Business Analytics in the Context of Big Data: A Roadmap for Research

By: Gloria Phillips-Wren, Lakshmi S. Iyer, Uday Kulkarni, and Thilini Ariyachandra.

Phillips-Wren, G., Iyer, L.S. Kulkarni, U., & Ariyachandra, T. (2015). "Business Analytics in the Context of Big Data: A Roadmap for Research," *Communications of the AIS, Vol. 37, #23.*

Abstract:

This paper builds on academic and industry discussions from the 2012 and 2013 pre-ICIS events: BI Congress III and the Special Interest Group on Decision Support Systems (SIGDSS) workshop, respectively. Recognizing the potential of "big data" to offer new insights for decision making and innovation, panelists at the two events discussed how organizations can use and manage big data for competitive advantage. In addition, expert panelists helped to identify research gaps. While emerging research in the academic community identifies some of the issues in acquiring, analyzing, and using big data, many of the new developments are occurring in the practitioner community. We bridge the gap between academic and practitioner research by presenting a big data analytics framework that depicts a process view of the components needed for big data analytics in organizations. Using practitioner interviews and literature from both academia and practice, we identify the current state of big data research guided by the framework and propose potential areas for future research to increase the relevance of academic research to practice.

Keywords: Business Intelligence | Business Analytics | Big Data | Decision Support | Data Governance | Unstructured Data | Framework | Data Scientist

Article:

1. Introduction

Business intelligence (BI), decision support, and analytics are core to making business decisions in many organizations. Recently, traditional approaches to using organizational data have been questioned as companies embrace voluminous, high-velocity data in a variety of formats (i.e., multi-structured) that is generally framed as "big data" (Barton & Court, 2012). Increased competitiveness and productivity in industry has provided the groundwork for big data analytics and its technologies. Interest in big data research is growing exponentially as evidenced by the increase in the number of papers, tracks, and mini- tracks focused on analytics and big data in leading IS conferences.

The Association for Information Systems (AIS) Special Interest Group in Decision Support and Analytics (SIGDSA, formerly SIGDSS) and the Teradata University Network have organized pre-International Conference on Information Systems (ICIS) events since 2009 on data analytics (e.g., BI Congresses I, II, III (2009, 2010, 2012) and SIGDSS 2013 pre-ICIS workshops) to promote theoretical, design science, behavioral research and innovative applications in emerging areas of BI, analytics, decision support, and knowledge management. The increasing level of practitioner involvement and sponsorships associated with these events indicates the interest of the larger community in opportunities associated with big data analytics. The events addressed questions such as how can organizations innovate through big data and how can academic research further innovative thinking in this area.

The plan to develop a framework focused on identifying research opportunities in big data stemmed from the 2012 BI Congress and 2013 pre-ICIS SIGDSA events just as academic research directly focused on big data was beginning to emerge in early 2012. Since advancements in big data were being led by practitioners, these two events aimed to foster "active collaboration between academia and industry to advance the teaching and use of business intelligence and analytics" (Wixom, et al., 2011, 2014, p. 4). The events' themes were innovation through big data and decision support from BI and social media. In light of these goals, academic and industry experts were invited to both events to address the following topics:

a) Industry views of big data: TED-like talks by industry experts on big data to expose academics to thought leadership from leading analytics organizations to inspire academic research efforts in big data
b) Developing the next generation big data workforce
c) Reshaping customer relationship in the context of BI and social media, and
d) Explaining why traditional analytics is not enough to capitalize on big data opportunities.

Industry experts at both events represented the following organizations: AT&T Bell Labs, Credito Emiliano, Deloitte, IBM, International Data Corporation (IDC), SAP, SAS, and Teradata. Participants at both events included IS academics and industry members.

Frameworks for specific types of information systems (IS) are useful to conceptualize their primary components, relationships between components, and processes. For example, Sprague (1980) presented an early framework for a decision support system (DSS) that shows an underlying structure of a database, model base, and user-system interface that influenced research and instruction. With the DSS's evolution into the BI field, Watson (2009) published a framework implementing a data warehouse and data marts as central components. As BI evolved to deal with big data, Eckerson (2011) presented a "new BI architecture" to describe the integration of platforms to handle structured data in traditional data warehouses with emerging data sources. While Eckerson provides a technology view of BI including big data assets, we extend these frameworks to emphasize the analytical processes enabled by big data, the human resources necessary to use it, and the governance processes necessary to manage it.

Drawing on the presentations at the events detailed above, interviews with industry experts, prior work by Watson (2009) on the evolution of the traditional BI environment, and Eckerson's (2011) technical BI architecture, we first developed an integrative big data analytics framework. We consulted industry experts to whet the framework's initial developments and incorporated their viewpoints to arrive at the final one (see Figure 1). The proposed framework captures the analytical process of BI in the context of big data and helps guide our second objective (i.e., to create a roadmap for relevant big data research).

This paper proceeds as follows: In Section 2, we discuss our study's background. In Section 3, we describe the methodology we adopted for conducting the research. In Section 4, we then present our big data analytics framework. In Section 5, we map BI and big data research evident in representative academic journals, practitioner publications, and practitioner interviews to our framework. Based on this mapping, in Section 6, we identify potential research questions that can advance our understanding. Finally, in Section 7, we summarize the state-of-the-art and the research opportunities. We hope this paper will aid researchers in identifying and exploring fruitful big data research ideas and will increase the relevance of academic research to practice.

2. Background

Davenport and Harris (2007) described how some companies gained a sustainable competitive advantage through analytics. For example, Progressive Insurance predicts risk associated with granular cells of customer segments; Harrah's predicts which customer's business is waning and which campaigns will revive it; Marriott dynamically computes the optimal room price; Wal-Mart and Amazon simulate supply chain flows and reduce inventory and stock-outs; UPS predicts which customers will defect to a competitor. During the last decade, these and many other companies across a spectrum of industries have systematically constructed complex models to make data-driven business decisions in their strategic processes to gain stronger competitive positions in their industries. McAfee and Brynjolfsson (2012) provide evidence that data-driven companies perform significantly better on both financial and operational measures. As the use of analytics has become more and more mainstream, the competition based on analytics has intensified.

Big data adds new dimensions to analytics. It offers enhanced opportunities for insight but also requires new human and technical resources due to its unique characteristics. Although practitioners sometimes describe big data as data that are beyond the capabilities of the organization to store or analyze for accurate and timely decision making (Kulkarni, 2013), the term has been characterized in the literature as having one or more of four dimensions: volume, velocity, variety, and veracity (Laney, 2001; IBM, 2014; Goes, 2014). Volume indicates the huge and growing amount of data being generated, with more data often at higher granularity. Velocity indicates the speed at which data are being generated from digital sources such as sensors and electronic communication, which offers the potential for real-time analysis and agility. Variety refers to the variation in types of data from internal and external sources. Veracity is a measure of accuracy, fidelity, or truthfulness of data to guard against the biases, noise, and abnormalities associated with big data. Although other Vs have been suggested, including value, visualization, and volatility, we address the four generally accepted characteristics by discussing a framework for big data analytics (McAfee & Brynjolfsson, 2012; Goes, 2014).

Traditionally, online retailers have tracked what customers bought, what others like them also bought, and, based on analyzing similarities between customer purchase behaviors, offered the most-likely-to-be- bought products to a browsing customer. Big data presents a potentially transformational opportunity (Gillon, Aral, & Lin, 2014). Beyond transactional data, online businesses can know what customers browsed and how long they stayed, along with their exact click-stream and location. They can track reactions to suggestions, responses to dynamically generated promotions, contributions to and influences from reviews. In addition, they can access masses of external data from social network interactions, and blog sites where rich sentiments are expressed. This explosion of data and its analysis has not just changed the answers to the question: what will this customer buy next? It has changed the questions themselves to: what is the potential value of this customer? How influential is this person? How should we communicate with them, and which channel should we use to build a long-term relationship with them? How can we engage with this person through products and services that the customer himself has not yet thought of?

Organizational interest in big data is spurred by opportunities to use these new data sources to make faster and better decisions through sophisticated analytics. The literature provides evidence of significant improvements from using big data for better customer knowledge, customized and personalized outreach to customers, and economic benefit (Davenport & Harris, 2007; Davenport, Harris, & Morison, 2010; McAfee & Brynjolfsson, 2012; Davenport, 2013; Thaler & Tucker, 2013; Roski, Bo-Linn, & Andrews, 2014). Estimates by the McKinsey Global Institute (Manikya et al., 2011) indicate that many government and industrial sectors in Europe and the US could benefit substantially from big data analytics: US healthcare could realize an efficiency and quality value of $300 billion, US retailers could increase their operating margin by up to 60 percent, European governments could save more than €100 billion in operational efficiency, and the services sector using personal location data could recover $600 billion in consumer surplus with the use of big data analytics.

In the future, it is not just the nature of questions that can be answered with big data that will change but also business models, the nature of expertise, the value of experience, business processes, and the decisions we make (Holsapple, Lee-Post, & Patkath, 2014). Thus, businesses find themselves in a situation where opportunity from big data exists but analytical talent and, to some extent, technology is lagging. What is also lagging is the business acumen to understand what questions can be answered and what problems can be solved by analysis of big data that will make business sense now and in future. A big data analytics framework can assist the academic community in identifying research opportunities relevant to practice. With this paper, we take a step in that direction.

3. Methodology

3.1 Practitioner Interviews

The BI Congress and workshops demonstrated practitioner interest in partnering with the academic community around big data concepts. Beginning with sponsors of those workshops and later expanding to a broader community of big data practitioners from university advisory boards and research contacts, we conducted semi-structured interviews to arrive at a generalized big data framework in organizations and to identify research gaps. We continued meeting with practitioners throughout the research project using both structured written interviews and verbal semi-structured interviews.

Table 1 describes the practitioners we interviewed in terms of company/industry and responsibility level. Organizations that provided interviews and agreed to be named are recognized in the acknowledgements, although some organizations requested anonymity. Interviewees are directly responsible for big data analytics either as implementers in their own organization or as consultants advising another industry. We conducted interviews in three rounds: (1) as we were developing and modifying the framework, we asked practitioners to critique it; (2) we circulated the final framework to practitioners for general consensus; and (3) we conducted interviews to augment practitioner literature on research gaps in big data to identify current thoughts in the field. We conducted these interviews to determine if emerging academic research is relevant and aligned with industry best practice and to locate areas in big data analytics that need further exploration useful to both academics and practitioners.

