Communications of the Association for Information Systems

Volume 37 Article 39

10-2015

Big Data and Analytics: Issues, Solutions, and ROI

J. P. Shim
Georgia State University, jpshim@gsu.edu

Aaron M. French
University of New Mexico

Chengqi Guo
James Madison University

Joey Jablonski
Dell, Inc.

Communications of the Association for Information Systems

Panel Report ISSN: 1529-3181

Big Data and Analytics: Issues, Solutions, and ROI

J. P. Shim
Department of Computer Information Systems, Georgia State University
jpshim@gsu.edu

Aaron M. French
Department of Management Information Systems, University of New Mexico

Chengqi Guo
Department of Computer Information Systems & Business Analytics, James Madison University

Joey Jablonski
Director of Product Management for Analytics & Big Data at Dell, Inc.

Abstract:

Recently, the topic of big data and analytics has received renewed attention from academia and practitioners. There has been an increase in demand for skills in big data and analytics due to the increasing speed, variety, and volume of information. Several research reports have shown that big data and analytics remain top priority for CIOs. A recent study shows how a company accurately predicted a teen girl's pregnancy via the company's big data algorithm. However, there are dark sides to big data and analytics. A panel discussion addressed topics concerning how companies ensure that big data projects clearly define measurable goals up front, methods that companies use to ensure maximum return and most effectively, and ways that companies evolve culture, processes, and technology to simultaneously maximize return. Most companies are looking at how they can effectively manage their business more through using their data assets. Companies today target an average return of $3.50 dollars for every dollar spent on big data projects. However, most are only returning a fraction of that today, which leaves room for improvement and the possibility that organizations will push back against new analytic technologies. In this paper, we cover these topics that a panel of researchers at AMCIS 2014 in Savannah, GA, discussed.

Keywords: Big Data, Analytics, Structured and Ill-Structured Data, Specific Issues, Big Data Projects, Return on Investment (ROI).

This manuscript underwent editorial review. It was received 05/25/2015 and was with the authors 1 month for 1 revision. Matti Rossi served as Associate Editor.

1 Introduction

Big data and business analytics are popular topics gaining significant attention from practitioners and scholars alike (Chen, Chiang, & Storey, 2012). A recent paper has showed that 2.5 exabytes (2.5 billion gigabytes) of data is created every day (McAfee& Brynjolfsson, 2012). A great number of firms are looking at how they manage their business more effectively through using their data assets. Companies today target a return of $3.50 for every dollar spent on big data and analytics projects. However, many projects only yield returns of $0.50, which leaves room for improvement and the possibility that organizations will push back against new analytic technologies. Analytics and Big data can be a revolutionizing force for numerous industries. Though website analytics itself has been a topic of discussion for over a decade, 90 percent of the world's data has been created in the past several years. This significant increase is due to the increased sophistication of mobile technology and social media networks. Further, as these mobile computing devices expand globally, the volume of this digital content will only increase. Various platforms allow individuals to be tracked along with real-time updates. The massive amount of customer data that is generated and collected on a daily basis holds the potential to drive profits. Customers are more educated than ever before and currently have a vast amount of information readily available. As a result, businesses are increasingly reliant on big data and analytics.

Big data can have a puzzling element because not all data is easily discernible. One can categorize data into structured and ill-structured information. Structured data mirrors that of information obtained from a direct transaction, while ill-structured data comes in a form often obtained from social media, such as Twitter feeds, "retweets", Facebook posts, and "likes". Previous information systems used by organizations, such as data warehouses and ERP systems, typically processed structured for reporting and decision making, while semi-structure and ill-structured data was left behind (Negash, 2004). With the increased sophistication of data storage and processing tools, this data that was once unusable is providing valuation information and resulting in new business intelligence. As social media platform users grow, so will the 3Vs (volume, variety, and velocity) of this information. These data components offer key performance indicators in their own way. Knowing one's customers, their behaviors, and markets and altering quickly to accommodate changes are imperative to adapting to instantaneous issues that arise and to ensure customer satisfaction and increase profitability.

The future of all industries relies heavily on firms positively leveraging big data and analytics. In our digital society, one can purchase goods and services from a remote location. The way in which firms interact with customers will have to incorporate specific tailoring of their products, platforms, and internal management. Such firms can invest in big data, analytics, and other technologies to track the customers, provide easily accessible platforms on various devices, store this gathered data, and adapt business practices to create experiences specifically to accommodate various micro-market segments. Leveraging big data can also increase our ability to research problems at the macro level (society) by evaluating large amounts of comprehensive data at the micro level (individuals) by using new tools equipped for handling semi-structured and unstructured data (Agarwal & Dhar, 2014).

