Big Data: The Structure & Value of Big Data Analytics

Indicate Submission Type: Emergent Research Forum papers

Hak J. Kim
Hofstra University
hak.j.kim@hofstra.edu

Abstract

The term Big Data is intuitively appealing and increasingly well accepted in academics as well as practices. Firms readily see the possibility of new business value from big data and future business opportunities. Although they are good understanding what Big Data captures that conventional data do not, the journey for Big Data is difficult and deeply frustrating, as widely known, because of its volume, variety, and velocity. They also get stuck how to collect and analyze Big Data because how-to advice is scarce on this subject and mostly aimed at experts. As a result, Big Data Analytics are considered difficult to implement. The paper discusses that big data have business value and develop a model for measuring its value. We also attempt to design an implementing framework for big data collection as the first step for big data analytics. This paper can contribute to provide a guideline for studying big data analytics.

Keywords

Mobile Cloud Computing, Big Data, Business Insight, Analytical Platform

Introduction

Today's world is full of data from a variety of sources including web, social media, smartphone records, medical records, sensors and surveillance cameras. The enormous volumes of data in different forms are generated every day. Rather, significant values are being extracted from this data deluge with extremely high velocity. According to IDC 2011 Digital Universe Study (EMC, 2011), the amount of data created and replicated will surpass 1.8 zettabytes (1.8 trillion gigabytes), growing by a factor of nine in just five years. And metadata, data about data, is growing twice as fast as the digital universe as a whole.

For decades, the Internet has been explosively growing and generating tremendous economic benefits to our world. As pointed out by Werbach (1997), the Internet's open and flexible architecture provides the endless spiral of connectivity; that is, any form of network could connect to and share data with other networks through the Internet. As known, the client/server platform (Abdul-Fatah, 2002) has been the main architecture of the Internet. However, the client/server platform could not accommodate the deluge of data from diverse devices and services. There are emerged a number of different computing platforms at the same time; whether it is cloud, whether it is mobility, and whether it is social.

Cloud Computing (Kuo, 2011; Mell & Grance, 2011) is gradually taking over the position of client/server platform in the Internet. Cloud computing refers to enable on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort.. People use it conveniently instead of their own PC with expensive applications and storage. Mobile Computing (Bresnahan & Greenstein, 2014) is the use of mobile devices with mobile technologies to enable people to access the Internet from

anywhere, anytime, and any person. It supports communication, application, and services in mobile devices.

The purpose of this paper is to discuss that big data have business value and develop a model for measuring its value. We also attempt to design an implementing framework for big data collection as the first step for big data analytics. The second section describes social media and smartphone as a typical source of big data. And then we attempt to establish an analytical framework of social media data.

Business Insights of Big Data

Social Media and Big Data

Smartphones (i.e., iPhone and Galaxy) with their rich application support are one of the fastest growing fields in mobile industry. People can communicate and interact with anyone, anytime, and anywhere using smartphones. It also creates new service deployment models. Unlike traditional cellular phones as a communication tool, today's smartphones are used for sharing information (i.e., social networking and geographic location services) and enjoying entertainment (i.e., games and sports), which is called '*Infotainment*'. The wide adoption of smartphones has opened new opportunities to business organizations, driving innovation in business. As a consequence of these initiatives, the business firms will experience better productivity and increased efficiency.

Smartphone users are sharing a lot more information, personal information about activities, and what they are doing through social media, such as Twitter and Facebook. With the popularity of smartphones, social media are becoming more prevalent and emerged as a new way of life to the people (Hathi, 2009). IBM (2013) reported more than two billion Internet users, 4.6 billon mobile phones, more than 500 million Facebook users, and about 340 millions of data every day in Twitter. These phenomena create new term which is called '*Big Data*'. Originally it is used for the large volume of data, but expands its scope to variety and velocity. It is not only meaning data itself, but also new paradigm. It drives to develop new methods and technologies for obtaining, processing, and analyzing unprecedented data (i.e., unstructured data). *Big Data* pose grand challenges as opportunities to advance business intelligence; for example, new computing architecture for processing *Big Data*, interoperability issues for sharing and managing *Big Data*, new methods for processing and visualizing *Big Data*.

Enterprise and Big Data

According McKinsey Global Institute report (Manyika et al., 2011), a retailer embracing big data has the potential to increase its operating margin by more than 60 percent. Companies are leveraging these and many other sources of data to achieve a better understanding of their customers, employees, partners and operations, with an eye towards improving every aspect of business. In fact, the Leadership Council for Information Advantage anticipates that data will generate a similar productivity boost to the enterprise that IT has over the past 20 years. Big data has the potential to redefine business, and companies that understand this stand to become leaders in the global marketplace.

