



Big Data

Making Sense of Structured and Unstructured Data

Author: Edgar Germer, Risk Control Specialist

Published: December 2013

Executive Summary

Are you being bombarded with emails about “big data”, weekly if not daily? Do you hear the term big data being tossed about as if everyone should know what it is? Then surely, you must be asking yourself – What is big data anyway? This whitepaper will help answer that question through a high level overview of this concept addressing:

- What are big data and big data analytics?
- What is meant by structured versus unstructured data?
- What technologies, applications and economic opportunities are presented by big data and big data analytics?

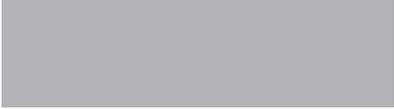
What is Big Data?

No one is quite sure who coined the term “big data” and began using it in the context we see today. The earliest hints are attributed to John Mashey at Silicon Graphics in the early 1990s.¹ In its simplest form, big data is nothing more than a large collection of data residing in storage databases. The key attribute is that the amount of data stored exceeds the organization’s storage and computing capacity, overwhelming the organization and inhibiting the ability to analyze the data for decision-making purposes.

The magnitude or extent of big data is dependent on an individual organization’s computing and storage capacity. Today typical big data magnitudes are in the order of terabytes (trillion or 1,000,000,000,000 or 10^{12} bytes) and petabytes (10^{15}), with exabytes (10^{18}) and zettabytes (10^{21}) not far away. To put it in perspective, 1 terabyte equates to the data from 2,000 hours of CD-quality music while 10 terabytes equates to the data found within the entire U.S. Library of Congress’ print collection. Two petabytes could store all U.S. academic research libraries. For a deeper and somewhat fun perspective on what these magnitudes represent click [here](#).

Industry expert IDC (International Data Corp) describes the growth in data as being a five dimensional phenomenon²:

- **Volume of data:** Greater volumes of data generated from transactions, text, social media, sensors and so on.
- **Variety of data:** Data today comes in many formats – traditional data, text documents, email, sensor and meter-collected data, video, audio, photo, etc. By some estimates, 80 percent of an organization's data is not numeric³.
- **Velocity of data:** How fast data is being produced and how quickly it must be processed to meet demand.
- **Variability of data:** Data generation varies and is inconsistent. Daily, seasonal (Mother’s Day), and event-triggered (Arab Spring and Uprising) peaks in data loads can be challenging to manage – especially when social media forums such as Twitter are involved.



- **Complexity of data:** Large volumes and varieties of data from multiple sources make it challenging to link, match, cleanse and transform data across systems.

Structured vs. Unstructured Data

Today, data is viewed quite differently than years ago. Traditionally, data had to be provided in a “structured” (numerical) format in order to be mined and analyzed. By some estimates, structured data today represents about 10 – 30% of a company’s data. The remaining 70 – 90% of data is considered “unstructured,” meaning freeform text, images, audio and video⁴. Unstructured data comes from websites, correspondence, customer service center records, social media, customer complaints and many other sources. It is contained in document repositories, emails, spreadsheets, audio/visual files, social media sites and texting channels. The availability of unstructured data and the ability to extract meaningful information from it is a significant difference adding to the big data/big analytics phenomenon.

What is Big Data Analytics?

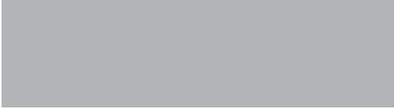
Due to relatively limited computing capacity historically available, organizations have been confined to analyzing subsets of data for decision-making purposes; or they were limited to simplistic analysis because the volume of data overwhelmed their processing platforms. Organizations could be overlooking important trends due to finite processing power, storage capacity or tools to effectively analyze the extent of such data. Today, however, we have specialized software tools and extensive computing processing power to tackle this problem thanks to the emergence of big data analytics.

Big data analytics is the expedient processing of large volumes of data using new hardware and software technologies to extract meaningful information to make better decisions⁵. Correlations that were never possible may be developed if there is a large enough dataset, appropriate analytical tools and sufficient computing power. Hidden patterns emerge allowing for better decision making and assessment in fields such as business, healthcare, epidemiology and so on.

Enabling Technologies

A number of new technological advancements that are enabling organizations to make the most of big data and big data analytics include⁶:

- **Cheap, abundant storage and server processing capacity:** The cost of a gigabyte of storage has dropped from approximately \$16 in 2000 to less than \$0.07 as of November 2011⁷.
- **Faster processors:** Based on Moore’s law, processor speeds continue to double at least every 18 months, if not faster.
- **New technologies:** Storage and processing technologies designed specifically for large data volumes, including unstructured data. Three key big data processing architectures include:
 - **Grid computing:** A centrally managed grid infrastructure provides dynamic workload balancing, high availability and parallel processing for data management, analytics and reporting.
 - **In-database processing:** Reduces the time needed to prepare data and build, deploy and update analytical models.
 - **In-memory processing:** Allows processing of data in memory rather than on a disk making data computing faster and more efficient.
- **Cloud computing:** Allows big data analytics to be delivered as a service through cloud-based storage and high speed connectivity.



- **Text analytics for unstructured data:** Software that identifies, extracts, and interprets relevant data and structures it to reveal patterns, sentiments and relationships within documents⁸.
- **Smart Filters:** Natural Language Processing (NLP) that allows for the processing and interpretation of nonstructured data for additional analytics.

The benefit of this capability is better business decisions through the analysis of whole datasets instead of smaller subsets in a fraction of the time – in minutes or hours compared to days or weeks.

