ARTICLE IN PRESS

TFS-18447; No of Pages 11

Technological Forecasting & Social Change xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations

Yichuan Wang [a,*], LeeAnn Kung [b], Terry Anthony Byrd [a]

[a] Raymond J. Harbert College of Business, Auburn University, 405 W. Magnolia Ave., Auburn, AL 36849, USA
[b] Rohrer College of Business, Rowan University, 201 Mullica Hill Road, Glassboro, NJ 08028, USA

ARTICLE INFO

Article history:
Received 17 June 2015
Received in revised form 11 November 2015
Accepted 12 December 2015
Available online xxxx

Keywords:
Big data analytics
Big data analytics architecture
Big data analytics capabilities
Business value of information technology (IT)
Health care

ABSTRACT

To date, health care industry has not fully grasped the potential benefits to be gained from big data analytics. While the constantly growing body of academic research on big data analytics is mostly technology oriented, a better understanding of the strategic implications of big data is urgently needed. To address this lack, this study examines the historical development, architectural design and component functionalities of big data analytics. From content analysis of 26 big data implementation cases in healthcare, we were able to identify five big data analytics capabilities: analytical capability for patterns of care, unstructured data analytical capability, decision support capability, predictive capability, and traceability. We also mapped the benefits driven by big data analytics in terms of information technology (IT) infrastructure, operational, organizational, managerial and strategic areas. In addition, we recommend five strategies for healthcare organizations that are considering to adopt big data analytics technologies. Our findings will help healthcare organizations understand the big data analytics capabilities and potential benefits and support them seeking to formulate more effective data-driven analytics strategies.

1. Introduction

Information technology (IT)-related challenges such as inadequate integration of healthcare systems and poor healthcare information management are seriously hampering efforts to transform IT value to business value in the U.S. healthcare sector (Bodenheimer, 2005; Grantmakers In Health, 2012; Herrick et al., 2010; The Kaiser Family Foundation, 2012). The high volume digital flood of information that is being generated at ever-higher velocities and varieties in healthcare adds complexity to the equation. The consequences are unnecessary increases in medical costs and time for both patients and healthcare service providers. Thus, healthcare organizations are seeking effective IT artifacts that will enable them to consolidate organizational resources to deliver a high quality patient experience, improve organizational performance, and maybe even create new, more effective data-driven business models (Agarwal et al., 2010; Goh et al., 2011; Ker et al., 2014).

One promising breakthrough is the application of big data analytics. Big data analytics that is evolved from business intelligence and decision support systems enable healthcare organizations to analyze an immense volume, variety and velocity of data across a wide range of healthcare networks to support evidence-based decision making and action taking (Watson, 2014; Raghupathi and Raghupathi, 2014). Big data analytics encompasses the various analytical techniques such as descriptive analytics and mining/predictive analytics that are ideal for analyzing a large proportion of text-based health documents and other unstructured clinical data (e.g., physician's written notes and prescriptions and medical imaging) (Groves et al., 2013). New database management systems such as MongoDB, MarkLogic and Apache Cassandra for data integration and retrieval, allow data being transferred between traditional and new operating systems. To store the huge volume and various formats of data, there are Apache HBase and NoSQL systems. These big data analytics tools with sophisticated functionalities facilitate clinical information integration and provide fresh business insights to help healthcare organizations meet patients' needs and future market trends, and thus improve quality of care and financial performance (Jiang et al., 2014; Murdoch and Detsky, 2013; Wang et al., 2015).

A technological understanding of big data analytics has been studied well by computer scientists (see a systemic review of big data research from Wamba et al., 2015). Yet, healthcare organizations continue to struggle to gain the benefits from their investments on big data analytics and some of them are skeptical about its power, although they invest in big data analytics in hope for healthcare transformation (Murdoch and Detsky, 2013; Shah and Pathak, 2014). Evidence shows that only 42% of healthcare organizations surveyed are adopting rigorous analytics approaches to support their decision-making process; only 16% of them have substantial experience using analytics across a broad range of functions (Cortada et al., 2012). This implies that healthcare

* Corresponding author.
E-mail addresses: yzw0037@auburn.edu (Y. Wang), kung@rowan.edu (L. Kung), byrdter@auburn.edu (T.A. Byrd).

http://dx.doi.org/10.1016/j.techfore.2015.12.019
0040-1625/© 2016 Elsevier Inc. All rights reserved.

Please cite this article as: Wang, Y., et al., Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2015.12.019

practitioners still vaguely understand how big data analytics can create value for their organizations (Sharma et al., 2014). As such, there is an urgent need to understand the managerial, economic, and strategic impact of big data analytics and explore its potential benefits driven by big data analytics. This will enable healthcare practitioners to fully seize the power of big data analytics.

To this end, two main goals of this study are: first, to identify big data analytics capabilities; and second, to explore the potential benefits it may bring. By doing so, we hope to give healthcare organization a more current comprehensive understanding of big data analytics and how it helps to transform organizations. In this paper, we begin by providing the historical context and developing big data analytics architecture in healthcare, and then move on to conceptualizing big data analytics capabilities and potential benefits in healthcare. We conducted a content analysis of 26 big data implementation cases in health care which lead to the identification of five major big data analytics capabilities and potential benefits derived from its application. In concluding sections, we present several strategies for being successful with big data analytics in healthcare settings as well as the limitations of this study, and direction of future research.