Big Data is a symbolic terminology as a new wave of analytic innovation to support business decision-making. Gartner (2013) says worldwide business intelligence, analytics and performance management software market surpassed the $12 billion mark in 2012. It is becoming a new way of life to the people, such as multi real-time access behavior. With this big data analytics, the world is changing and becoming more intelligent and interconnected. These phenomena have become a revolutionary driving force for the development of new digital era through analytic thinking. People are becoming to enjoy intelligent digital life.

In enterprise, *Big Data* is to impact on information systems and data management approaches (i.e., database and applications). For example, the traditional data warehouse does not accommodate big data. So the enterprise should abandon it and introduce new system for big data. Current the increase of data volume in enterprise is challenge, but also big opportunities to capitalize this data for enterprise competitive advantage. For example, the company can gain the benefit of timely, differentiated insight from big data analysis.

These data may provide very little value. The key is to be able to turn this data into information that can help to predict the future and makes decisions that will have a positive impact. Enter the world of big

data. The most successful technology companies today have been using big data to generate value for their organizations. The new mega-rich of Silicon Valley, first at Google and now Facebook, are masters at harnessing the data of the Web- online searches, posts and messages—with Internet Advertising (Lohr, 2011). The idea of using data analytics to make decisions will create new jobs and change the way the world works.

For the spread of the data is not the end in sight. Tens of petabytes of enterprise data volumes further, business and IT leaders in the past to move terabytes of data for competitive advantage are facing a unique opportunity. In order to accommodate and take advantage of big data processes, operations and corporate culture sort of company in a timely manner will gain insight into the distinctive falling by the wayside who does not take the risk.

Developing Analytic Framework

Linking Big Data to Business Value

By mapping the characteristics (i.e., volume, velocity, variety, and variability) of the Big Data opportunities onto the business value (Figure 1), we can obtain a model of the project that combines its characteristics with the structure of business analytics.

- Volume to Granularity: How much of volume can be collected and used for analysis ? More volume, in general, is more valuable.
- Velocity to Timeliness: How much of time can be needed for analysis ? More fast is more valuable.
- Variety to Option Value: How many of types data can be used for analysis ? More data types are more valuable.
- Variability to Reliability: How much reliable of the collected data can be used for analysis ? More reliable is more valuable.

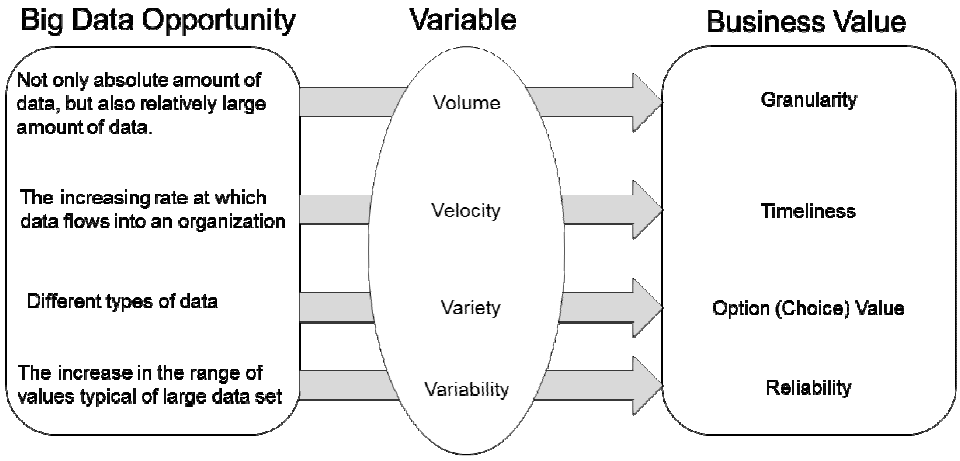

Figure 1. Mapping Big Data to Business Value

Implementing Data Collection for Analytics

Step 1: Understanding the types of data

There are a variety of data types of data from different sources including web, social media, sensor network, and mobile application services. We should understand data types to avoid privacy issues before collecting data. For example, Twitter data is free but there is a lot of it. It is not easy to identify relevant stuff and also is usually not available history. Facebook (public) data is free through search. Its data is available through application. More sophisticated data is available with permissions using some tools.

Step 2: Deciding the level of data

We can see two very different levels of information provided from sources. For example, when we focus on Twitter and Facebook, Twitter provides only basic, low level data, while Facebook provides much more complex, rational data.

