View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarWorks at WMU

Western Michigan University
ScholarWorks at WMU

Transactions of the International Conference on Health Information Technology Advancement

Center for Health Information Technology Advancement

10-2013

The Impact of Big Data on the Healthcare Information Systems

Kuo Lane Chen
University of Southern Mississippi, Kuo.chen@usm.edu

Huei Lee
Eastern Michigan University, huei.lee@emich.edu

Follow this and additional works at: http://scholarworks.wmich.edu/ichita_transactions

Part of the Health Information Technology Commons

WMU ScholarWorks Citation

Chen, Kuo Lane and Lee, Huei, "The Impact of Big Data on the Healthcare Information Systems" (2013). *Transactions of the International Conference on Health Information Technology Advancement*. 25.
http://scholarworks.wmich.edu/ichita_transactions/25

This Article is brought to you for free and open access by the Center for Health Information Technology Advancement at ScholarWorks at WMU. It has been accepted for inclusion in Transactions of the International Conference on Health Information Technology Advancement by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.

The Impact of Big Data on the Healthcare Information Systems

Kuo Lane Chen
School of Computing,
University of Southern Mississippi
kuo.chen@usm.edu

Huei Lee
Department of Computer Information Systems
Eastern Michigan University
Ypsilanti, MI 48187
Huei.Lee@emich.edu

Abstract: This article explores the possible impact of big data on healthcare information systems. Possible research issues include: 1). What applications in healthcare information systems are impacted most? 2). What algorithm/programs will be used for big data? 3). What privacy, security, and ethical issues are there for big data? In the biology area, big data becomes the newest technology for genomics. Other possible areas include pharmacovigilance, patient care, and medical supply chain management.

INTRODUCTION

Due to the explosion of information, big data is one of the hottest topics in data analytics and healthcare information systems today. The U.S. government announced the Big Data Research and Development Initiative for 200 million in 2012. The initiative explores the possibility of using big data or a large-scale database to solve significant problems that the government is facing. The funding will be provided to six departments/agencies which include National Institutes of Health. In a recently published *Harvard Business Review* article, McAfee and Brynjolfsson (2012) listed two scenarios for the application of the big data on predication and sales. The purpose of the paper is to explore the possible impact of big data on healthcare information systems.

WHAT IS BIG DATA?

Big data is similar to business analytical and business intelligence but scale of big data is much larger. Three "V" dimensions are typically used to describe big data: volume, velocity, and variability. Figure 1 illustrates the traditional relationship among healthcare information systems and business intelligences. Several reasons account for the evolution of big data. The first is the rise of in-memory computing. Traditional computers include central processing units (CPU) and secondary storage device (such as a hard drive). Data transfer occurs between the CPU and the secondary storage during data processing. Although it is an economical way to use computers, the processing speed of the computers tends to be slow. SAP Hana and Oracle design new computer systems which use in-memory systems (Monk & Wagner, 2013). Therefore, today's computer systems are more capable of handling large-scale data than before.

The second reason for the rise of big data has been social networking. In traditional data mining and business intelligence, the data are based on the internal data generated from internal enterprise resource planning systems (ERP) or healthcare information systems. The data used in these systems are also referred to as "structured data" which is still limited. Social networking became popular merely six years ago but already generates a large amount of data. Some data are very useful for data analysis. These external data or some of the unused internal data referred to as unstructured data. As shown in Figure 2, big data analysis combines structure data and unstructured data which come from internal unused data, social media and other external data sources.

WHAT CAN BIG DATA DO FOR HEALTHCARE INFORMATION SYSTEMS?

An excellent example of a real-world case for data analysis was provided by Duhigg (2012) when he "revealed that the retail giant Target can figure out whether a girl was pregnant by tracking her consumer behavior (Trank, 2013)."

(Lee, Chen, & Kao, 2013). With these data, Target would be able to "predict her future consumption of infant-related items such as diapers." Hill (2012) explained how Target figured out whether a girl was pregnant using data mining tools. "Since Target hired a statistician Andrew Pole to use data mining tools to predict consumer behavior, the revenue grows from 44 billion in 2002 to 67 billion in 2010." (Lee, Chen, & Kao, 2013).