2. Background

2.1. Big data analytics: past and present

The history of big data analytics is inextricably linked with that of data science. The term "big data" was used for the first time in 1997 by Michael Cox and David Ellsworth in a paper presented at an IEEE conference to explain the visualization of data and the challenges it posed for computer systems (Cox and Ellsworth, 1997). By the end of the 1990s, the rapid IT innovations and technology improvements had enabled generation of large amount of data but little useable information in comparison. Concepts of business intelligence (BI) created to emphasize the importance of collection, integration, analysis, and interpretation of business information and how this set of process can help businesses make more appropriate decisions and obtain a better understanding of market behaviors and trends.

The period of 2001 to 2008 was the evolutionary stage for big data development. Big data was first defined in terms of its volume, velocity, and variety (3Vs), after which it became possible to develop more sophisticated software to fulfill the needs of handling information explosion accordingly. Software and application developments like Extensible Markup Language (XML) Web services, database management systems, and Hadoop added analytics modules and functions to core modules that focused on enhancing usability for end users, and enabled users to process huge amounts of data across and within organizations collaboratively and in real-time. At the same time, healthcare organizations were starting to digitize their medical records and aggregate clinical data in huge electronic databases. This development made the health data storable, usable, searchable, and actionable, and helped healthcare providers practice more effective medicine.

At the beginning of 2009, big data analytics entered the revolutionary stage (Bryant et al., 2008). Not only had big-data computing become a breakthrough innovation for business intelligence, but also researchers were predicting that data management and its techniques were about to shift from structured data into unstructured data, and from a static terminal environment to a ubiquitous cloud-based environment. Big data analytics computing pioneer industries such as banks and e-commerce were beginning to have an impact on improving business processes and workforce effectiveness, reducing enterprise costs and attracting new customers. In regards to healthcare industry, as of 2011, stored health care data had reached 150 exabytes (1 EB = 10^{18} bytes) worldwide, mainly in the form of electronic health records (Institute for Health Technology Transformation, 2013). However, most of the potential value creation is still in its infancy, because predictive modeling and simulation techniques for analyzing healthcare data as a whole have not yet been adequately developed.

More recent trend of big data analytics technology has been towards the use of cloud in conjunction with data. Enterprises have increasingly adopted a "big data in the cloud" solution such as software-as-a-service (SaaS) that offers an attractive alternative with lower cost. According to the Gartner's, 2013 IT trend prediction, taking advantage of cloud computing services for big data analytics systems that support a real-time analytic capability and cost-effective storage will become a preferred IT solution by 2016. The main trend in the healthcare industry is a shift in data type from structure-based to semi-structured based (e.g., home monitoring, telehealth, sensor-based wireless devices) and unstructured data (e.g., transcribed notes, images, and video). The increasing use of sensors and remote monitors is a key factor supporting the rise of home healthcare services, meaning that the amount of data being generated from sensors will continue to grow significantly. This will in turn improve the quality of healthcare services through more accurate analysis and prediction.

2.2. Big data analytics architecture

To reach our goals of this study which are to describe the big data analytics capability profile and its potential benefits, it is necessary to understand its architecture, components and functionalities. The first action taken is to explore best practice of big data analytics architecture in healthcare. We invited four IT experts (two practitioners and two academics) to participate in a five-round evaluation process which included brainstorming and discussions. The resulted big data analytics architecture is rooted in the concept of data life cycle framework that starts with data capture, proceeds via data transformation, and culminates with data consumption. Fig. 1 depicts the proposed best practice big data analytics architecture that is loosely comprised of five major architectural layers: (1) data, (2) data aggregation, (3) analytics, (4) information exploration, and (5) data governance. These logical layers make up the big data analytics components that perform specific functions, and will therefore enable healthcare managers to understand how to transform the healthcare data from various sources into meaningful clinical information through big data implementations.

2.2.1. Data layer

This layer includes all the data sources necessary to provide the insights required to support daily operations and solve business problems. Data is divided into structured data such as traditional electronic healthcare records (EHRs), semi-structured data such as the logs of health monitoring devices, and unstructured data such as clinical images. These clinical data are collected from various internal or external locations, and will be stored immediately into appropriate databases, depending on the content format.

2.2.2. Data aggregation layer

This layer is responsible for handling data from the various data sources. In this layer, data will be intelligently digested by performing three steps: data acquisition, transformation, and storage. The primary goal of data acquisition is to read data provided from various communication channels, frequencies, sizes, and formats. This step is often a major obstacle in the early stages of implementing big data analytics, because these incoming data characteristics might vary considerably. Here, the cost may well exceed the budget available for establishing new data warehouses, and extending their capacity to avoid workload bottlenecks. During the transformation step, the transformation engine must be capable of moving, cleaning, splitting, translating, merging, sorting, and validating data. For example, structured data such as that typically contained in an eclectic medical record might be extracted from healthcare information systems and subsequently converted into a specific standard data format, sorted by the specified criterion (e.g., patient name, location, or medical history), and then the record

Please cite this article as: Wang, Y., et al., Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2015.12.019