Step 3: Selecting the way to data access

Social media data can collect several ways including within the social media itself, across multiple social media, and between social media and enterprise systems. The most secure forms of linking are to use unique references, such as email addresses, IP addresses, and telephone numbers. This can be supported by direct access methods; for example, asking users for their Twitter name or persuading them to 'Like' companies on Facebook from within a known environment.

Step 4: Selecting the approach to data collection

There are several approaches to collect the data. For example, one is the open approach adopted by Twitter. Users on Twitter are warned that their posts are visible to anyone who can find them. The other is the extensive privacy model adopted by Facebook. Data needs to be accessed using authorizations from likes and games.

Step 5: Evaluating the confidence of data

The first step for exploring data is to evaluate the confidence of data sources. For example, the data in social media is user provided and may be wrong in some cases. The notion of a soft match should be adopted; that is, the data is the same person, but we cannot be sure. In case of enterprise data evaluation, we think the value for companies to go beyond surface level analytics of likes, followers and friends, and challenges you to ask deeper and more important questions.

Concluding Remarks

Data analytics are not new. In the past, there was a lot of work to analyze on the flow of society, but that relies on the analysis of specific areas and also was difficult to provide a quantitative basis. But the big data analytics are possible to take advantage of the more objectifying data. For example, we can ask what you think about this summer. People may answer that this summer is very hotter than last year. Instead of that, you may answer that this year is 2.5 times hotter in summer than last year. The answer is more specific and objective supported by the analytical tools. This is *Big Data Analytics*.

Analytics is increasingly delivered to users at the point of action and in context. With the improvement of performance and costs, IT leaders can afford to perform analytics and simulation for every action taken in the business. The mobile client linked to cloud-based analytic engines and big data repositories potentially enables use of optimization and simulation everywhere and every time. This new step provides simulation, prediction, optimization and other analytics, to empower even more decision flexibility at the time and place of every business process action.

The future research will develop a theory of data analytics and build a generalized model for analyzing Big Data (unstructured data) in several industry sector, and then integrated with traditional data (structured data) . This process will contribute for better decision-making.

REFERENCES

Abdul-Fatah, I. 2002. "Performance of CORBA-based client-server architectures," *IEEE Transactions on Parallel and Distributed Systems*, 13 (2), pp. 111-127.

Bresnahan, T. and Greenstein, S. 2014. "Mobile Computing: The Next Platform Rivalry." *American Economic Review*, (104:5), pp475-480.

Butler, J.G. 1996. "Mainframe to Client/Server Migration: Strategic Planning Issues and Techniques," in *Computer Technology Research Corp.*

EMC. 2011. "*2011* IDC *Digital Universe Study," Digital Universe Report.*

Davenport, T. H. 2006. "Competing on Analytics," *HarvardBusiness Review*, vol. 84, no. 1, pp. 98-107.

Fernando, N., Loke, S.W., and Rahayu, W. 2013. "Mobile cloud computing: A survey," *Future Generation Computer Systems* (29:1), pp. 84–106.

Gartner. 2012 "The Nexus of Forces: Social, Mobile, Cloud and Information," *Special Report.*

Hathi, S. 2009. "How Social Networking Increases Collaboration at IBM," *Strategic Communication Management* (14:1), pp. 32-35.

IBM. 2013. "What can you do with big data?," IBM, retrieved from (http://www.ibm.com/big-data/us/en/.

Kuo, C.J. 2011. "Paradigm Shifts in Modern ICT Era and Future Trends," in *Proceedings of the 10th International Symposium on Signals, Circuits and Systems* (ISSCS), Iasi, Romania, pp.1-2.

Lohr, S. 2011. "The age of big data," *The New York Times*. Retrieved from http://wolfweb.unr.edu/homepage/ania/NYTFeb12.pdf

Manyika, J, Chui, M., Brown, B., Bughin, J., Doobs, R, Roxburgh, C., and Byers, A.H. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity," *McKinsey Global Institute report*.

Mell, P. and Grance, T. 2011. "The NIST Definition of Cloud Computing," *National Institute of Standards and Technology*, U.S. Department of Commerce, NIST Special Publication 800-145.

Subramanian, A., and Lacity, M.C. 1997. "Managing Client/Server Implementation: Today's Technology, Yesterday's Lessons," *Journal of Information technology* (12:3), pp.169-186.

Werbach, K. 1997. "Digital Tornado: The Internet and Telecommunications Policy," *Working Paper ,* FCC Office of Plans and Policy, No. 29.