Table 1. Organization and Positional Levels of Interviewees

Organization	Interviewee's position
Booz Allen Hamilton	Principal—data science team leader
Financial Services Firm (anonymous)	Associate director and group manager, decision science and big data team
IBM	Data science team leader; team member (2 interviews)
MicroStrategy	Program manager
NetApp	Program manager
Northrup Grumman	CIO-level
Price Waterhouse Coopers	CIO-level; data science team leader (2 interviews)
SAS	Data science team leader; data science team member (2 interviews)
Large search engine firm (anonymous)	Data science team member
Sogeti	Division-level manager
Teradata	Marketing director
UnderArmour	CIO-level
Apollo Education Group	VP, enterprise performance management

3.2 Survey of Published Academic Literature

Our study's academic portion is based on a survey of representative published academic literature on big data analytics during 2011-2014 in the Senior Scholars' basket of journals (AIS, 2011). We included two additional academic journals (*Communications of the AIS* and *Decision Support Systems*) because they publish related research. We used these journals to provide a representative sample to identify research gaps, not to undertake an exhaustive study of the literature. Beyond business research, we also referred to the large body of methodological research related to big data from computer science and engineering fields. To identify research papers, we used a keyword search in each journal for the terms: big data, social media, analytics, business intelligence, distributed computing, Hadoop, analytics discovery, and data scientist. Tables 2 and 3 list the journals and the papers that we reviewed, grouped by the research method they employed.

Table 2. Academic Journals Reviewed for Big Data Papers

Source	Number of papers (2011-2014)
Communications of the AIS	10
Decision Support Systems	9
European Journal of Information Systems	2
Information Systems Journal	2
Information Systems Research	5
Journal of AIS	3
Journal of Information Technology	1
Journal of MIS	1
Journal of Strategic Information Systems	1
MIS Quarterly	6

Table 3. Academic Research Papers and Methods Used in Reviewed Journals from 2011-2014*

Method	Papers
Overview	Wang & Zhang (2012), Basole, Seuss, & Rouse (2013), Lycett (2013), Aral, Dellarocas, & Godes (2013), Sundararajan, Provost, Oestreicher-Singer, & Aral (2013), Lacity, Solomon, Yan, & Willcocks (2011), Chen, Chiang, & Storey (2012), Shmueli & Koppius (2011), Hosack, Hall, Paradice, & Courtney (2012), Watson (2014), Wixom et al. (2011, 2014)
Theory	Wang & Zhang (2012), Chen & Sharma (2013), Hu, Poston, & Kettinger (2011), Elbashir, Collier, Sutton, Davern, & Leech (2013), Deng & Chi (2012), Oh, Agrawal, & Rao (2013), Li, Hsieh, & Rai (2013), Chae (2014), Kane, Alavi, Labianca, & Borgatti (2014)
Case study	Deng & Chi (2012), Oh et al. (2013), McGrath & Elbanna (2012), Chau & Xu (2012), Pan, Pan, & Leidner (2012), Koch, Leidner, & Gonzalez (2013), vom Brocke, Debortoli, Müller, & Reuter (2014)
Data analysis	da Silva, Hruschka, & Hruschka (2014), Lau, Li, & Liao (2014), Basole et al. (2013), Deng & Chi (2012), van Valkenhoef, Tervonen, Zwinkels, de Brock, & Hillege (2013), Li et al. (2013), Wakefield (2013), Ferguson, Soekijad, Huysman, & Vaast (2013)
Survey	Chen & Sharma (2013), Elbashir et al. (2013), Hu et al. (2011), Wixom et al. (2011, 2014)
Method/Algorithm	Lau et al. (2014), Deng & Chi (2012), Chau & Xu (2012), Evangelopoulos, Zhang, & Prybutok (2012), Zheng, Fader, & Padmanabhan (2012)
Framework	Wang & Zhang (2012), Lacity et al. (2011), Chau & Xu (2012), Chang, Kauffman, & Kwon (2014), Demirkan & Delen (2013), Kane et al. (2014), Holsapple et al. (2014)
Panel/Editorial	Androile (2012), Gillon et al. (2014), Lycett (2013), Beath, Berente, Gallivan, & Lyytinen (2013), Agarwal & Dhar (2014), Goes (2014)
Research directions	Wang & Zhang (2012), Aral et al. (2013), Sundararajan et al. (2013), Chang et al. (2014), Demirkan & Delen (2013), Gillon et al. (2014)

*Note: a publication may be classified under multiple headings

3.3 Survey of Practitioner Literature

Similar to the academic literature survey, we reviewed a representative sample of practitioner literature. We surveyed two sources each in the broad categories of information technology research and advisory firms (Gartner and TDWI), comprehensive online information technology resources (BeyeNETWORK and Information Management), and management consulting organizations (Booz Allen Hamilton and McKinsey & Company). In addition, although we recognize the valuable research contributions of many other companies, to maintain independence and neutrality, we avoided vendors of information technology products and services. Table 4 lists the sources that we reviewed along with a brief description. As we previously indicate, practitioner interviews expanded and informed our understanding of concepts and ideas that have not made their way into the published literature, particularly in regards to the big data framework and future research needs.

Table 4. Practitioner Literature Reviewed

Resource	Description
BeyeNETWORK	Online resources that provide news, expert opinion, on-demand business intelligence education, and resources for business intelligence and data warehousing. Founded in 2005.
Booz Allen Hamilton	Publically traded management consulting firm founded in 1914. Published *The Field Guide to Data Science*.
Gartner	Publically traded, independent information technology research and advisory firm founded in 1979 and targeted at senior information technology leaders. Known for Hype Cycles and Magic Quadrants to visualize technology market in different sectors.
Information Management	Online resources for news, original reporting, online radio programming, Web seminar programming, white papers, education, commentary and feature content serving the information technology and business community. Debuted in 1997 and re-launched in 2009.
McKinsey & Company	Global management consulting firm founded in 1926 that conducts qualitative and quantitative analysis to provide strategic advice to corporations and other organizations. Private corporation with shares owned exclusively by McKinsey employees.
The Data Warehousing Institute (TDWI)	A premier provider of in-depth education and research in business intelligence, data warehousing, and analytics for executives and information technology professionals. Founded in 1995.

4. Big Data Analytics Framework

4.1 Frameworks in the Literature

Frameworks play an important role in helping an organization effectively plan and allocate resources for information systems tasks (Gorry & Scott Morton, 1971). They can help an organization identify components and relationships between parts to understand an otherwise complex system (Sprague, 1980). The frameworks for management information systems (Gorry & Scott Morton, 1971) and for decision support systems (Sprague 1980) are early major frameworks that guided organizations in implementing systems to support decision making. They have also assisted academics in mapping research trends and identifying gaps in research.

As information systems have evolved, numerous frameworks have emerged to inform practice and to provide research insights to academics. For instance, the Zachman framework (Zachman, 1987; Sowa & Zachman, 1992) provides a means of understanding the integration of all components of a system independent of its variety, size, and complexity. In the decision support area, the executive information systems (EIS) development framework (Watson, Rainer, & Koh, 1991) presents a structural perspective of EIS elements, their interaction, and the EIS development process. Since the seminal work of Sprague's DSS framework (1980), the decision support arena has grown and matured (Hosack et al., 2012) to include platforms for executive information systems, group decision support systems, geographic information systems, and, more recently, for business intelligence and big data.

Along with the evolution of DSS, new frameworks to understand the various categorizations of decision support have emerged. The business intelligence framework presented by Watson (2009) describes the components and relationships that may assist in a traditional business intelligence implementation with a data warehouse and one or more data marts at the center of its decision support architecture. However, the changing landscape of BI has brought about the need for alternate platforms for dealing with big data and integrating certain processes that are missing in the traditional BI context. Eckerson (2011) presents a "new BI architecture" describing the various platforms that might be integrated and used to handle traditional structured data sources and a wide variety of new data sources that include big data. Our framework extends these frameworks to emphasize the analytical processes enabled by big data.

4.2 Big Data Analytics Framework

Figure 1 shows our proposed framework for big data analytics. A process view is shown across the top of the diagram, initiated with data sources and proceeding through data preparation, data storage, analysis, and data access and usage. The left hand side shows possible types of data sources. The center section proposes a unified data exchange (UDE) with the components for big data analytics. UDE spans multiple processes including data preparation, storage, and analysis, which tend to overlap in the big data environment. This is distinct from traditional BI, which focused on bringing data from all sources into an integrated or enterprise data warehouse (EDW) and then making it available for analysis. Big data environment requires specialized technical platforms and software integrated into a comprehensive process to support complex BI needs. We note that several organizations and consulting firms are experimenting with alternative concepts such as the UDE, which we discuss in Section 5.

The right hand side of the framework shows the user groups with a range of skills needed to analyze and use big data. At the bottom of the diagram are big data management and governance processes. We use the framework to organize the remainder of this paper and address each component.

Figure 1 shows our proposed framework for big data analytics. A process view is shown across the top of the diagram, initiated with data sources and proceeding through data preparation, data storage, analysis, and data access and usage. The left hand side shows possible types of data sources. The center section proposes a unified data exchange (UDE) with the components for big data analytics. UDE spans multiple processes including data preparation, storage, and analysis, which tend to overlap in the big data environment. This is distinct from traditional BI, which focused on bringing data from all sources into an integrated or enterprise data warehouse (EDW) and then making it available for analysis. Big data environment requires specialized technical platforms and software integrated into a comprehensive process to support complex BI needs. We note that several organizations and consulting firms are experimenting with alternative concepts such as the UDE, which we discuss in Section 5.

The right hand side of the framework shows the user groups with a range of skills needed to analyze and use big data. At the bottom of the diagram are big data management and governance processes. We use the framework to organize the remainder of this paper and address each component.