Analytics & Applications of Big Data

Big data and big data analytics are application independent meaning that the data can be generated and the analytics can be used by a variety of applications.⁹ The vision for big data is that organizations will harness more relevant data, apply analytical tools and use the results to make the best decision. According to a survey conducted by industry expert Economist Intelligence Unit in June 2011, there is a strong link between an organization’s effective data management and its financial performance. Eric Brynjolfsson, an economist at the Sloan School of Management at the Massachusetts Institute of Technology, found that companies that adopted data-driven, decision making achieved productivity boosts of 5-6%¹⁰.

Examples of applications of big data analytics include:

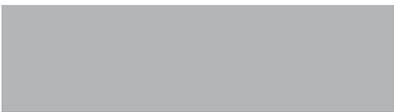
- Analyze millions of SKUs to determine optimal pricing and maximize profits.
- Calculate entire risk portfolios in minutes and understand future possibilities to mitigate risk.
- Mine customer data for insights that drive new strategies for customer acquisition, retention, campaign optimization and development of next-best offers.
- Quickly identify customers who matter the most.
- Send tailored recommendations to customer’s mobile devices at the optimal time - while they are in the right location to take advantage of offers. Analyze data from social media to detect new market trends and changes in demand.
- Determine root causes of failures, issues and defects by investigating user sessions, network logs and machine sensors.
- Government intelligence to track terrorists¹¹. Big data analytics were used to track down the Boston Marathon Bombers¹².
- Insurance companies and financial institutions (MasterCard, Visa) use big data analytics to identify fraud¹³.
- Health industry derives meaning from unstructured data to uncover disease patterns, causes and effects to improve healthcare¹⁴.

Economic Opportunities

Aside from potential increases in profits, efficiency and national safety, big data is giving rise to many economic opportunities such as:

- **New professions:** There is a shortage of individuals with skills to manage data effectively. As a result, universities are working with private industry to address their specific needs¹⁵.
- **Growth in private industry:** New products and service opportunities for software and hardware companies specializing in data management and storage. Startups and existing firms developing new products to capitalize on the growing need for better and faster analytics and data management.

This article is provided for general informational purposes only and does not constitute and is not intended to take the place of legal or risk management advice. Readers should consult their own counsel or other representatives for any such advice. Any and all external websites or sources referred to herein are for informational purposes only and are not affiliated with or endorsed by OneBeacon Insurance Group. OneBeacon Insurance Group hereby disclaims any and all liability arising out of the information contained herein.



- **Growth of cloud service providers:** Increase in data storage needs will fuel the growth of such providers.
- **Growth of revenue:** Generate significant financial value across sectors¹⁶.
 - \$300 billion value per year in U.S. health care
 - \$100 billion+ revenue for service providers in global personal location data
 - 60+% increase in net margin possible for U.S. retail

Conclusion

As the old adage goes “knowledge is power.” Big data and big data analytics are tools that facilitate the acquisition of knowledge so clearly are a powerful resource delivering game-changing insights and competitive advantages. As big data changes the rules of the game for organizations of all sizes it has become an industry unto itself. Companies have no choice but to participate in the opportunity or risk losing market share and falling behind.

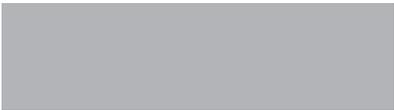
Looking ahead, this is a trend that’s here to stay. The relative affordability and availability of such analytical tools allows organizations of all sizes to significantly improve their decision-making and business plan execution. Big data means big opportunities, and the companies willing to leverage these tools will be positioning themselves to best compete within their industry segments both in the near and long term.

Contact Us

To learn more about how OneBeacon Technology Insurance can help you manage online and other technology risks, please contact Dan Bauman, Vice President of Risk Control for OneBeacon Technology Insurance at dbauman@onebeacontech.com or 262.966.2739.

References

- ¹ Lohr, Steve (February 1, 2013). “The Origins of ‘Big Data’: An Etymological Detective Story.” New York Times. Retrieved October 14, 2013. http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/?_r=0
- ² “What is Big Data?” SAS Website. Retrieved October 2013. <http://www.sas.com/big-data/>
- ³ Ibid 2
- ⁴ Pope, David. “From Big Data to Meaningful Information”. SAS. Retrieved October, 2013. http://www.sas.com/resources/whitepaper/wp_59833.pdf
- ⁵ “Big Data Analytics – Why is it important?” SAS. Retrieved October 2013. <http://www.sas.com/big-data/big-data-analytics.html>
- ⁶ Ibid 2
- ⁷ Ramanathan, Deepak (2012). “Big Data Meets Big Analytics”. SAS. Retrieved October 2013. <http://de.slideshare.net/deepakramanathan/big-data-meets-big-analytics>
- ⁸ Ibid 4
- ⁹ Ibid 2
- ¹⁰ “Big Data Harnessing a game changing asset”. (September 2011). Economist Intelligence Unit. Retrieved October 2013. http://www.sas.com/resources/asset/SAS_BigData_final.pdf
- ¹¹ Felman, Susan and others (June 2012). “Unlocking the Power of Unstructured Data”. IDC Health Insights. Retrieved October 2013. <http://www-01.ibm.com/software/ebusiness/jstart/downloads/unlockingUnstructuredData.pdf>
- ¹² Rathnam, Lavanya (June 7, 2103). “How Big Data was used to find the Boston Bombers”. iCrunchData News. Retrieved October 2013. <http://news.icrunchdata.com/post/2013/06/07/big-data-boston-bomber>



¹³ “Building Believers – how to expand the use of predictive analytics in claims”. SAS. Retrieved October 2013. http://www.sas.com/resources/whitepaper/wp_59831.pdf

¹⁴ Ibid 11

¹⁵ Ibid 10

¹⁶ Manyika, James. (May 2011). “Big data: The next frontier for innovation, competition, and productivity”. McKinsey Global Institute. Page 8. retrieved October 2013. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation