

: Does the term big data have utility for managers?

by Daniel J. Power

Editor, DSSResources.COM

Big data is a colorful phrase for a significant change in data capture, retrieval and storage. Each day, every one of us generates very large amounts of digital data. We send and receive email, visit Web sites and make online purchases, use tools like Google Docs, make phone calls, upload photos to Facebook, use Google Search, chat with friends, take our cars for service, workout at a gym on a machine with an Internet connection, pay bills online, and have our utilities, Internet and cable usage recorded. This data and much more from our activity is recorded and often backed-up in the Cloud. Now we can capture, store and perhaps analyze the data incidental to personal and organization activities and actions. Machine data can also be stored and analyzed.

Some information technology vendors regularly over promote technology opportunities and that has happened with Big Data and analytics. Some managers very quickly get disillusioned and that is happening with the ambiguous concept of Big Data. It is a fact that data is accumulating faster than we can analyze it. Like Coleridge's ancient mariner who could not drink the vast amounts of salty ocean water, often we can not use and interpret the vast amounts of machine and unstructured data that is now available.

Machine data is a major contributor to the Big Data revolution. Machine data is all of the data generated by a computing machine while it operates. Examples of machine data include: application logs, clickstream data, sensor data and Web access logs (cf., Power, 2013b).

Devlin (March, 2013) identifies 3 categories of Big Data "process-mediated data - the well-defined, well-structured and well-managed data residing in current operational and informational systems - is growing fast and can drive significant new value through operational analytic approaches. Human-sourced information - currently mostly about social media - and machine-generated data are emerging and rapidly growing sources of knowledge about people's behaviors and intentions."

Aziza, Ehrenberg, Franks, Morris and others argue that the potential of Big Data for improving our personal lives, helping businesses compete, and governments provide services is unbounded. According to Ehrenberg, "Greater access to data and the technologies for managing and analyzing data are changing the world." Somehow Big Data will lead to better health, better teachers and improved education, and better decision-making.

Provost and Fawcett (2013) define big data as "datasets that are too large for traditional

: Does the term big data have utility for managers?

data-processing systems and that therefore require new technologies" with names like Hadoop, Hbase, MongoDB or CouchDB. Ehrenberg notes that when he first used the term 'big data' in 2009 to label a new ventures fund that the term "implied tools for managing large amounts of data and applications for extracting value from that data". Venture Capitalist Bryce Roberts reminds us "Data, big, medium or small, has no value in and of itself. The value of data is unlocked through context and presentation." How data is presented can change behavior.

Digital data is massive. For example, an *Economist* magazine special report notes, Wal-Mart "handles more than 1 million customer transactions every hour, feeding databases estimated at more than 2.5 petabytes — the equivalent of 167 times the books in America's Library of Congress ..." Data comes from both new and old sources and the increased volume of data led some vendors and industry observers to proclaim the era of 'Big Data'. IBM researchers (Zikopoulos et al, 2013) describe Big Data in terms of four dimensions: Volume, Velocity, Variety, and Veracity. See [IBM "What is big data?"](#)

Effectively using Big Data involves managing 1) the platform for storing and accessing data, 2) the analytics, BI and decision support capabilities, and 3) the policies and procedures for governing and managing data including issues of privacy, ethical use and retention (cf., Dyche, 2013). Dyche asserts the "hard part of big data is managing it."

Recently some bloggers have become disillusioned by the term Big Data, but realize the potential. For example, Barry Devlin (2013) argues "Big data as a technological category is becoming an increasingly meaningless name." De Goes (2013) further asserts "The phrase 'big data' is now beyond completely meaningless." Sorofman (2013) considers Big Data "a cute way of describing the idea of data processed at massive scale and speed, where the trail thrown off by all of our varied digital interactions and experiences becomes the fuel for decisions, insights and actions." The ongoing challenge for decision support and information technology researchers is identifying use cases or user examples for analyzing the large volume of semi- and unstructured data that is accumulating.

Managers need to understand what to do with new data sources and few managers want to blindly hire high salary data scientists to work magic and find new strategic insights. Managers want to understand what a data scientist will do and why someone is needed in that role. Managers also seem reluctant to purchase more expensive hardware and software to store data that may not be useful. Big is not necessarily needed or better data. Aziza in his critique notes "we need a different and more mainstream way to think about Big Data".

Lopez (2013) reports Economist Prassanna Tambe "found that the use of big data technologies correlates with significant additional productivity growth". It is unclear what study by Tambe shows

: Does the term *big data* have utility for managers?

such a positive result for the ambiguous Big Data concept. Tambe and Hitt (2012) examined IT returns from 2000-2006, a period prior to the Big Data phenomenon. It is doubtful they would make such a broad, unscientific claim.

An *Economist* (2010) special report cautions us about misanalysis of Big Data. The report explains that "During the recent financial crisis it became clear that banks and rating agencies had been relying on models which, although they required a vast amount of information to be fed in, failed to reflect financial risk in the real world. This was the first crisis to be sparked by big data — and there will be more."

Information technology educators need to help prepare data scientists who have the skills of a database designer, software programmer, statistician and storyteller. Davenport and Patil (2012) describe the job of a data scientist in more detail. In general, data scientists prepare three major types of analyses with Big Data (Power, 2013a):

- 1) Retrospective data analysis — uses historical data and quantitative tools to understand patterns and results to make inferences about the future. This is the area of business intelligence.
- 2) Predictive data analysis — uses simulation models to generate scenarios based on historical data to understand the future. Predictive means "looking forward" and making known in advance.
- 3) Prescriptive data analysis - uses planned, quantitative analyses of real-time data that may trigger events. Prescriptive means recommending.

The terms "big data" and "analytics" have created very high expectations for better decisions. Big data is useful only if the data called "big data" is appropriately used in analysis. The term has limited usefulness as a descriptive label. We need to explore and document business use cases of machine, social media and other data sources and help prepare professionals to manage and analyze these new types of data. The term Big Data is increasingly meaningless and the expectations for analytics improving decisions are too high, but extensive digital data can be captured and analyzed and in many companies we should definitely do so.

References

: Does the term big data have utility for managers?

Aziza, B., "Big Data "A-Ha" Moment?" Forbes, 2/25/2013 at URL
http://www.forbes.com/sites/ciocentral/2013/02/25/big-data-a-ha-moment/?goback=%2Egde_4732551_member_218623809 .

Davenport, T. H., P. Barth and R. Bean, "How 'Big Data' is Different," MIT Sloan Management Review, Vol. 54, No. 1, Fall 2012.

Davenport, T.H. and D.J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review, October 2012.

De Goes, J. "'Big data' is dead. What's next?" Venturebeat.com guest blog post, February 22, 2013, at URL
http://venturebeat.com/2013/02/22/big-data-is-dead-whats-next/?goback=%2Egde_62438_member_217099766

Devlin, B. "Big Analytics rather than Big Data," B-eye-Network blog, February 5, 2013 at URL
http://www.b-eye-network.com/blogs/devlin/archives/2013/02/big_analytics_r.php

Devlin, B. "Big Data - Please, Drive a Stake through its Heart!," B-eye-Network blog, March 4, 2013 at URL http://www.b-eye-network.com/blogs/devlin/archives/2013/03/big_data_-_plea.php

Dyche, J., "Big Data's Three-Legged Stool," Information Management, March 13, 2013 at URL
<http://www.information-management.com/news/big-data-three-legged-stool-10024077-1.html> .

Ehrenberg, R. "What's the big deal about Big Data?" InformationArbitrage.com blog post, January 19, 2012 at URL
<http://informationarbitrage.com/post/16121669634/whats-the-big-deal-about-big-data>.

Franks, B., Taming the Big Data Tidal Wave, Hoboken, NJ: Wiley, 2013.

: Does the term big data have utility for managers?

IBM, "What is big data?" accessed 3/6/2013 at URL <http://www-01.ibm.com/software/data/bigdata/>

IDC. "Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO," September 2011.

Ignatius, A., "From the Editor: Big Data for Skeptics," Harvard Business Review, October 2012.

Klein, P., "Tom Davenport on Building Big Data and Analytics Capabilities," The MIT Center for Digital Business blog, August 27, 2012 at URL http://digitalcommunity.mit.edu/community/featured_content/blog/2012/08/27/tom-davenport-on-building-big-data-and-analytics-capabilities.

Lopez, I., "Data Science and the Decision-maker in the Machine," datanami, February 25, 2013 at URL http://www.datanami.com/datanami/2013-02-25/data_science_and_the_decision-maker_in_the_machine.html

Manyika, J., Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, May 2011 at URL http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation .

Morris, J. , "Top 10 categories for Big Data sources and mining technologies," ZDNet, July 16, 2012 at URL <http://www.zdnet.com/top-10-categories-for-big-data-sources-and-mining-technologies-7000000926/> .

Power, D. "What is Hadoop?" Decision Support News, Vol. 12, No. 23, November 13, 2011 at URL <http://dssresources.com/faq/index.php?action=artikel&id=235>.

Power, D., Decision Support, Analytics, and Business Intelligence, Second Edition, New York, NY:

: Does the term big data have utility for managers?

Business Expert Press, 2013a.

Power, D. "What is machine data?" Decision Support News, Vol. 14, No. 02, January 20, 2013b at URL <http://dssresources.com/faq/index.php?action=artikel&id=255>.

Provost, F. and T. Fawcett, Data Science for Business: Fundamental principles of data mining and data-analytic thinking, O'Reilly, 2013 (<http://people.stern.nyu.edu/fprovost/>).

SAS, "Big Data – What Is It?" last accessed March 5, 2013 at URL <http://www.sas.com/big-data/>.

Sorofman, J., "Data, Data Everywhere," Gartner Blog Network, February 28, 2013 at URL <http://blogs.gartner.com/jake-sorofman/data-data-everywhere/>

Roberts, B. "Data Data Everywhere and Not a Drop of Value," <http://bryce.vc> blog, February 2012 at URL <http://bryce.vc/post/15300645787/data-data-everywhere-and-not-a-drop-of-value> .

Scoble, R. (2010) Interview with Cloudera CEO Mike Olson, What is Hadoop?, YouTube video, March 4 retrieved from <http://youtu.be/S9xnYBVqLws>.

Special Report, "Data, data everywhere," *Economist* (print and digital), February 25, 2010 at URL <http://www.economist.com/node/15557443> .

Tambe, P. and L. M. Hitt, "The Productivity of Information Technology Investments: New Evidence from IT Labor Data," Information Systems Research, 23(3-1), 599-617, 2012 at URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1180722 .

Vesset, D. B. Woo, H. D. Morris, R. L. Villars, G. Little, J. S. Bozman, L. Borovick, C. W. Olofson, S. Feldman, S. Conway, M. Eastwood, N. Yezhkova (March 2012) Worldwide Big Data Technology and Services 2012–2015 Forecast, International Data Corporation (IDC) at URL <http://www.idc.com/getdoc.jsp?containerId=233485> .

: Does the term big data have utility for managers?

West. G. (2013) Big Data Needs a Big Theory to Go with It, Scientific American, May 5, last accessed April 20, 2013 at URL <http://www.scientificamerican.com/article.cfm?id=big-data-needs-big-theory> .

Zikopoulos, P., D. deRoos, K. Parasuraman, T. Deutsch, J. Giles, and D. Corrigan, Harness the Power of Big Data: The IBM Big Data Platform, New York, NY: McGraw Hill, 2013.

What are the dimensions of Big Data?