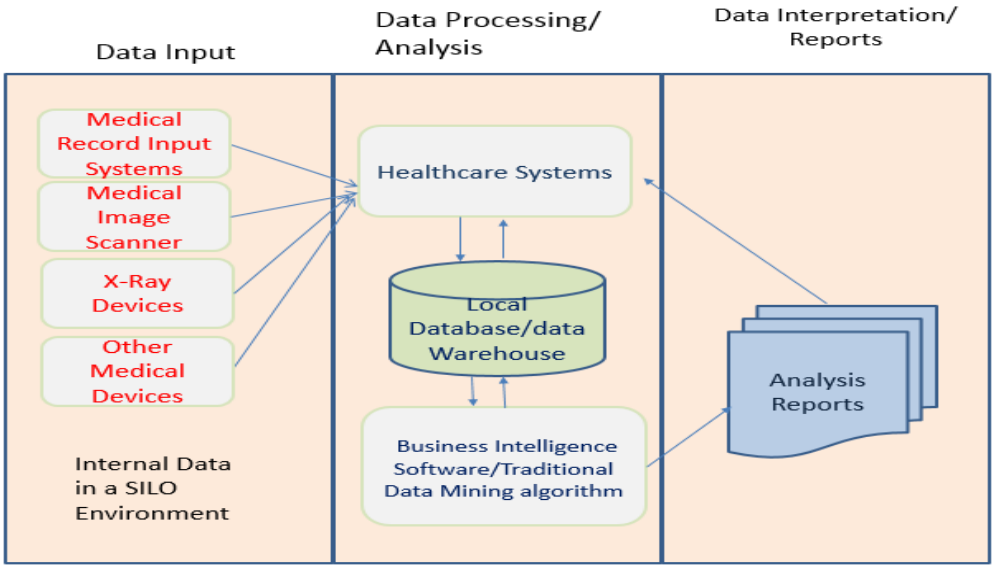

Figure 1. Traditional relationship between healthcare information systems and business intelligence

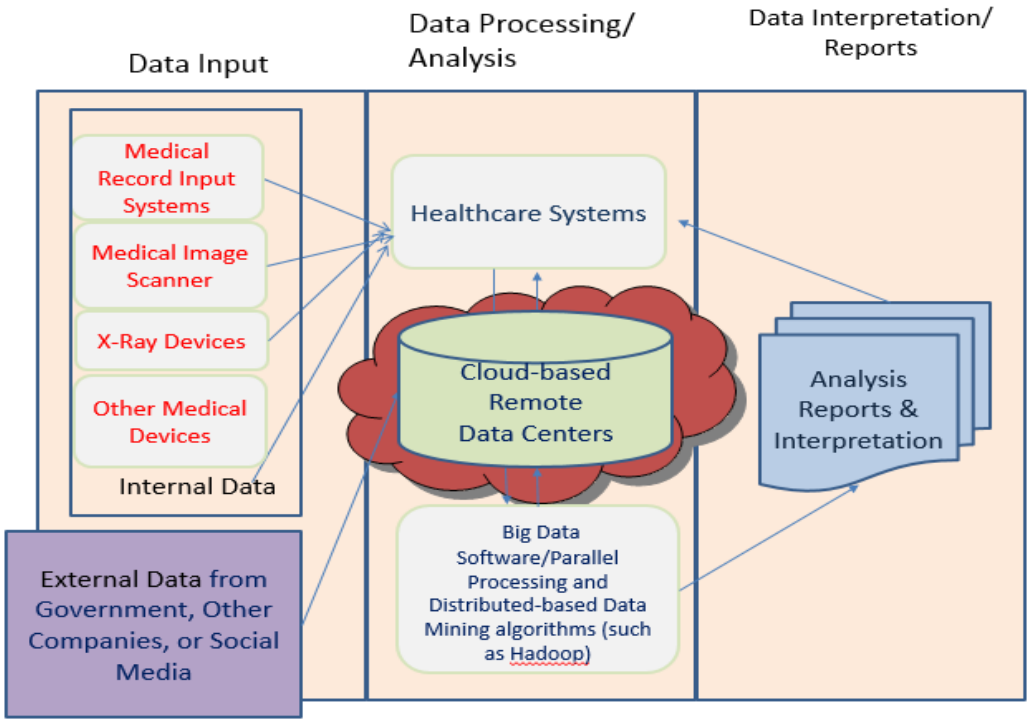

Figure 2. Relationship between healthcare information systems and big data

Clearly, there is a great deal of potential for big data in healthcare information systems. For one, such mainframe-based business intelligence (BI) or data-mining programs as SAS or SAP BW should have the capability to be upgraded to handle the big data analysis. Several ERP and healthcare systems vendors and major IT companies like IBM, Microsoft, SAP, Oracle, and SAS have already worked on various big data projects (Lee, Chen, & Kao, 2013). Research into the specific applications big data is still in the early stage and under development, but several general applications are emerging. In the biology area, big data has become the newest technological tool for genomics. Marx (2013) mentioned that biologists use big data to examine "everything from the regulation of genes and the evolution of genomes to why coastal algae bloom, what microbes dwell where in human body cavities and how the genetic make-up of different cancers influences how cancer patients fare." It took twelve years for the Human Genome Project to collect, analyze, and interpret the huge amount of data needed to produce a map of the about 20,000–25,000 genes, but it may take just one single day to use new big data technologies to achieve the same result (Ross, 2012).