In the past years, there has been an explosive growth of user data in terms of volume because of social media's proliferation. Such a data avalanche poses serious challenges and significant competitive advantages to business entities that rely on distributed or cloud computing to handle user requests and respond rapidly. "Perception is reality" seems to have become the slogan of big data practices, which differentiates them from traditional database sectors (e.g., relational database) where consistency requirement is ubiquitous.

However, there are dark sides to big data. While big data has received a lot of attention for its potential, we must face several of its challenges. For instance, privacy is a major concern in terms of big data. The massive amounts of data that organizations collect has led to the development of digital dossiers at a level of detail that we have never seen before. One can use these digital dossiers to uncover intimate details about an individual such as sexuality, menstrual cycles, and whether a woman is pregnant or not. It raises ethical concerns whether or not companies have a right to mine for such personal information that can be used for marketing purposes. Additional concerns could be the misuse of data, which can result in misleading truths or the introduction of the digital divide 2.0 (i.e., difference between those who have access to big data and those who do not). In addition, one of the most important issues with regards to big data is whether companies are seeing a return on their investments (ROI) in it.

This paper proceeds as follows: in Section 2, we discuss the current status of big data and analytics. In Section 3, we describe how to interpret big data's benefits. In Section 4, we discuss the dark side of big data, including big data cases, digital dossiers, ethics and privacy, and the digital divide 2.0. In Section 5, we discuss maximizing the return for big data projects, which includes improving ROI for big data projects and big data project fundamentals. Finally, in Section 6, we concludes with recommendations for future of big data and analytics.

2 Current Status of Big Data and Analytics

2.1 Timeline and Data Storage with Complexity

The big data timeline illustrates how the first big data problem began in 1890 when the U.S. Government Census was carried out. In 1965, the first big data center was established when the U.S. Government needed a place to store 742 million tax returns and 175 million sets of fingerprints. The birth of the World Wide Web in 1989 was a milestone for the big data that was housed and accessed on the Internet on a daily basis. Big data was defined around 1997-2001. By 2002, the world contained in excess of three billion documents with an estimated 80 percent of it being ill-structured and predictions of the number of ill-structured documents doubling every eight months (Negash, 2004). In 2004, big data tools, such as Hadoop, helped individuals and firms to actually understand big data and realize its potential (Alteryx.com)

Big data is often characterized by volume, velocity, and variety (the three Vs) (Gillon, Aral, Lin, Mithas, & Zozulia, 2014; Goes, 2014; Hashem et al., 2015; Lycett, 2013; McAfee & Brynjolfsson, 2012; Wixom, Ariyachandra, Douglas, Goul, & Gupta, 2014). Researchers have also extended this model to include veracity (Gillon et al., 2014; Goes, 2014) and value (to make five Vs) (Hashem et al., 2015; Lycett et al., 2013). In detail:

- Volume refers to the amount of data being collected ("how much"). Data sets today can extend beyond exabytes of data.
- Variety refers to the type of data being collected ("in what forms"). Sensor data, mobile technology, and social networking have increased the types of data to include text, images, videos, location information, and more.
- Velocity refers to the speed at which data is generated ("how fast"). Ubiquitous technologies have generated continues flows of data at rates well beyond anything seen in history to date.
- Veracity refers to the data's integrity ("how reliable"); that is, its accuracy, truthfulness, precision, and other such factors.
- Value refers to the ability to use the data to extract information of value to the organization ("how valuable") that will result in business intelligence and assist decision making.

Following the hierarchy of data → information → knowledge → intelligence (Goes, 2014), we can classify the five Vs of big data into two subgroups: big data characteristics and big data processing. One can argue that we do not know data's validity (veracity) or value until it has been processed. After data has been processed, we can extract knowledge and intelligence from the information produced. Figure 1 displays the characteristics of big data and its processing.

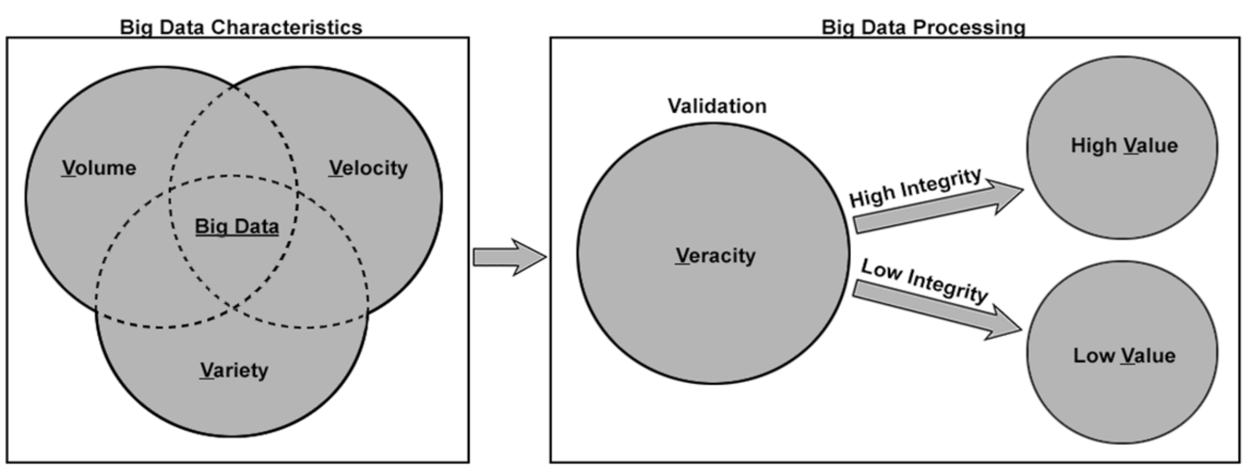

Figure 1. Characteristics and Processing of Big Data

By definition, big data is characterized by the large volumes of various types of data generated at a high rate. Just as the name implies, big data is a lot of data. We cannot know the data's veracity or value until it has been processed. However, big data with low veracity and low value is still big data. The variety of data ranges from structured to unstructured (Hashem et al., 2015). The continued growth of illstructured data increases the difficulty of processing and extracting useful information, which results in text mining and image processing becoming an important new frontier of research (Agarwal & Dhar, 2014). New data classification and analysis tools such as MapReduce and Hadoop have been developed to process and manage these large repositories of data that are too large for traditional storage methods (i.e., relational database) to handle (Ferrera, De Prado, Palacios, Fernandez-Marquez, & Serugendo, 2013). The open source solution Hadoop has led to a plethora of big data processing tools such as Sawzall, FlumeJava, Pig, Hive, Jaql, and Cascading that all contain specialized features ranging from SQL-style data manipulation to Java-based APIs serving a wide range of users and skills (Ferrera et al., 2013).

Big data comes in several types; 1) Web and social media data, including clickstream and interaction data from social media; 2) machine-to-machine (M2M) data, including readings from sensors, meters, and other devices; 3) big transaction data, including healthcare claims, telecommunications call detail rerecords, and utility billing records; 4) biometric data, including fingerprints, genetics, handwriting, retinal scans, and similar types of data; and 5) human-generated data, including vast quantities of unstructured and semi-structured data such as call center agents' notes, voice recordings, emails, paper documents, surveys, and electronic medical records (Gartner, 2013). Figure 2 displays various systems along with information used and the data analyzed as the variety of data and complexity increases compared with the amount of data being collected (i.e., petabytes, exabytes).