- **Data volume** - measures the units of data storage on various media.
- **Data variety** - refers to the many formats of digital data including photos, email and text documents.
- **Data velocity** - according to Gartner "means both how fast data is being produced and how fast the data must be processed to meet demand."
- **Data variability** - according to SAS means "data flows can be highly inconsistent with periodic peaks".
- **Data complexity** - according to SAS means data is from multiple sources and it is difficult and challenging to link, match, cleanse and transform data across systems.

Based on Gartner, IDC, IBM and SAS

: Does the term big data have utility for managers?

According to SAS, the following are possible uses of Big Data with appropriate analytics.

Analyze millions of SKUs to determine optimal prices that maximize profit and clear inventory.
Recalculate entire risk portfolios in minutes and understand future possibilities to mitigate risk.
Mine customer data for insights that drive new strategies for customer acquisition, retention, campaign optimization and next best offers.

Quickly identify customers who matter the most.

Generate retail coupons at the point of sale based on the customer's current and past purchases, ensuring a higher redemption rate.

Send tailored recommendations to mobile devices at just the right time, while customers are in the right location to take advantage of offers.

Analyze data from social media to detect new market trends and changes in demand.

Use clickstream analysis and data mining to detect fraudulent behavior.

Determine root causes of failures, issues and defects by investigating user sessions, network logs and machine sensors.

More big data use cases:

May 1, 2013 by Sushil Pramanick

http://thebigdatainstitute.wordpress.com/2013/05/01/big-data-use-cases-banking-and-financial-services/?goback=%2Egde_4495391_member_237331147

Big Data use-cases in Banking & Financial Services

1. Fraud Detection:

"One of the large credit card issuing bank has implemented fraud detection system that would disable your card if they see suspicious activity based on your past history with spending patterns and trends. In addition to the transaction records for authorization and approvals, banks and credit card companies are collecting lot more information from location, your life style, spending patterns. Credit card companies manage huge volume of data from individual Social Security number and income, account balances and employment details, and credit history and transaction history. All this put together helps credit card companies to fight fraud in real-time. Big Data architecture provides that scalability to analyze the incoming transaction against individual history and approve/decline the transaction and alert the account owner."

: *Does the term big data have utility for managers?*

2. Customer Segmentation

"In Banking & Financial industry, customer segmentation is a key tool in risk scoring analysis and for sales, promotion and marketing campaigns."

3. Crowdsourcing

"Some of the larger institutions have realized they can use analytics to learn about new lines of business and products, to ask customers what they think, and to get ideas. In a move to expand its utility beyond simply finding better answers to known statistical problems, data science startup Kaggle is now letting its stable of expert data scientists compete to tell companies how they can improve their businesses using machine learning."

4. Sales and Marketing Campaigns

5. Call Center Analysis

"For decades, companies have been analyzing call center data for staffing, agent performance, network management. But with big data age, many new interesting software are being implemented today in attempt to take unstructured voice recordings and analyze them for content and sentiment. Banks are applying text and sentiment analysis to this unstructured data, and looking for patterns and trends. Many banks are integrating this call center data with their transactional data warehouse to reduce customer churn, and drive up-sell, cross-sell, customer monitoring alerts and fraud detection."

The claim: "Big data is an emerging paradigm applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Such datasets are often from various sources (Variety) yet unstructured such as social media, sensors, scientific applications, surveillance, video and image archives, Internet texts and documents, Internet search indexing, medical records, business transactions and web logs; and are of large size (Volume) with fast data

: Does the term big data have utility for managers?

in/out (Velocity). More importantly, big data has to be of high value (Value) and establish trust in it for business decision making (Veracity). Various technologies are being discussed to support the handling of big data such as massively parallel processing databases, scalable storage systems, cloud computing platforms, and MapReduce. Big data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make business more agile, and to answer questions that were previously considered beyond our reach."

Author: Daniel Power

Last update: 2014-04-10 06:00