4.3 Data Sources

Big data are characterized by variety in types of data that can be processed for analysis. "Data sources" in Figure 1 indicate the types of data available to organizations

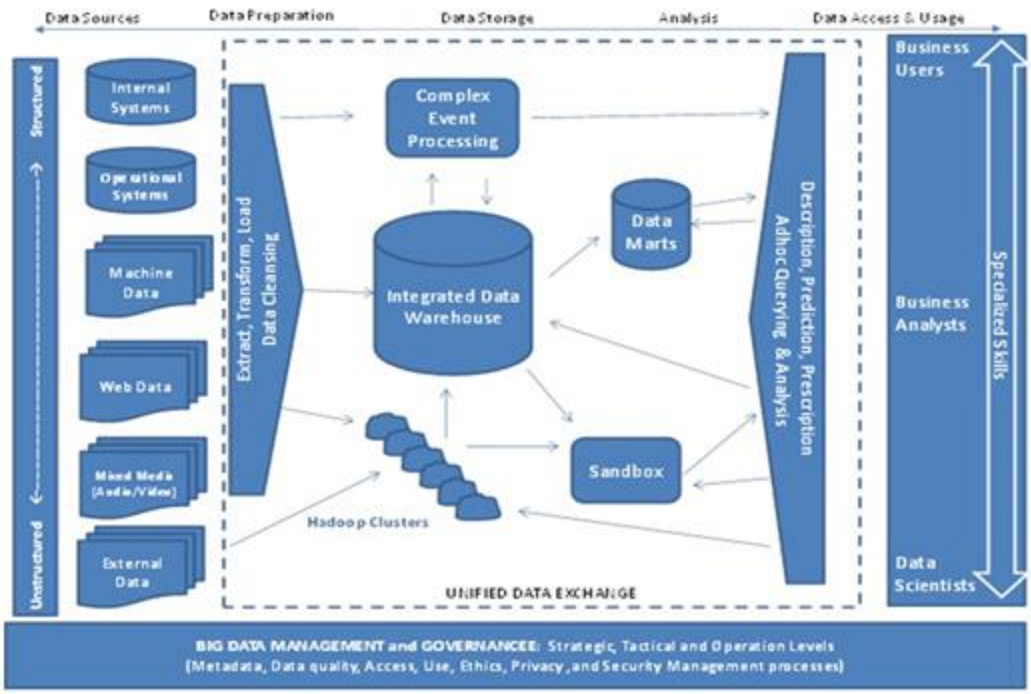

Figure 1. Big Data Analytics Framework

Structured data still represent the majority of data used for analytics according to surveys (Russom, 2011). Structured data reside in spreadsheets, tables, and relational databases corresponding to a data model that addresses the properties and relationships between them. They have known data lengths, types, and restrictions. They can be easily captured, organized, and queried due to the known structure. Figure 1 shows structured data coming from sources such as internal systems producing reports, operational systems capturing transaction data, and automated systems capturing machine data such as customer activity logs.

Increasingly, semi-structured data are used for analytics (Russom, 2011). These data lack a strict and rigid structure but have identifiable features. For example, photos and images can be tagged with time, date, creator, and keywords to assist users to find and organize them; emails have fixed tags such as sender, date, time, and recipient attached to the contents; and webpages have identifiable elements that allow companies to exchange information with their business partners. Industry standards such as Extensible Markup Language (XML) enable computing devices to identify these data by defining a set of rules for processing.

Unstructured data, primarily in the form of human language text, are growing in importance for analytics (Russom, 2011). These data are ill-defined and include images, video, audio, emails, presentations, wikis, blogs, webpages, and text documents. Tools such as text mining or text analytics are maturing and enabling people to analyze unstructured data. For example, hospitals can search physician instructions, patients' charts, and prescription information to identify potential adverse drug interactions. These data are primarily from external sources such as social media, the Web, and sensors.

4.4 Data Preparation

Data preparation includes extracting, transforming, and loading (ETL) data and data cleansing. ETL processes involve expert judgment and are essential as foundations for analysis. Once data are identified as pertinent, a data warehouse team extracts data from primary sources and transforms them to support the decision objective (Watson & Wixom, 2007). For example, a customer-centric decision may require consolidating records from different sources, such as an operational transaction processing system and social media customer complaints, and linking them through a customer identifier such as a zip code. Source systems can be incomplete, inaccurate, and difficult to access, so data are cleansed to ensure data integrity. Data may need to be transformed to be useful in analysis such as creating new fields to describe customer value. Data may be loaded into a traditional data warehouse or in Hadoop clusters. Loading can occur in a variety of methods with a data warehouse either sequentially or in parallel by tasks such as overwriting existing data, updating data hourly or weekly.

4.5 Data Storage

Traditionally, data are loaded "into a data store that is subject-oriented (modeled after business concepts), integrated (standardized), time-variant (permits new versions), and nonvolatile (unmodified and retained)" (Watson & Wixom, 2007). Thus, loading requires an established data dictionary and a data warehouse that serves as the storage location for verified data that the organization will use for analysis. Data related to specific uses or business departments might be consolidated into a data mart for ease of access or to restrict access. However, moving and processing extremely large amounts of data as a single dataset with a single server is not feasible with current technology. Thus, storing and analyzing big data requires processing to be split across networked computers that can communicate and coordinate their actions. Hadoop is an open-source framework that permits distributed processing of data across small to large clusters of computers using local computation and storage. Hadoop is not an ETL tool; it supports ETL processes running in parallel with, and complementary to, the data warehouse (Awadallah & Graham, 2011). Results from Hadoop cluster may be passed to the data warehouse or analyzed directly.

4.6 Analysis

Analysis spans a wide range of activities that may occur at various stages in managing and using data (Kulkarni, 2013). Querying data is often the first step in an analysis process and is a predefined and often routine call to data storage for a particular piece of information; by contrast, ad hoc querying is unplanned and used as the need arises for data. Descriptive analytics is a class of tools and statistics to describe the data in summary form. For example, analysts may report on the number of occurrences of different metrics such as number of clicks or number of people in certain age groups, or they may use summary statistics such as means and standard deviations to characterize data. Descriptive analytics may use exploratory methods to attempt to understand data; for example, clustering can identify affinity groups. Exploratory analytics is often helpful in identifying a potential data item of interest for future study or guiding the selection of variables to include in an analysis. Predictive analytics refers to a group of methods that use historical data to predict or forecast the future for a specific target variable. Some of the better-known predictive methods are regression and neural networks. Prescriptive analytics is an emerging field that has received more attention with the advent of big data since more future states and a wider variety of data types can be examined than in the past. This analysis attempts to examine various courses of actions in order to find the optimal one by anticipating the result of various decision options (Watson, 2014).

Many of these processes have been standard in data analysis for a long time. What is different in the case of big data is the larger amount and variety of data under consideration and, possibly, the real-time nature of data acquisition and analysis. For example, Hadoop can be used to process and even store raw data from supplier websites, detect patterns indicative of fraud, and develop a predictive model in a flexible and interactive manner. The predictive model could be developed on Hadoop and then copied in the data warehouse to find sales activity with the identified pattern. A fraudulent supplier would then be further investigated and possibly discontinued (Awadallah & Graham, 2011). As another example, graphic images of items for sale could be analyzed to identify tags that a consumer is most likely to use to search for an item. The results might result in improved labels to increase sales.

The “analytics sandbox” shown in Figure 1 is a scalable, developmental platform for data scientists to explore data, combine data from internal and external sources, develop advanced analytics models. and suggest alternatives without modifying an organization’s current data state. The sandbox can be a standalone platform placed in the Hadoop cluster or be a logical partition in the enterprise data warehouse (Kobielus, 2012). For example, eBay provides virtual sandboxes inside the enterprise data warehouse to allow employees to explore or manipulate data or to even combine new data sets to encourage experimentation in a managed environment (Laskowski, 2012).

Conventional architectures use a save-and-process paradigm in which data are first saved to a device and then queried (Buytendijk, 2014). Complex event processing (shown in Figure 1) is a proactive process-first monitoring of real-time events based on data such as operational systems to enable organizations to make decisions and respond quickly to events as they occur such as potential threats or opportunities (Chandy & Schulte, 2009; Buytendijk, 2014). The software gathers information from selected data sources, identifies patterns, and notifies other systems or people. Events cannot always be predicted. In complex event processing, the event acts as a trigger, and organizations that respond to events are referred to as event-driven (Luckham, 2002). For example, a regional sales manager who is notified that a particular item such as a medication is suddenly in high demand could possibly adjust inventory to respond in a timely way. Complex event processing enables accurate and actionable information for appropriate response.

The combination of real-time event processing, data warehousing, data marts, Hadoop clusters, and sandbox provide a data analysis and storage infrastructure that supports a stable environment while enabling innovation and real-time response.

4.7 Data Access and Usage

In the final stage of the big data process, users and analysts use data by querying, accessing reports, and performing analytics. The Eckerson (2011) framework categorizes users into two groups: casual users and power users. Casual users—executives, managers, front-line workers—are users who use the basic capabilities of the system. The system's reporting functionality may be used as and when needed, or analytical processing may be integrated into the workflow of these users. For example, a call center operator, while talking to a customer, may have a display of the customer's value, preferences, and potential offers for cross-selling. Power users—business analysts, analytical modelers and data scientists—exploit the full capabilities of the BI systems available to them. They have good knowledge of the system's features, capabilities, and limitations and a deep understanding of business processes and the data that sit behind those processes.

Our framework identifies three types of users: business users, business analysts, and data scientists. Business users have basic skills and domain-based needs. They comprise the casual users in Eckerson's (2011) framework, but they also include external users such as customers and suppliers who may connect via applications that depend on analytical processing. For example, an airline customer building and pricing a multi-city itinerary may be using a sophisticated scheduling application with a dynamic pricing engine without being aware of the complex processing involved. Business analysts are users who have more analytical skills than business users: they can analyze data, understand how data is organized, retrieve data via ad hoc queries, produce specialized reports, build what-if scenarios, and interactively perform deeper analysis to support their decision-making.

While these two roles roughly correspond to the two types of users in Eckerson's (2011) framework, our framework identifies data scientists as different, more advanced, data users. A data scientist has a strong background in mathematics, statistics, and/or computer science, an equally strong business acumen, and an ability to communicate with both business and IT leaders in a way that can influence how an organization approaches its business challenges with the help of data. A data scientist can develop descriptive and predictive models (perhaps using the discovery platform; e.g., Sandbox), evaluate models, and deploy and test them through controlled experiments. In the context of big data, data scientists may advise organizations in interpreting rich data, managing large amounts of data, integrating data from multiple sources, and creating visualizations to aid in understanding data. They may also participate in communicating the insights/findings not only to the specialists and scientists on their team but also to business leaders, and, if required, to a non-expert audience.

4.8 Big Data and Management

Data management and governance are integral to any business and especially to BI (Watson, 2009). Due to the increase in complexity of issues related to big data, organizations face new ethical, legal, and regulatory challenges with big data management and governance (Ballard et al., 2014). They have to balance their data governance process to manage top-down and bottom-up needs (Eckerson 2011). The big data management and governance component identified in our framework suggests a comprehensive data management approach that address issues at the strategic, tactical, and operational levels. At a strategic level, a successful data governance process should span the entire spectrum from obtaining data to its consumption and ensure that big data efforts undertaken align with business strategy. Decisions include deciding what internal and external data sources to use, selecting and deploying appropriate big data technologies for storing data and for unified data exchange, and investing in training programs to have the appropriate skill sets to make informed, timely decisions. In the context of big data, organizations store more data than what their immediate needs might be, which can expose them to more privacy and security risks. Appropriate governance mechanisms that ensure regulatory and legal compliance is crucial.

In contrast to traditional BI where most business units and users are provided with appropriate reports/data for decision making, in the context of big data, many organizations now allow their business units to find ways to use and analyze data to better serve their needs. Thus, it is not uncommon to see big data projects stem from various business units. Hence, managing big data projects is critical. At a tactical level, good governance process should include ways to prioritize big data projects, set metrics to assess projects and their usefulness, and deploy knowledge management processes so that there is effective sharing of resources in the organization as it relates to big data efforts.

Another major change in the big data context is the management efforts at the operational level. Latency (i.e., speed to access data) is critical. Since data being used by organizations is both internal and external, decisions need to be made at the operational level on how to handle data from disparate sources (e.g., about how to structure unstructured data, how to ensure data quality (i.e., master data management), what in-memory databases to use for storage, and what no-SQL approaches will be used to access data).

5. Academic Research and Practitioner Perspectives

We use the process view of our framework to structure the review starting with data sources and ending with big data management and governance. In each category, we highlight contributions from both academic research and practitioner literature and interviews. The research findings and other observations presented in these works led us to identify opportunities for future investigations.

5.1 Data Sources

5.1.1 Academic Research

The term big data is often used to refer to unstructured data, particularly from social media sources such as Facebook, Twitter, and blogs (Androile, 2012). The literature that we reviewed reveals research aimed at interpreting the meaning of human written language (Kane et al., 2014) and assessing the relevance of textual language to a particular problem or situation (Chau & Xu, 2012; Oh et al., 2013) along with the reliability of that information as a precursor to trusting it (Gefen, Benbasat, & Pavlou, 2008). Kane et al. (2014) identify key differences between offline social networks and online social media networks. Oh et al. (2013) use rumor theory to examine the reliability of community intelligence obtained from the online community of amateur reporting during crisis. They found that source ambiguity is the most important rumor-causing factor, which suggests that meta-data associated with source data provides additional information that should be considered when judging information quality or veracity.

5.1.2 Practitioner Perspectives

The practitioner literature shows an evolution in the concepts of structured/unstructured data with new open standards. The World Wide Web Consortium (2014) describes a Web of data referred to as the semantic Web that provides a common framework to allow data to be shared and reused across applications, enterprises, and communities. "Vocabularies" are used to define concepts and relationships to enable data integration and organize knowledge; they are the basic building blocks for inference techniques. According to some practitioners, these evolving standards may affect the manner in which data sources are used in the big data implementation process and challenge our ideas of data blending. Some practitioners argue that there is no such thing as unstructured data since all data has some structure; the structure is simply more flexible than that demanded by a relational database (Porter, 2014).

Another way to approach data sources is to classify them as one of three corporate data types: structured corporate data, unstructured repetitive corporate data, and unstructured non-repetitive corporate data (Inmon, 2014a). Structured corporate data can be stored in a traditional database. An example of unstructured repetitive corporate data is records of telephone calls; the repetitive nature of the data provides heuristics for obtaining value from the data. The last category of unstructured non-repetitive corporate data (e.g., corporate contracts, warranty claims, email contents, healthcare data) is the most challenging since the content and structure of each item could be different. As such, an organization needs different approaches to obtain value from data depending on the data types themselves (Inmon, 2014a, 2014b, 2014c).

Practitioner interviews and literature revealed an increasing interest in the "Internet of things" and the associated brontobytes (1000 yottabytes or 10^27 bytes) of data (Zaslavsky, 2014; Grover & John, 2015). For example, the *Commonwealth Scientific and Industrial Research Organisation (CSIRO)* in Australia is developing technology to increase crop yield by performing sensor-based monitoring of plants, soil, and environmental conditions at high resolution (Zaslavsky, 2014). Real-time, online analysis of sensor data permits one to interactively assess crop performance and crop selection based on expected conditions, irrigation, and fertilizer. Embedded sensors allow one to precisely measure the use of physical objects such as drilling equipment, which creates opportunities for alternative pricing strategies and new business models (Grover and John, 2015).

5.2 Data Preparation and Storage

5.2.1 Academic Research

The literature on data preparation and storage mainly includes research commentaries that help identify future trends and research questions (e.g., Gillon et al., 2014; Andriole, 2012). Andriole (2012) surveyed practice on future trends, and Gillon et al. (2014) provides a panel report on analytics that describes the need for change in the existing data preparation and storage architecture in organizations. The vast digital trail triggered by the Internet and accelerated by the growth in social media has led to a big data explosion, which has created a need to change traditional data structures. The demand for real-time processing of a variety of data types and from diverse sources can provide organizations with immediate insights on customer behavior and business transactions and create both incremental and radical change to typical decision support frameworks (Gillon et al., 2014).

However, the combination of technologies required to collect and combine the variety of structured and unstructured data to gain insights is still not established. Many major organizations are currently experimenting with a variety of storage technologies available to enable big data analysis such as in- memory and in-database processing solutions, streaming engines, Hadoop clusters, and traditional data warehousing frameworks (Watson 2014). Gillon et al. (2014) describe the current growth in technologies that support big data as the "wild west of technologies" with no proven set of tools to handle data storage and processing needs for big data. As a result, many research questions in this space focus on the evolution of traditional data architectures and on identifying the right mix of platforms that would meet the analytics needs of the organization in an efficient and cost-effective manner (Andriole 2012; Watson 2014).

The sampling of research literature provides further evidence that data preparation and storage to accommodate big data is still being developed. Empirical research focusing on the multi-platform framework required for big data storage is still in its infancy. As a result, much of the academic research still focuses on traditional structured data architectures or different data preparation and storage platforms in isolation. For instance, van Valkenhoef et al. (2013) have looked at creating standardized data models and data standards for new data sources such as clinical trial data. However, as the research commentaries suggest, innovation and evolution are expected in data storage with the dawn of the era of big data (Watson, 2014). For instance, advances in data storage techniques and analytics has also created the need for automation of decision support processes that historically relied on human intuition (Gillon et al., 2014). Future research in this area will address methods of combining structured and unstructured data and the right combination of storage options to optimize analytics.

5.2.2 Practitioner Perspectives

The practitioner literature indicates a shift in the mindset that data storage should also make accommodations for a discovery platform in addition to storing data. One approach for data warehousing to deal with exponential increases in data is to move from extract-transform-load (ETL) to a new design of extract-load-transform (ELT) (Subramanian, 2013). In ELT, raw data is loaded directly to the target and transformed there to reduce the time between data extraction and availability to users. Thus, the need for discovery at the cost of data quality attributes (such as standardization) has become a theme in big data research. In the context of big data, practitioners suggest flexible data storage such as "data lakes" or "data clouds" that are easily expandable and can accommodate a variety of data formats (Roski et al., 2014). Data lakes store near-exact or even exact copies of the source format in order to present an unrefined view of the data independent of any of the compromises made in storing data in traditional architectures such as a data warehouse (Schlegel, 2014). Data lakes also enable flexible analytics schemas that can be developed "on the fly" to answer a particular question without affecting the raw data (Herman et al., 2013).

Data architectures in most organizations today rely on hybrid architectures involving a myriad of database and file management systems, and data is stored and managed in a variety of locations: in memory, on disk, or in the cloud. Research is needed on data-intensive distributed systems such as Hadoop that separate how data is stored from the way it is processed. The integration of Hadoop capabilities with existing enterprise architectures is key to enterprise adoption (Potter, 2014). Practitioners also find that most data warehouses have reached maximum storage capacity and will need expensive upgrades to accommodate big data (Swoyer, 2010). In addition, the underlying data warehousing platform is not scalable enough to support new sources of data (internal or external) and maintain adequate query performance. Thus, many companies are implementing new, specialized analytical platforms designed to accelerate query performance when running complex functions against large volumes of data (Swoyer, 2010).

Companies are realizing that, with the growth of big data, there is a need to have a data storage metric such as dollar-per-TB (Swoyer, 2013). Currently, more and more organizations are considering analytics technologies that actually work best off of raw source data as opposed to squeaky-clean data loaded into the warehouse. It may be cost effective to store raw data in a Hadoop distributed file system (HDSF) and move analytical computing closer to the data (Evans, 2013). It is considered a favorable choice in terms of matching a workload to a platform best suited to it (Swoyer, 2013).

5.3 Analysis

5.3.1 Academic Research

Current research on methods to analyze big data are focused in four areas: extending intelligent methods developed to analyze large datasets, developing methods to analyze unstructured data, investigating methods that combine structured and unstructured data to improve model prediction, and visualizing data and analyses to aid interpretation. Much of the analytical methods research is found in the computer science and engineering literatures rather than the IS literature. We briefly describe data analysis research and refer the reader to alternate references for details.

Intelligent methods have been developed and expanded over the past few years with an emphasis on big data. For example, the IEEE (2014) conferences and workshops on big data provide a forum for computer science and engineering researchers to focus on analyses of these data. In general, these approaches use mathematics and artificial intelligence to develop predictive models. Methods include descriptive analytics, such as clustering or network analysis, and predictive analytics, such as regression, logistic regression, decision trees, and neural networks.