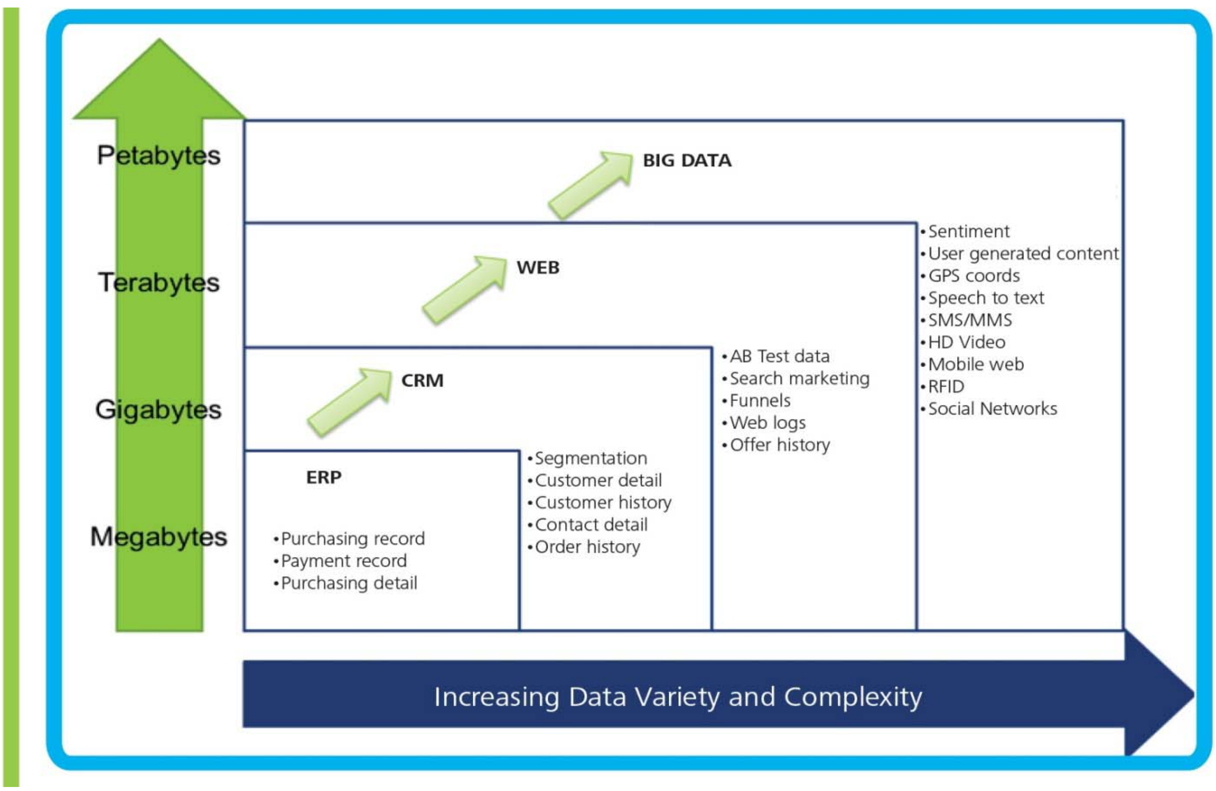

Figure 2. Data Storages and Increasing Data Variety and Complexity

2.2 Real-world Applications

There are numerous real-world big data applications in major firms. Among these, several interesting cases have been illustrated (Davenport & Dyche, 2013): UPS, a global leader in logistics, has been using big data for over the past two decades. "UPS is no stranger to big data, having begun to capture and track a variety of package movements and transactions as early as the 1980s… Much of its recently acquired big data comes from telematics sensors" (Davenport & Dyche, 2013). In particular, UPS's on-road integrated optimization and navigation (ORION) and DIAD (handheld device) are a legacy to big data at UPS.

Caesars Entertainment, formerly known as Harrah's, established itself as a leader in big data and analytics. For instance, Caesars has data about its customers from its "total rewards loyalty program, web clickstreams, and from real-time play in slot machines" (Davenport & Dyche, 2013). Like most other entertainment industry, Caesars analyze mobile data for spotting service (i.e., use of real-time or near real-time trend spotting with visualization tools).

Prior to the development of big data tools, many companies stored data that provided little to no information. It was reported that one telecommunications provider contained up to 10,000 phone conversations per day with customers but were unable to evaluate them; they could only measure the end result displaying if a phone plan was changed or not (Negash, 2004). With new analytical tools to evaluate this qualitative information, managers would not only know the results of a conversation but also the underlying data the led to the positive result.

3 Interpreting Big Data Benefits

In general, big data refers to all the data/information a business has. Companies, either profit or non-profit, have faced the fundamental question "how do I make this work?". They need data/information that is relevant to ad hoc and long-term decision making so that organizations are built for data-driven decision making, which is important because decisions made without data-driven answers will likely fail. Big data has been around for many years and there are two aspects in terms of how the public perceives it. First, big data refers to the large amount of data assets that information workers store and manage. Second, big data means the techniques used to process and analyze these data assets in real-time for valuable information. In the late 1970s, when point of sale (POS) system was introduced in supermarkets and supercenters, there was a huge growth of consumer data, which essentially led to major paradigm shifts in manufacturing companies' marketing and advertising strategies (see Figure 3) (Fulgoni, 2013). POS data were obtained and analyzed by constantly evolving software applications (e.g., data warehousing) to generate business value. In a nutshell, big data's most fundamental advantage lies in its capability to build the foundation for information workers who identify interesting patterns among data at an extreme scale in terms of volumes and formats. Doing the same is difficult and not cost-effective for conventional database technologies (e.g., relational database management systems (RDBMS)).

