Journal of Accounting Education 38 (2017) 9–22

Contents lists available at ScienceDirect

Journal of Accounting Education

journal homepage: www.elsevier.com/locate/jaccedu

Mystery, Inc.: A Big Data case

Kathryn Enget [a], Gabriel D. Saucedo [b,*], Nicole S. Wright [c]

[a] Department of Accounting & Law, University at Albany – SUNY, 1400 Washington Avenue, Albany, NY 12222, United States
[b] Department of Accounting, Albers School of Business and Economics, Seattle University, 901 12th Avenue, Seattle, WA 98122, United States
[c] D'Amore-McKim School of Business, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, United States

ARTICLE INFO

Article history:
Received 29 June 2016
Received in revised form 17 December 2016
Accepted 17 December 2016
Available online 2 February 2017

Keywords:
Big Data
Data analytics
AU 316 (SAS 99)
Journal entry testwork

ABSTRACT

This case introduces you to the concepts of Big Data and data analytics in conjunction with AU 316, *Consideration of Fraud in a Financial Statement Audit*, and journal entry testwork. You will assume the role of a staff auditor and assist the engagement senior with journal entry testwork and data analyses for Mystery, Inc.'s 20x3 fiscal year. In the first part of the case, you will perform journal entry planning and scoping procedures, which will then be passed on to the engagement data analytics team. In the second part of the case, you will analyze the Big Data output prepared by the data analytics team to identify anomalies and to propose follow-up audit fieldwork. To prepare you for the case and to build your understanding of Big Data and data analytics, the engagement partner has also provided you a few pages of background reading on the subject matter. By completing the Mystery, Inc. case, you will gain an understanding of Big Data and AU 316 within the context of a financial statement audit, and supplement your analytical and higher order thinking skillset.

1. Background and introduction

1.1. Big Data

> Big Data is the new oil. The companies, governments, and organizations that are able to mine this resource will have an enormous advantage over those that don't.
>
> [Deloitte, 2012, pg. i]

Big Data consists of datasets so voluminous they cannot be reasonably analyzed using traditional database management systems or software programs. Furthermore, Big Data consists of structured and unstructured data (about 90% is unstructured) and includes soft information such as email messages, social media postings (e.g., blogs, tweets, Facebook entries), phone calls, website traffic, and video streams (Syed, Gillela, & Venugopal, 2013). Examples of Big Data include items such as: more than 1 million customer transactions every hour at Walmart and more than 50 billion photos on Facebook (Cao, Chychyla, & Stewart, 2015; Cukier, 2010). What is considered Big Data differs across various domains, and whether particular data are big or not is determined by whether these data push the capability limits of the information systems that work with these data (Vasarhelyi, Kogan, & Tuttle, 2015). As difficult as it is to truly define Big Data, there are four specific features of Big Data that challenge the capabilities of modern information systems (IBM, 2012; Laney, 2001; Zhang, Yang, & Appelbaum, 2015). These features, known as the 4 Vs, include:

* Corresponding author.
E-mail addresses: kenget@albany.edu (K. Enget), saucedog@seattleu.edu (G.D. Saucedo), n.wright@northeastern.edu (N.S. Wright).

http://dx.doi.org/10.1016/j.jaccedu.2016.12.003
0748-5751/

1. huge *Volume* – the massive size of a typical database;
2. high *Velocity* – data added on a continuous basis;
3. huge *Variety* – types of data, both structured and unstructured, and;
4. uncertain *Veracity* – reliability, authenticity, and validity of data.

Big Data now permeates almost all aspects of major companies' decision making and business strategies. In addition, the data ecosystem is exponentially expanding, and this environment presents a dynamically changing set of characteristics. For example, organizations collected more data during the past two years than in the previous 2,000 years combined (Syed et al., 2013). As corporations continue to expand the scope of their information systems from traditional data processes to automated data capture, it becomes more and more evident that those organizations which devote the right resources to analyzing relevant data will benefit from a more informed perspective that will help them make better business decisions and, ultimately, achieve the measured business outcomes they desire (KPMG, 2012). Simply stated, companies that effectively and efficiently utilize Big Data have the potential to gain significant competitive advantages including cost avoidance, increased profits, clear(er) thinking, and new product/service development (Dennehy, 2016).