Methods to analyze unstructured data, particularly in the form of textual data, have been investigated and expanded in response to the availability of electronic text communication (Parameswaran & Whinston, 2007). Evangelopoulos, Zhang, and Prybutok (2012) discuss latent semantic analysis (LSA) to describe the semantic content in textual data as a set of vectors. The meaning of a passage of text is related to patterns of presence or absence of words. Applications include analyzing customer feedback, interpreting social media, and managing knowledge repositories. Research is needed in intelligent selection of parameters for analysis, interpretation of unstructured data from various sources, combination of multiple approaches to analysis, and the finding of new application areas (Evangelopoulos et al., 2012).

Zheng et al. (2012) illustrate how to investigate methods that combine structured and unstructured data to improve model prediction. They describe methods to derive competitive intelligence in BI systems, such as customer visits and purchasing behavior in e-commerce without the need for individual transaction history. They use a constrained optimization model to infer key competitive measures (i.e., penetration, market share, share of wallet) using less granular and aggregate data than has been used in past studies, and they combine data about online retailers with transactional data from syndicated data providers.

The significant growth in use of social media data has spurred research in analytic methods that can help firms monitor public sentiment of their brand and products (da Silva et al. 2014). This research can also aid consumers search for appropriate products based on sentiment analysis (Lau et al., 2014).

Lycett (2013) points out that data visualization is maturing to present complex data quickly and clearly in ways that are informative while also being pleasing to the viewer. Tufte (1997) investigates the presentation of information and “visual explanations” or “visual displays of data” to aid interpretation. He demonstrates that context and interpretation of the visual display by a decision maker need to be considered in creating a visualization. Although Tufte’s (1997) body of research is applicable to all types of data, large and small, representing big data requires more-sophisticated approaches than that found in traditional spreadsheets.

5.3.2 Practitioner Perspectives

The practitioner literature points out that gaining value from big data requires a new way of thinking about and analyzing data (Herman & Delurey, 2013; Hey, 2010; Grover & John, 2015). With the concept of a data lake containing all of an organization’s data regardless of type (see section on data storage), we need new methods of connecting data sources. Extract/transform/load processes are tied to structured databases and data warehouses, while relationships are derived at the time of analysis in a data lake. Metadata tags provide descriptions of data elements and become connectors between data so that an analyst can retrieve any relevant data, regardless of data type, to answer a particular question.

We need new techniques to develop these flexible schemas “on the fly” to answer a particular question without affecting the raw data (Herman et al., 2013). For example, real-time processing permits one to identify opportunities or threats so that they can be acted on more quickly, which would make complex event processing (shown in Figure 1) more effective (Schlegel, 2014). Ultimately, we need to understand the true relationship between the measures of highest interest and related events. Such relationships, or “analytical pathways”, are key to understanding cause-and-effect as part of analytics (O'Rourke, 2013).

Although techniques for analyzing structured data are well developed, robust approaches are needed for other big data assets such as images, video, human language, sound, and even three-dimensional objects (Schlegel, 2014). Interpreting these data types is often context dependent (Feigenbaum, 2014), so analytical strategies are needed that are inspired by the way the human brain processes information, draws conclusions, and makes decisions (i.e., cognitive analytics) (Chandrasekaran, 2014). Even if data are interpreted, fusing data on a massive scale by incorporating various data types is a persistent problem (Herman et al., 2013). For example, "social analytics" (i.e., social filtering, social network analysis, social channel analysis, sentiment analysis, social media analytics) needs to be combined with other types of data that describe human behaviors (Buytendijk, 2014).

Big data analytics require new ways to facilitate the discovery of hidden patterns in large, complex datasets without traditional model building or algorithmic development (Schlegel, 2014). Sometimes referred to as "smart pattern discovery", data visualization enables users less skilled than data scientists to interact with and explore data (Schlegel, 2014). We need techniques that increase accessibility of data analytics to a larger number of users. New dashboard designs are being developed with new requirements such as interactivity and data flexibility (Porter, 2014). We are moving toward data discovery and visual analytics with an emphasis on storyboarding, scenarios, and use-cases that enable business hypothesis testing and insight generation (Mohanty, 2014).

Technologies to deliver these capabilities are constantly evolving. In-memory analytics, in which all data to be analyzed are loaded into memory and analyzed there, is many times faster than traditional disk-read systems (vom Brocke et al., 2014). This technology is becoming mainstream, driven by declining memory prices coupled with the widespread adoption of 64-bit computing (Buytendijk, 2014). Cloud-based analysis of big data is being accomplished with technologies that permit large-scale batch processing across hundreds or thousands of computers, each with several processers to permit parallel processing (e.g., Hadoop) (Buytendijk, 2014). Hadoop is a core component of big data analytics, but we need a more effective toolkit to reduce the complexity of the MapReduce syntax and help developers generate and execute Hadoop jobs and abstracts (Narayanan, 2013).

5.4 Data Access and Usage

5.4.1 Academic Research

The literature on data access and usage focuses on users and their roles and how they interact with data, including big data. Andriole (2012), while discussing technology trends, predicts that users will be segmented into a myriad of roles with a need for different devices (e.g., tablets, smartphones, hybrids, laptops, desktops, and heavy devices), technology platforms, and access mechanisms to fit their roles.

Demirkan and Delen (2013) suggest an architecture for delivering data, information, and analytics as a cloud-based service, which is possibly a necessity in the context of big data. Their conceptual framework proposes tailoring these services to the types of users and their specific decision and analytical needs, such as ad hoc querying, reporting, OLAP, dashboards, intra- and inter-net searching for content, data and information mashups, optimization, data mining, text mining, simulation, and automated decision systems.

Research on data access and usage focuses on post-adoptive behavior of users in actually using the data and information, especially when using BI systems is voluntary. For example, Deng and Chi (2012) examine the problems and causes that deter the use of implemented BI systems by both regular and power users. Li et al. (2013) use motivation theory to predict the impact of rich intrinsic motivation on routine and innovative use of BI systems. Other researchers have prescribed how data modeling and analysis can be integrated into the decision making process with examples of specific applications. For example, van Valkenhoef et al. (2013) suggest a data model, a decision support system, and an analysis technique for conducting clinical trials. Brydon and Gemino (2008) show how data mining can be integrated into the decision making process of selecting which video games to develop based on prior blockbuster performance data.

5.4.2 Practitioner Perspectives

The practitioner literature reiterates the evolving roles of the users of BI systems with the data scientist emerging as distinct from business analyst and statistician (Laney & Kart, 2012; Davenport & Patil, 2012). The implications of this evolution are many: training and educating data scientists, creating new organizational structures, and transforming business processes for a data science teams (Herman et al., 2013). The literature also points out that big data and related technologies are new to BI and data warehouse professionals. Experience with early adopters has shown that learning (e.g., Hadoop and related technologies such as Hive, HBase, MapReduce, Java programming, etc.) takes significant time and training (Russom, 2014).

Ten years ago, a dashboard visually depicted a tabular report. Today, dashboards have evolved to provide visual analysis, data discovery, and self-service BI. In the context of big data, business users will be more and more empowered to explore data without necessarily knowing what they are looking for.

Furthermore, mobility, data freshness, and collaboration have become commonplace in dashboard requirements (Porter, 2014).

The practitioner literature also emphasizes the need to provide data access via mobile devices to users who may not be traditional analysts. At the same time, users need to be trained to use the data effectively (Powell, 2013). Moreover, making the right data available for access by any device at any time is a challenge. Sophisticated data visualization tools are especially required for interpreting big data (Herman et al., 2013). The literature points out the need to investigate the role of human behavior in big data access and usage. There are many factors to investigate: from understanding the big data vision and motivations to developing trust and managing conflict that may arise from lack of such understanding (TWDI, 2013).

As we mention earlier, our concept of a unified data exchange represents integration of multiple processes. With big data analytics, the lines between data preparation, storage, and analysis are becoming blurred. For example, some type of data may not go through the traditional ETL process and may reside directly in Hadoop clusters where transformation and analyses may take place simultaneously on the fly. Complex event processing may take place with real-time data freshly extracted and analyzed along with business rules or profiles stored in the traditional data warehouse. A data scientist may use the models and analytics tools available in the sandbox to "play" with the streaming in-memory data along with the data sets in the warehouse. These examples show that the analytics processes in the UDE, preparation, storage, and analysis are not necessarily sequential, especially in the case of big data analytics. Thus, UDE represents a conceptual integration of otherwise distinct processes. Note that several organizations and consulting firms are experimenting with alternative representations of this concept and these representations are evolving as our understanding of the components and their capabilities evolves.

The practitioner literature points to new business models that may result from "hyperscale" businesses (i.e., big data businesses that exploit immense digital data stores) (Chui & Manyika, 2015). For example, companies that sell physical assets could use machine-to-machine data to evolve into service businesses based on usage charges. Rapid growth is possible since the networks can be easily expanded and the marginal cost of adding additional devices or users is small. Hyperscale businesses can disrupt traditional business models by automating process improvements and quickly experimenting with customer preferences on a large scale. Companies with access to large amounts of data may be able to compete at hyperscale in some segments of their business (Chiu & Manyika, 2015).

5.5 Big Data Management and Governance

5.5.1 Academic Research

Academic research on big data management and governance is at a very early stage and has primarily focused on its importance. In a recent panel report, Gillon et al. (2014) stress that organizations need to rethink the governance and management of big data because there are significant changes in internal environment. Many organizations are still in the process of building the needed skills and technical capabilities to tackle big data. Based on panel discussions, they present a 4D framework for analytics: decisions rights (who will own decision rights and how to centralize/decentralize the rights), department role and configuration (determining role of IT in analytics and how it interfaces with other units), dollars (making monetary decisions on projects), and delivery (securing and training staff).

In his tutorial on big data analytics, Watson (2014, p. 1264) states that some organizations are creating "analytics centers of excellence" for providing strategic direction for big data analytics, building appropriate capabilities for use of analytics, establishing guidelines and standards, and prioritizing projects. In identifying seven technology trends for 2015, Andriole (2012) recognizes that organizations will see a dramatic shift in governance due to the change in velocity of business. As technologies for enabling data analytics are radically changing and evolving combined with dynamic data needs of business units, appropriate mechanisms to manage and govern the change is important. Recent research also confirms the increasing need for governance and compliance as the analytics landscape is changing (Chae, 2014; Demirkan & Delen, 2013). The importance of governance and management, and the lack of detailed research in the area, provides opportunities for academic research that are identified in a later section.