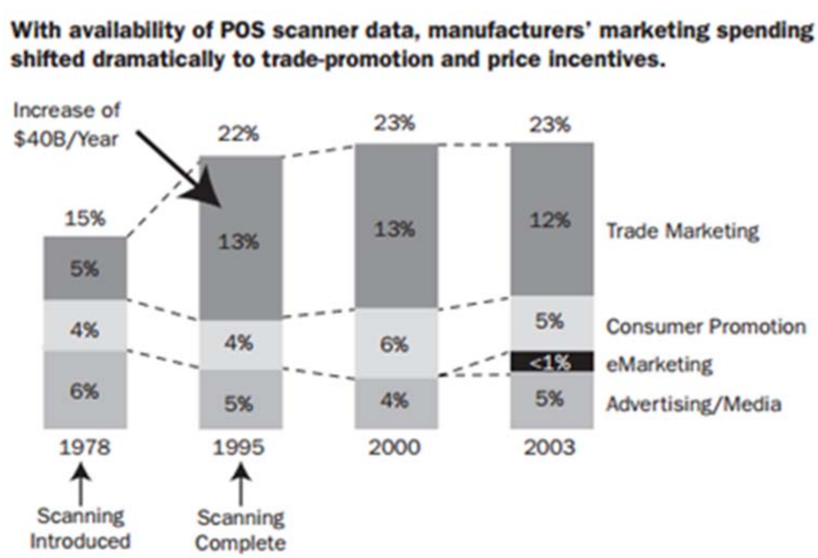

Figure 3. Paradigm Shift of Marketing due to POS Data Increase (Fulgoni, 2013)

In recent decades, the world economy has become increasingly dependent on knowledge/business intelligence and well-informed decisions (Kabir & Carayannis, 2013). This trend has stimulated the rapid development of computing technologies for collecting and analyzing data, which generate the data tsunami we witness today. Correspondingly, the challenge has transformed from not having enough data to dealing with too much data. Although the word "big" stresses the volume characteristic, big data requires more than just volume management. In fact, in Mark Beyer's (2011) three Vs model (volume, velocity, and variety), big data displays conspicuous advantages over traditional databases (Stephens, 2013; Vriens & Brazell, 2013).

3.1 Perception is Reality: The Choice between Consistency and Speed

"Reality is merely an illusion, albeit a very persistent one." —Albert Einstein

Customers demand convenient and customized data visualization that helps them to perform daily tasks more efficiently. To them, everything happens at their fingertips and that is all that matters. The internal database structure, storage mechanism, and cross validations of query accuracy are invisible to end users and, therefore, all considerations should give way to data availability and query response speed. This notion, therefore, introduces a give-and-take scenario: should we sacrifice data consistency or reliability to obtain a boost in query speeds? For RDBMS developers, this choice is difficult because the atomic consistency isolation durability (ACID) requirements guarantee reliability in a rigorous way. Plus, the benefits of ACID have been touted by RDMBS manufacturers for years. For big data engineers, they have more flexibility when implementing the solutions thanks to the "eventually consistent" technique. The following list, although incomplete, describes important advantages of big data's schema-less design:

- Not ACID compliant: non-relational has been proven to be valuable in contemporary Web-based database environments such as MongoDB and HBase. According to Oracle, any database that isn't RDBMS upholds schema-less structures and is generally relaxed on ACID transactions. The schema-free model promises high availability and support for large data sets in distributed environments.
- Horizontally scalable: big data has progressed hand-in-hand with cloud computing. The Hadoop file system first maps the data and then reduces it based on nodes to distribute tasks in a large network of servers. A major advantage of this feature, when combined with schema-less data structures, is superior performance without committing to a significant system upgrade. Such an advantage also involves large volumes of structured/semi-structured/unstructured data and, thus, answers the challenge of working with a large variety of data.
- Eventually consistent: eventually consistent, or lazy consensus, is a popular workaround to resolve the conflict between availability and consistency in big data solutions. With shared data (e.g., shards) and distributed systems, big data platforms sacrifice consistency rigor to some degree because they do not prioritize the demand of serialized operations. In return, the speed of data generation, query performance, and availability are significantly improved. Once a user request has been recorded, it takes effect in one of the data replicas and, eventually, such updates will be validated across all nodes. Some big data flavors prefer consistency over availability and, according to the CAP theorem, no database solution can guarantee consistency, availability, and partition tolerance together. Therefore, one should evaluate the business requirements to plan for two and make the best of the third.