While more and more companies are harnessing the power of Big Data, there are still challenges encountered when handling such information including data consistency, integrity, identification, aggregation, and confidentiality (Zhang et al., 2015). Data consistency relates to the interdependence of data between applications and across an organization (Sheth & Rusinkiewicz, 1990). As there are an increasing number of different data sources in today's era of Big Data, the frequency of data conflicts has inevitably increased (Zhang et al., 2015). Data integrity relates to the ability to identify whether individual data or larger sets of data have been modified, deleted, hidden, or destroyed because of intentional acts or unintentional error (Menezes, Oorschot, & Vanstone, 1996; Motro, 1989). Data identification refers to records that link two or more separately recorded pieces of information about the same piece of data. Data aggregation is the operation of summarizing and simplifying Big Data that is likely coming from different sources. Data confidentiality infers that certain data are sensitive and cannot be released to others (Ciriani et al., 2009).

Brown-Liburd, Issa, and Lombardi (2015) also warn users of Big Data, specifically auditors, to be mindful of information processing biases and the limitation of human cognition when facing excessive information from Big Data. These breakdowns in information processing heuristics may result in information overload, inability to identify relevant information, and difficulty in recognizing patterns in financial, nonfinancial, and ambiguous data. Effective judgment and decision-making is one of the most critical tools available to the modern accounting professional (Fay & Montague, 2015). Consequently, low-quality judgments can have significant ramifications for an accounting firm, such as greater liability exposure and a heightened risk of financial losses. For individual accountants, low-quality judgments can have negative performance evaluation, compensation, job retention, and promotion effects (Bonner, 2008).

Big Data will specifically play an important role in auditing because it complements traditional evidence with sufficient, reliable, and relevant information (Yoon, Hoogduin, & Zhang, 2015). Littley (2012) comments that Big Data has the potential to improve forecasts of estimates, going concern, fraud, and other such variables that are of concern to internal and external auditors. Given the criticism the audit profession has received for their perceived failings ranging from Enron and WorldCom to the financial crisis (Association of Chartered Certified Accountants (ACCA), 2011; Institute of Chartered Accountants in England & Wales (ICAEW), 2010), auditors may eagerly embrace Big Data as a way of increasing the effectiveness and credibility of their work (Alles, 2015). External auditors may also see Big Data as a way to reduce the costs of their audits and enhance profitability, while internal auditors may see Big Data through the lens of cost effectiveness (Littley, 2012).

1.2. Data analytics

Data analytics is the specific process of examining raw data with the purpose of drawing conclusions about that information (i.e., conclusions about Big Data) (Deloitte, 2016; Ernst & Young (EY), 2015; KPMG, 2012; PricewaterhouseCoopers (PwC), 2015). Data analytics is used in many businesses to make better business decisions and in the sciences to verify or disprove existing models or theories. Data analytics is distinguished from data mining by the scope, purpose, and focus of the analysis. Data miners sort through huge datasets using sophisticated software to identify undiscovered patterns and establish hidden relationships, whereas data analytics focuses on inference, the process of deriving a conclusion based solely on what is already known by the researcher (Deloitte, 2016; EY, 2015; KPMG, 2012; PwC, 2015). More specifically, data analytics is a complete cycle of understanding, normalizing,[1] visualizing, interpreting, and communicating data.

New graduates in accounting require new perspectives and training in data analytics to handle what is becoming an important discipline permeating the profession (Coyne, Coyne, & Walker, 2016). A recent white paper, *Reimagining Auditing in a Wired World*, produced by the Emerging Assurance Technologies Task Force of the American Institute of Certified Public Accountants (AICPA), points out that if auditing were to be invented today, the processes would be designed to make best use of the available technology and information so that auditors could provide the most effective and efficient services possible. Instead, auditors today use legacy processes and procedures (e.g., data sampling) that are not much different from 50 years

[1] Normalizing the data involves sorting and organizing the data files so that they are in a format useful for analysis.