5.5.2 Practitioner Perspectives

It was evident from the practitioner literature and interviews that governance and management of big data need new paradigms (O'Neil, 2012). With increased use of internal and external data in decision making, organizations face increased risks. Data inaccuracies can lead to poor decision making. As organizations make extensive use of social media data to make important decisions, veracity of data is one of their biggest concerns. Appropriate policies and strategies for data management, big data project management, the securing of data both in house and in cloud, compliance and control to meet regulatory requirements, and training of employees to ensure effective use of data all fall under the purview of overall data management and governance. Mergers and acquisitions can pose more challenges with integrating data architectures if organizations do not have well-defined management mechanisms (Powell, 2013).

Another way to manage big data is using data lake (Schlegel, 2014; Violino, 2014). As Schlegel (2014) notes:

> *By its definition, a data lake accepts any data, without oversight or governance. Without descriptive metadata, and a mechanism to maintain it, the data lake risks turning into a data swamp. Moreover, every subsequent use of data means analysts start from scratch, like a form of data amnesia.*

A data lake also adds more security and access control challenges as the technologies handling data lakes are still in their early forms. This lack of maturity reinforces the need for good data management and governance mechanisms.

If organizations use cloud service providers, then they need to ensure those providers have up-to-date security and policies to share data and collaborate across organizations. Practitioners also stress the importance of addressing organizational culture in the context of big data as attitudes on ethics, privacy, and security can vary significantly across organizations. For example, in the healthcare sector, the policy implications of using big data are that many current practices and policies related to data use, access, sharing, privacy, and stewardship need to be revised to ensure protection of patients' confidentiality (Roski et al., 2014). Big data research continues to identify cybersecurity threats. Gartner reports that 25 percent of companies will have adopted big data analytics for at least one security and fraud detection use case (Rivera, 2014). Identification prior to attack will continue to evolve as a potential research area. While there are practitioner blogs, interviews, and white papers on the topic, systematic research is needed to understand appropriate privacy and security policies for different industry sectors.

However, the biggest challenge for businesses is to develop a simple big data plan "for how data, analytics, frontline tools, and people come together to create business value" (Biesdorf, Court, & Willmott, 2013). The plan should provide a common language for executives, managers, and data scientists to assess opportunities for business value and identify priorities. A successful plan will focus on three elements: assembling and integrating data with associated governance, developing advanced analytic models, and creating intuitive tools that integrate data insights into business decisions. Big data planning differs from traditional business intelligence plans in integrating data across company divisions and requiring investment in new data architectures and analytics. It is “at least as much a management challenge as a technical one” (Biesdorf et al., 2013), and we need new organizational skills and thought processes for effective implementation. A 50-50 ratio of data and modeling to training is suggested for planning purposes for big data.

6. Research Directions Based on Big Data Analytics Framework

We identified potential research questions in the field of big data using representative literature and practitioner interviews as we discuss above. We list the research opportunities in each component of the big data framework in Table 5. Because these opportunities are also evident from the practitioner literature, we trust that further academic research in these areas will help increase the relevance of academic research to practice. While we do not claim that the list provided is exhaustive, we hope it will aid researchers in investigating relevant topics for research.

Table 5. Research Opportunities Identified from Representative Literature and Practitioner Interviews

Components of the big data analytics framework (Figure 1)	Potential research questions
Data sources	What external data sources make strategic sense for an organization? What steps should organizations take to evaluate the usefulness of external data sources? How should relevant data sources (e.g., textual, video, voice, imagery) for a given problem be identified before retrieval? What metrics should be used to identify relevant data sources for a problem? How will new standards for data affect the manner in which data sources are used in the big data implementation process? How do open standards affect the notion of structured/unstructured data (e.g., Semantic web consisting of vocabularies and ontologies)? As the notion and definition of big data evolves, are there other ways to classify data? How can relevant data be identified from the Internet of Things (IoT)?
Data preparation	What are the challenges of data blending? Can the reliability of various data types be assessed as a precursor to trusting it? What techniques can be used to prepare real-time streaming data for analysis? What details need to be stored permanently? How can one determine the frequency and granularity of snapshots of data streams? What types of algorithms are necessary for detecting patterns in data streams? Given the question of veracity of data sources, what technologies and methods can help to prepare data for analysis? How can organizations identify the right mix of platforms that would meet the analytics needs of an organization in an efficient and cost effective manner?
Data storage	What are the components of a hybrid architecture (e.g., data lake) for big data? What combination of data architectures (e.g., data warehouse and Hadoop clusters) would fit different contexts? How will the use of Hadoop as data staging impact traditional data warehousing infrastructure? When is it appropriate to use in-memory and in-database processing solutions, streaming engines, Hadoop clusters, and/or traditional data warehousing frameworks? What kind of flexible storage can accommodate undefined data formats? What metrics should be used for the cost and efficiency of data storage in the era of big data? How can virtualization impact/improve the big data process in organizations? How can an organization assess the risk and benefits associated with storing data in the cloud?
Analysis	What intelligent methods can be developed to assist with big data analytics? How can different data types be fused on a massive scale? How can cause and effect relationships between various big data measures (i.e., analytical pathways) be discovered/validated? What visualization methods are most useful for decision support with big data? How can other big data assets, such as images, video, sound and even three-dimensional object models be used to support decision making? What new techniques can be developed to push data to nodes for processing (e.g., mappers, partitioners, combiners, and reducers)? What new tools can assist with abstracting code and reducing complexity? How will cognitive analytics affect the interpretation of big data for decision making? How does an analyst's perception influence cognitive analytics and its effectiveness? How can real-time processing permit identification of opportunities or threats so that they could be acted upon quickly and effectively? How can data assets be augmented with annotations using natural language processing, machine learning, tagging, and so on to assist in drawing useful inferences? What types on flexible schemas can be developed "on the fly" to answer a particular question without affecting the raw data?

Table 5. Research Opportunities Identified from Representative Literature and Practitioner Interviews

Data access and usage	How should data science teams be constructed and organized? What skill mix is needed for effective data science teams in organizations? What kind of education and training is needed for analytics teams of the future? What kind of talent management processes need to be instituted? How can user engagement in implementation and usage of analytics systems influence building of big data capabilities? What types of devices are appropriate for different types of analytical needs by users (for efficiency / effectiveness gain)? What incentive mechanisms are needed for users to effectively adopt analytics systems in their routine and non-routine decision making tasks? How can data modeling and analysis be integrated into the decision-making process? How can cloud-based services be effectively used to deliver analytics as a service, especially in the context of big data? What new business models could result from "hyperscale" use of data?
Big data management and governance	How can an organization build an effective big data governance strategy? How should big data be governed? What type of framework can guide big data governance? What capabilities (technical and non-technical) should an organization acquire to succeed in big data efforts? What controls and assurance mechanisms should be adopted to ensure appropriate access and use of big data? What strategies and policies should organizations have in place to ensure privacy and security of big data? How can organizations prioritize big data projects? What metrics are appropriate for evaluating big data projects? What strategies work best for building and retaining a workforce with the various skills needed for big data? What organizational structures encourage the use of big data for decision making? Should efforts be centralized or decentralized, or are there alternative structures that work better for big data management? What ethical issues arise with the use of big data? How can ethical issues be effectively managed? What kind of knowledge management strategies and practices should organizations adopt to enable sharing big data know-how that is potentially dispersed across business units? How does big data analytics impact organizational processes? What type of competitive advantages are organizations realizing with the use of big data?

7. Summary

We propose an updated framework for the BI environment in the context of big data. Academic and practitioner panels and discussions from the 2012 and 2013 pre-ICIS analytics events and previously published frameworks (Watson, 2009; Eckerson, 2011) provided the initial ideas for building the framework. We had leaders in the practitioner community whet the initial ideas. The resulting big data analytics framework depicts a process view of the various components that form the analytics process including sourcing, preparation, storage, analysis, access, and usage. In addition, the framework represents three types of users: business users, business analysts, and data scientists. It also captures organizational issues in big data management and governance at the strategic, tactical, and operational levels.

We reviewed a significant sample of top academic and practitioner literature in the context of the components of the framework. This structured review of current literature helped us identify the gaps in research and also propose new avenues for research. We find that the challenges presented by the nature of big data offer unique opportunities for research in each component of our big data analytics framework.

Questions regarding sourcing of big data revolve around what data sources are strategically important and identifying such sources because the overabundance of data will make it necessary to tune in on those data that are most beneficial. Eventually, standards and metrics will need to be developed to measure progress in this direction.

Preparing big data for storage and organization is a major challenge. Research questions include topics such as preparing data versus keeping it in raw form, ensuring quality and preparation of real-time streams, and building technology platforms needed to prepare data on the fly. The diversity with which big data will be used (including its unspecified use) necessitates delaying its preparation into a rigid structure, so that flexibility is maintained as long as possible.

Big data storage is one of the most technologically advanced research areas in practice with architectures such as Hadoop clusters becoming commonplace. Because of the different ways in which big data can be potentially used, questions still persist around how to strike the right balance between in-memory, in- database, traditional data warehouses, and on-demand data storage in the cloud. Perhaps the solutions will be in replicating big data in multiple formats.

The analysis area poses perhaps the most interesting unanswered questions because intelligent methods to assist big data analytics on a massive scale are yet to be developed. Semantic merging of different types of big data streams is an untapped area both technically and for decision support. Moreover, analyzing complex data objects such as images, video, sound, and three-dimensional models is still in its infancy.

On the usage side of the spectrum, behavioral research questions abound. The understanding of what problem can be solved by big data and the expertise needed to solve them is still evolving. Hence, questions focusing on the composition of data science teams, user engagement in building big data capabilities, and integration of big data analytics in user workflow will need to be answered. Moreover, BI as a service will become even more relevant in the context of big data as expert groups specialize in, for example, different types of service (storage/ processing/ analytics), types of industry, or problem type.

Finally, managing and governing big data is replete with challenging research questions. Although these questions are similar to the ones we face in the case of traditional BI, some are unique to big data, such as strategies and policies needed to identify, hire, train, and retain a big data workforce; managing security and privacy of sensitive big data; and determining parameters for storage and disposal of big data.

