 12, 2012

14. "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review, Oct. 2012

15. "Solving the Unstructured Data Problem," NGE Think Tank, Jan. 2011

16. BankTech 2012 Conference

17. "Enhancing Public Policy Decision Making Using Large-Scale Cell Phone Data," Sept. 4, 2012

18. TED: The Beauty of Data Visualization, Aug. 2010

19. "New U.S. Research Will Aim at Flood of Digital Data," NY Times, March 29,2012

20. "Excel: Great Hammer, Lousy Screwdriver," Tableau Software

21. "Gartner Predicts 2012: Four Forces Combine to Transform the IT Landscape," Gartner Research

22. "Big Data Success Stories: Tableau Software," The Wall Street Journal, Nov. 23, 2011

23. "Big Data: The Next Frontier for Innovation, Competition and Productivity," McKinsey Global Institute

24. Hans Rosling, chairman of Gapminder Foundation

25. TED: The Beauty of Data Visualization, Aug. 2010

26. "Why is Data Visualization So Hot?" Visual.ly Blog, March 2012

27. "The Big Data Challenge: How to Develop a Winning Strategy," CIO, June 14, 2012

About Tableau

Tableau Software helps people see and understand data. According to IDC in its 2012 report, Tableau is the world's fastest-growing business intelligence company, Tableau helps anyone quickly analyze, visualize and share information. More than 10,000 organizations get rapid results with Tableau in the office and on-the-go. And tens of thousands of people use Tableau Public to share data in their blogs and websites. See how Tableau can help you by downloading the free trial at *www.tableausoftware.com/trial*.

Tableau and Tableau Software are trademarks of Tableau Software, Inc. All other company and product names may be trademarks of the respective companies with which they are associated.