Schultz (2013) indicates big data can be used in the pharmaceutical development cycle in the following areas: 1) Genomics, 2) Clinical monitoring, and 3) Pharmacovigilance. Chawla and Davis (2013) developed a novel system called Collaborative Assessment and Recommendation Engine (CARE) for predicting personalized disease risk. Big data also facilitates the storage and processing of medical imaging data (Hay, 2011).

RESEARCH ISSUES FOR BIG DATA

We have conducted a literature study (Chawla & Davis, 2013; Jee & Kim, 2013) and make the following possible research issues for big data:

1. **What applications (modules) in healthcare information systems will be impacted most by big data?**
   In business works, marketing research, sales forecast, and finance are the most impacted areas. In healthcare, genomics, pharmacovigilance, and patient care are the most important areas.

2. **What algorithm/programs will be used for Big Data?**
   Traditional algorithms include statistics (regression analysis, time series, clustering, and sequential clustering), mathematics (neural network and Naïve Bayes), and other (decision tree and Structured Query Languages --SQL). What will be the algorithms for big data analysis? Traditional SQL is based on relational database. We will see more complex modes such as NoSQL (also referred to as "not only SQL") and Hadoop® for big data analysis (Schultz, 2013). For example, Hadoop includes MapReduce, a software framework for writing applications, which "processes vast amounts of data in parallel on large clusters consisting of thousands of nodes of commodity hardware in a reliable and fault-tolerant manner" (The Apache Software Foundation, 2013).

3. **What the privacy, security, and ethical issues emerge for big data?**
   Several concerns and privacy issues arise from the use of big data (Hill, 2012). For one, because of the Health Insurance Portability and Accountability Act (best known as HIPAA), healthcare providers and institutions are aware of the importance of security and privacy. Ethical issues are another concern. The case of Target's use of information mentioned earlier is an excellent example of how possible ethical issues call for the need to protect confidential patient and client information.

   The growing applications of big data mean the use of more central-controlled data centers and cloud computing. For example, Cook County Health and Hospitals System (CCHHS) in Illinois has installed a unified virtual data center infrastructure to serve the operational needs of more than twenty facilities across the Chicago area (Ritchey, 2012). The European Bioinformatics Institute (EBI) in UK, one of the world's largest biology-data repositories, presently stores 20 petabytes of data and back-ups about genes, proteins and small molecules (Marx, 2013). A security breach for data centers the size and nature of these could trigger both a financial and a non-financial disaster for the institutions, hospitals, and individuals not to mention the mental pain such disclosures could lead to.

ACKNOWLEDGEMENTS

The authors would like to thank one of the anonymous reviewers for his useful comments and suggestions. The authors also would like to thank Dr. John Waltman for his suggestions and corrections.

REFERENCES

The Apache Software Foundation (2013). *MapReduce tutorial*. Retrieved from http://hadoop.apache.org/docs/r1.0.4/mapred_tutorial.html

Chawla, N. V & Davis, D. A. (Sep, 2013). Bringing big data to personalized healthcare: A patient-centered framework. *Journal of General Internal Medicine*, 28(3), 660 – 665.

Duhigg, C. (2012). *The power of habit: Why we do what we do in life and business.* New York: Random House.

Hay, T. (2011). Big data: New advancements changing the game for medical imaging. *VentureWire*, Retrieved from http://ezproxy.emich.edu/login?url=http://search.proquest.com/docview/886015321?accountid=10650

Hill, K. (2012). How Target figured out a teen girl was pregnant before her father did. *Forbes*. Retrieved from http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

Jee, K. & Kim, G. (2013). Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system. *Healthcare Informatics Research*, 19 (2): 79-85.

Lee, H, Chen, K., & Kao, F. (2013). The impact of big data on the evolution of ERP systems, *Proceedings of the Annual Conference of ACME*, Las Vegas.

Marx, V. (June 13, 2013). The big challenges of big data, *Nature*, 498, 255-260.

McAfee, A. & Brynjolfsson, E. (Oct, 2012). Big data: The management revolution, *Harvard Business Review*, 61-68.

Monk, E. & Wagner, B. (2013). *Concepts in Enterprise Resource Planning*, 4th Edition, Independence, KY: Course Technology, Inc.

Office of Science and Technology Policy (2012), Obama administration unveils "big data" initiative: Announces $200 million in new R&D investments. Retrieved from: http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

Ritchey, D. (2012). Big data, big security. *Security*, 49(7), 28, 30.

Rose, S. (2012). Big data and the future. *Significance*, 9: 47–48.

Schultz, T. (2013). Turning healthcare challenges into big data opportunities: A use-case review across the pharmaceutical development lifecycle. *Bulletin of the American Society for Information Science and Technology* (Online), 39(5), 34-40.

Trank, M. (Apr/May 2013). Big Data: Where's the pot of gold for retail store designers? *Display & Design Ideas*, 25(3), 100.