Acknowledgments

The individuals that participated in practitioner interviews requested anonymity since they were expressing their personal opinions and experiences with big data that may not necessarily represent a consensus in their companies. We want to express our deepest gratitude and respect to them and their companies for their collaboration. Industry participation does not constitute endorsement, and any errors in this paper are ours alone. However, we would like to thank the following companies along with those companies who prefer to remain anonymous for providing interviews and comments during our research: Apollo Education Group, Booz Allen Hamilton, IBM, Micro Strategies, NetApp, Northrup Grumman, Price Waterhouse Coopers (PwC), SAS, Sogeti, Teradata, and UnderArmour. We also thank the authors of academic and practitioner literature who provided comments and research that we used in this study. We offer a special thanks to Barbara Wixom at the MIT Center for Information Systems Research (CISR) for her support and guidance on an earlier version of the paper.

References

Agarwal, R., & Dhar, V. (2014). Editorial: Big data, data science, and analytics: The opportunity and challenge for IS resesarch. *Information Systems Research*, *25*(3), 443-448.

Andriole, S. (2012). Seven indisputable technology trends that will define 2015. *Communications of the Association for Information Systems*, *30*, 61-72.

Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the special issue-social media and business transformation: A framework for research. *Information Systems Research*, *24*(1), 3-13.

Association for Information Systems. (2011). *Senior Scholars' basket of journals*. Retrieved from http://aisnet.org/?SeniorScholarBasket

Awadallah A., & Graham, D. (2011). *Hadoop and the data warehouse: When to use which* (White paper). Cloudera and Teradata. Retrieved from http://www.teradatauniversitynetwork.com

Ballard, C., Compert, C., Jesionowski, T., Milman, I., Plants, B., Rosen, B., & Smith, H. (2014). *Information governance principles and practices for a big data landscape*. IBM Redbooks. Retrieved from http://www.redbooks.ibm.com/abstracts/sg248165.html?Open

Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard Business Review*. Retrieved from https://hbr.org/2012/10/making-advanced-analytics-work-for-you

Basole, R. C., Seuss, C. D., & Rouse, W. B. (2013). IT innovation adoption by enterprises: Knowledge discovery through text analytics. *Decision Support Systems*, *54*(2), 1044-1054.

Beath, C., Berente, N., Gallivan, M., & Lyytinen, K. (2013). Expanding the frontiers of information systems research: Introduction to the special issue. *Journal of the Association for Information Systems*, *14*(4), i-xvi.

Biesdorf, S., Court, D., & Willmott, P. (2013). Big data: What's your plan? *McKinsey Quarterly*. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_whats_your_plan

Brydon, M., & Gemino, A. (2008). You've data mined. Now what? *Communications of the Association for Information Systems*, *22*, 603-616.

Buytendijk, F. (2014). Hype cycle for big data 2014. *Gartner*. Retrieved from https://www.gartner.com/doc/2814517/hype-cycle-big-data-

Chae, B. (2014). A complexity theory approach to IT-enabled services (IESs) and service innovation: Business analytics as an illustration of IES. *Decision Support Systems*, *57*, 1-10.

Chandrasekaran, S. (2014). How data virtualization turbocharges cognitive analytics and big data. Retrieved from http://tdwi.org/Articles/2014/08/12/Data-Virtualization-Cognitive-Analytics.aspx?Page=1

Chandy, K., & Schulte, W. R. (2009). *Event processing: Designing IT systems for Agile companies*. New York, NY: McGraw-Hill Osborne Media.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*, 67-80.

Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *MIS Quarterly, 3*6(4), 1189-1216.

Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1189.

Chen, R., & Sharma, S. (2013). Understanding member use of social networking sites: A value analysis. *Communications of the Association for Information Systems*, *33*, 97-114.

Chui, M., & Manyika, J. (2015). Competition at the digital edge: "Hyperscale" businesses. McKinsey Quarterly. Retrieved from http://www.mckinsey.com/insights/high_tech_telecoms_internet/competition_at_the_digital_edge_hyperscale_businesses

da Silva, N. F. F., Hruschka, E. R., Hruschka, E. R., Jr. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, *66*, 170-179.

Davenport, T. (2013). Analytics 3.0: In the new era, big data will power consumers products and services. *Harvard Business Review*. Retrieved from https://hbr.org/2013/12/analytics-30

Davenport, T., & Harris, J. (2007). *Competing on analytics*. Boston, MA: Harvard Business School. Davenport, T., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*. Retrieved from https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/

Davenport, T., Harris, J., & Morison, R. (2010) *Analytics at work*. Boston, MA: Harvard Business School.

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems, 55*(1), 412-421.

Deng, X., & Chi, L. (2012). Understanding postadoptive behaviors in information systems use: A longitudinal analysis of system use problems in the business intelligence context. *Journal of Management Information Systems*, *29*(3), 291-326.

Eckerson, W. (2011). Analytic architectures: Approaches to supporting analytics users and workloads. *BeyeNetwork.* Retrieved from http://www.b-eye-network.com/blogs/eckerson/archives/2011/10/

Elbashir, M. Z., Collier, P., Sutton, S. G., Davern, M. J., & Leech, S. (2013). Enhancing the business value of business intelligence: The role of shared knowledge and assimilation. *Journal of Information Systems*, *27*(2), 87-105.

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*, *21*(1), 70-86.

Evans, C. (2013). *Big data storage: Hadoop storage basics*. Retrieved from http://www.computerweekly.com/feature/Big-data-storage-Hadoop-storage-basics

Feigenbaum, L. (2014). *5 steps to turning big data into smart data*. Retrieved from http://tdwi.org/articles/2014/07/15/turning-big-data-into-smart-data-2.aspx

Ferguson, J., Soekijad, M., Huysman, M., & Vaast, E. (2013). Blogging for ICT4D: Reflecting and engaging with peers to build development discourse**.** *Information Systems Journal*, *23*(4), 307-328.

Gefen, D., Benbasat, I., & Pavlou, P. (2008). A research agenda for trust in online environments. *Journal of Management Information Systems*, *24*(4), 275-286.

Gillon, K., Aral, S., & Lin, C. (2014). Business analytics: Radical shift or incremental change? *Communications of the Association for Information Systems*, *34*, 287-296.

Goes, P. (2014). Editor's comments: Big data and IS research. *MIS Quarterly*, *38*(3), iii-viii.

Gorry, G. A., & Scott Morton, M. (1971). A framework for management information systems. *Sloan Management Review*. Retrieved from http://cpe.njit.edu/dlnotes/MIS645/Frame_Management_Info.pdf

Grover, P., & John, R. (2015). A virtuous cycle for top-line growth. *McKinsey Quarterly*. Retrieved from http://www.mckinsey.com/insights/marketing_sales/a_virtuous_cycle_for_top-line_growth

Herman, M., & Delurey, M. (2013). Turning big data into opportunity: The data lake. Retrieved from http://www.boozallen.com/content/dam/boozallen/media/file/VP_MakingTheCloudReign.pdf

Herman, M., Rivera, S., Mills, S., Sullivan, J., Guerra, P., Cosmas, A., Farris, D., Kohlwey, E., Yacci, P., Keller, B., Kherlopian, A., & Kim, M. (2013). The field guide to data science. Retrieved from http://www.boozallen.com/media/file/The-Field-Guide-to-Data-Science.pdf

Hey, T. (2010). The next scientific revolution. *Harvard Business Review*, *88*(11), 56-63.

Holsapple, C., Lee-Post, A., & Patkath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, *64*, 130-141.

Hosack, B., Hall, D., Paradice, D., & Courtney, J. (2012). A look toward the future: Decision support systems research is alive and well. *Journal of the Association for Information Systems*, *13*(5), 315- 340.

Hu, T., Poston, R., & Kettinger, W. (2011). Nonadopters of online social network services: Is it easy to have fun yet? *Communications of the Association for Information Systems*, *29*, 441-458.

IBM. (2014). *The four V's of big data*. Retrieved from http://www.ibmbigdatahub.com/infographic/four-vs-big-data

IEEE. (2014). Homepage. Retrieved from http://www.ieee.com

Inmon, B. (2014a). *Classification of corporate data*. Retrieved from http://www.b-eye-network.com/view/17251

Inmon, B. (2014b). *Characteristics of corporate data types*. Retrieved from http://www.b-eye-network.com/view/17253

Inmon, B. (2014c). *Deriving business value from unstructured nonrepetitive data.* Retrieved from http://www.b-eye-network.com/view/17254

Kane, G., Alavi, M., Labianca, G., & Borgatti, S. (2014). What's different about social media networks? A framework and research agenda. *MIS Quarterly*, *38*(1), 275-304.

Kobielus, J. (2012). *Data scientists: How big is your big data sandbox*? Retrieved from http://www.ibmbigdatahub.com/blog/data-scientists-how-big-your-big-data-sandbox

Koch, H., Leidner, D., & Gonzalez, E. (2013). Digitally enabling social networks: Resolving IT–culture conflict. *Information Systems Journal*, *23*(6), 501-523.

Kulkarni, R. (2013). *Transforming the data deluge into data-driven insights: Analytics that drive business*. Keynote speech presented at the 44th Annual Decision Sciences Institute Meeting, Baltimore, MD.

Lacity, M. C., Solomon, S., Yan, A., & Willcocks, L. P. (2011). Business process outsourcing studies: A critical review and research directions. *Journal of Information Technology*, *26*(4), 221-258.

Laney, D. (2001). 3-D data management: Controlling data volume, velocity and variety. *Gartner*. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling- Data-Volume-Velocity-and-Variety.pdf

Laney, D., & Kart, L. (2012). Emerging role of the data scientist and the art of data science. *Gartner*.

Laskowski, N. (2012). *Data sandboxes help analysts dig deep into corporate info*. Retrieved from http://searchbusinessanalytics.techtarget.com/feature/Data-sandboxes-help-analysts-dig-deep-into- corporate-info

Lau, R. Y. K., Li, C., & Liao, S. S. Y. (2014). Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis. *Decision Support Systems*, *65*, 80-94.

Li, X., Hsieh, J., & Rai, A. (2013). Motivational differences across post-acceptance information system usage behaviors: An investigation in the business intelligence systems context. *Information Systems Research*, *24*(3), 659-682.

Luckham, D. (2002). *The power of events: An introduction to complex event processing in distributed enterprise systems.* New York, NY: Addison-Wesley Professional.

Lycett, M. (2013). "Datafication": Making sense of (big) data in a complex world *European Journal of Information Systems*, *22*(4), 381-386.

Manikya, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60-66.

McGrath, K., & Elbanna, A. (2012). Exploring the democratic potential of online social networking: The scope and limitations of e-participation. *Communications of the Association for Information Systems, 30*, 239-254.

Mohanty, S. (2014). Analytics: *Detect the expected—discover the unexpected*. Retrieved from http://www.b-eye-network.com/print/17270

Narayanan, R. (2013). *Has Hadoop entered the mainstream?* Retrieved from http://www.informationmanagement.com/news/has-hadoop-entered-the-mainstream-10024948-1.html?zkprintable=1andnopagination=1

Oh, O., Agrawal, M., & Rao, H. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly*, *37*(2), 407-426.

O'Neil, K. (2012). *Big data: Governance is the critical starting point.* Retrieved from http://www.b-eye-network.com/blogs/oneal/archives/2012/11/big_data_govern.php

O'Rourke, K. (2013). *Best practices in BI: Let your business processes drive successful BI analytics*.
Retrieved from http://www.information-management.com/news/best-practices-in-bi-10025015-1.html?zkPrintable=1andnopagination=1

Pan, S., Pan, G., & Leidner, D. (2012). Crisis response information networks. *Journal of the Association for Information Systems*, *13*(1), 31-56.

Parameswaran, M., & Whinston, A. (2007). Social computing: An overview. *Communications of the Association for Information Systems*, *19*, 762-780.

Porter, R. (2014). *There's no such thing as unstructured data*. Retrieved from http://tdwi.org/Articles/2014/06/24/No-Such-Thing-as-Unstructured-Data.aspx?

Potter, B. (2014). *How dashboards have grown up*. Retrieved from http://tdwi.org/Articles/2014/03/04/Dashboards-Grown-Up.aspx?Page=1

Powell, R. (2013). *The convergence of BI, big data, analytics and mobile technologies: A spotlight Q&A with Michael Corcoran and Jake Freivald of Information Builders*. Retrieved from http://www.b-eye- network.com/view/17066

Rivera, J. (2014). By 2016, 25 percent of large global companies will have adopted big data analytics for atleast one security or fraud detection use case. *Gartner*. Retrieved from http://www.gartner.com/newsroom/id/2663015

Roski, J., Bo-Linn, G., & Andrews, T. (2014). Creating value in health care through big data: Opportunities and policy implications. *Health Affairs*, *33*(7), 1115-1122.

Russom, P. (2011). *Big data analytics*. Retrieved from http://teradatauniversitynetwork.com

Russom, P. (2014). *Getting started with big data*. Retrieved from http://tdwi.org/Articles/2014/08/26/Getting-Started-Big-Data.aspx?Page=1

Schlegel, K. (2014). Hype cycle for business intelligence and analytics. *Gartner*. Retrieved from https://www.gartner.com/doc/2813223/hype-cycle-business-intelligence-analytics

Shmueli, G., & Koppius, O. (2011). Predictive analytics in information systems research. *MIS Quarterly*, *35*(3), 553-573.

Sowa, J. F., & Zachman, J. A. (1992). Extending and formalizing the framework for information systems architecture. *IBM Systems Journal*, *31*(3), 590-616.

Sprague, R. (1980). A framework for the development of decision support systems. *MIS Quarterly*, *4*(4), 1- 26.

Subramanian, S. (2013). *ETL to ELT conversion testing in data warehouse engagements.* Retrieved from http://www.information-management.com/news/etl-to-elt-conversion-testing-in-data-warehouse-engagements-100238381.html?zkPrintable=1&nopagination=1

Sundararajan, A., Provost, F., Oestreicher-Singer, G., & Aral, S. (2013). Research commentary-information in digital, economic and social networks. *Information Systems Research*, *24*(4), 883- 905.

Swoyer, S. (2010). *What's essential—and what's not—in big data analytics*. Retrieved from http://tdwi.org/Articles/2010/08/18/Big-Data-Analytics-Essentials.aspx?s=BITW_180810andadmgarea=BITWandPage=1

Swoyer, S. (2013). Big data and meaningful storage metrics. Retrieved from http://tdwi.org/Articles/2013/06/18/Big-Data-Storage-Metrics.aspx?Page=1

TDWI. (2013). *Q&A: The human side of big data*. Retrieved from http://tdwi.org/articles/2013/04/02/Human-Side-of-Big-Data.aspx

Thaler, R., & Tucker, W. (2013). Smarter information, smarter consumers. *Harvard Business Review*, *91*(1), 45-54.

Tufte, E. (1997). *Visual explanations: Images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press.

van Valkenhoef, G., Tervonen, T., Zwinkels, T., de Brock, B., & Hillege, B. H. (2013). ADDIS: A decision support system for evidence-based medicine. *Decision Support Systems*, *55*(2), 459-475.

Violino, B. (2014). *Confusion surfaces about the data lake*. Retrieved from http://www.information-management.com/news/confusion-surfaces-about-the-data-lake-10025925-1.html

vom Brocke, J., Debortoli, S., Müller, O., & Reuter, N. (2014). How in-memory technology can create business value: Insights from the Hilti case. *Communications of the Association for Information Systems, 34*, 151-168.

Wakefield, R. (2013). The influence of user affect in online information disclosure. *Journal of Strategic Information Systems*, *22*(2), 157-174.

Wang, C., & Zhang, P. (2012). The evolution of social commerce: The people, management, technology and information dimensions. *Communications of the Association for Information Systems, 31*, 105- 127.

Watson, H. (2009). Tutorial: Business intelligence—past, present and future. *Communications of the Association for Information Systems*, *34*, 1247-1268.

Watson, H. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *25*, 487-510

Watson, H. J., Rainer, R. K., Jr., & Koh, C. E. (1991). Executive information systems: A framework for development and a survey of current practices. *MIS Quarterly*, 15(1), 13-30.

Watson, H., & Wixom, B. (2007). The current state of business intelligence. *Computer*, *40*(9), 96-99. Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L. S., Kulkarni, U., Mooney, J. G., Phillips-Wren, G., & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems, 34*, 1-13.

Wixom, B., Ariyachandra, T., Goul, M., Gray, P., Kulkarni, U., & Phillips-Wren, G. (2011). The current state of business intelligence in academia: Results from the BI congress. *Communications of the Association for Information Systems*, *29*, 300-312.

World Wide Web Consortium. (2014). *Semantic Web*. Retrieved from http://www.w3.org/standards/semanticweb/

Zachman, J. A. (1987). A framework for information systems architecture. *IBM Systems Journal, 26*(3), 276-292.

Zaslavsky, A. (2014). *Big data and the Internet of things*. Presentation at IFIP WG8.3 International Conference, Paris, France.

Zheng, Z., Fader, P., & Padmanabhan, B. (2012). From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data. *Information Systems Research*, *23*(3), 698-720.

About the Authors

Gloria Phillips-Wren is Professor and Chair of Information Systems and Operations Management at Loyola University Maryland. She is co-Editor-in-Chief of *Intelligent Decision Technologies* (IDT), Past Chair of SIGDSS (now SIGDSA), Secretary of IFIP WG8.3 DSS, and leader of a focus group for KES International. She serves as an Associate Editor of the *Journal of Decision Support*. She was co-chair of Business Intelligence Congresses II and III held in 2010 and 2012. She joined academia after a career in research and development for the federal government in advanced ballistic concepts. Her main areas of research are in decision making, decision support systems, analytics, big data, business intelligence, data mining, intelligent systems, and healthcare IT. She is co-author of a book on advanced propulsion, co- editor of several books on intelligent decision technologies, and has published in *the European Journal of Operations Research, Communications of the Association for Information Systems, Omega, IEEE Transactions on Magnetics,* and others.

Lakshmi Iyer is Associate Professor and Director of Graduate Programs in the Information Systems and Supply Chain Management Department at the University of North Carolina at Greensboro. Her research interests are in the areas of business intelligence, knowledge management, emerging technologies and its organizational impact, and gender gap in computing. Her research work has been published in *Journal of Association for Information Systems, European Journal of Information Systems, Communications of the Association for Information Systems, Communications of the ACM, Decision Support Systems, eService Journal, Journal of Electronic Commerce Research, International Journal of Business Intelligence Research, Information Systems Management,* and others. She is a Board member of Teradata University Network and recent past Chair of the Special Interest Group in Decision Support and Analytics (SIGDSA).

Uday Kulkarni is Associate Professor of Information Systems and founding director of the Masters of Science of Business Analytics (MSBA) degree program at Arizona State University. His research interests span the areas of business intelligence capability - how organizational factors act as its antecedents, knowledge management capability assessment of organizations, and ICT adoption in micro-finance institutions. His publications have appeared in leading IEEE and MIS journals such as *IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Software Engineering, Journal of MIS, Decision Support Systems, Communications of the Association for Information Systems*, and *European Journal of Operations Research*. He serves as Associate Editor for *the International Journal of Knowledge Management* and served as chair of SIGDSA, an international special interest group in decision support and analytics. He has developed graduate-level coursework in Business Intelligence and IT Strategy and has been recognized for teaching excellence on numerous occasions.

Thilini Ariyachandra is Associate Professor of Management Information Systems in the Williams College of Business at Xavier University in Cincinnati. Her research is focused on business intelligence, analytics and Big Data frameworks, design, implementation and success. She has published in journals such as the *Communications of the ACM, Decision Support Systems*, *Communications of the AIS*, *Business Intelligence Journal*, and *International Journal of Business Intelligence Research*. She was one of the youngest recipients of the O'Conor Professorship at Xavier University in recognition of her research and scholarship. She collaborates extensively with industry and academia to enhance curriculum, pedagogy and research in the field through survey work, consultancy and advisory roles. She is actively involved in the organization of various workshops and events that encourage BI and analytics research and teaching within the field. Currently she is the Program Chair for the Special Interest Group for Decision Support and Analytics (SIGDSA) for the Association for Information Systems.