4 The Dark Side of Big Data

While big data has received a lot of attention for its potential, we need to face some of its challenges. Privacy is one major concern. With the amount of data and rate at which it is being created continue to increase, organizations are able to create comprehensive digital dossiers for each customer with details beyond anything seen before. These digital dossiers can be used reveal intimate details about an individual such as sexuality, menstrual cycles, and whether a woman is pregnant or not. It raises ethical concerns about whether or not companies have a right to mine for such personal information that can be used for marketing purposes. Additional concerns could be the misuse of data resulting in misleading truths or the introduction of the digital divide 2.0 (i.e., difference between those who have access to big data and those who do not).

4.1 Big Data Cases

There are several cases where big data and the information mined from this data have been intrusive and even harmful to individuals. The retailer Target had a situation where their data mining proved to be an invasion of privacy into the life of a minor and informed her father of her untimely pregnancy prematurely

(Hill, 2012). Through data mining customers' shopping habits and trends, Target was able to identify 25 products that they classified as a "pregnancy prediction" for their shoppers. When the company identified customers as potentially being pregnant, it flagged the customer and sent coupons based on their pregnancy score. When a local teen in Minneapolis received coupons for baby products, her father was outraged and complained to the store. It turned out the teen was pregnant, but she did not intend her father to find out that way.

In January 2014, a family in Chicago received coupons from OfficeMax identifying their daughter as deceased from a car crash one year earlier (Merrick, 2014). Through data collections about their customers and other data sources, OfficeMax knew when their daughter had died and how. The coupons being sent were blamed on a data error but the collection of that information and its potential use for marketing raises serious ethical issues. The amount of data collected by these companies is small compared to the vast amount of user generated data collected through search engines and social media.

The amount of data being collected and stored by companies such as Google, Facebook and Twitter is beyond anything that has been seen to date. These companies received scrutiny for using public and private messages in their data mining for marketing purposes (Compeau, Haggerty, & Fraiha, 2011). Furthermore, Facebook has received complaints about privacy issues for using its technology called Beacon, which tracks its users' activity across the Web. Beacon collects user IP addresses from partner sites to match with IP addresses used on Facebook. Information provided by partner sites can be matched to a Facebook accounts, increasing the information known about individuals and the ability to target them for ads (Martin, 2010).

4.2 Digital Dossier

Companies have been collecting information for many years about their customers, which they use to develop dossiers that they can use for marketing. The use of digital technology has improved companies' ability to understand their customers through developing digital dossiers. With the advent of social media and the amount of information social networking users are willing to provide, these digital dossiers have reached a level of detail never before seen in history. While this information provides a powerful tool for marketers to understand their customers and relate to them on a micro-level, they need to consider ethical and privacy issues.

As the examples in Section 4.1 show, companies can effect negative outcomes by collecting big data. With Facebook's purchase of WhatsApp, the amount of information contained in the digital dossier of Facebook's user base increased exponentially. Facebook has made a strong attempt to collect their users' phone numbers in ways ranging from recommended security features and password recovery systems to their development of Facebook Messenger, which requires a phone number to register. WhatsApp messenger requires a valid phone number to register and is used to exchange private messages among its users. After being purchased by Facebook for USD$19 billion, Facebook obtained the rights to the company and all its information. By matching phone numbers from WhatsApp to phone numbers it had obtained itself, Facebook could link the accounts of both systems to create a more complete window into a person's life. But the privacy threats extend beyond textual messages posted both publically and privately.

If a picture is worth a thousand words, then one can imagine how many words can be generated by social networking users who have posted thousands of pictures to their accounts. With new technology such as facial recognition software, images shared can tell more of a story than the captions posted to describe them. New technology described by Facebook called DeepFace accurately has identified matching individuals 97.25 percent of the time (O'Toole, 2014). This type of technology could match individuals from the background of pictures taken by others they don't know. For instance, if person A was taking a vacation and appeared in the background of a picture taken by person B, then facial recognition software could identify person A and link them to a location identifying where they were traveling. Google chose not to include facial recognition software in its Google Glasses due to privacy concerns raised by privacy advocates such as the American Civil Liberties Union (Bawab, 2014). This technology could allow people wearing Google Glasses to identify others while walking down the street and conduct Google searches for more information on them.

4.3 Ethics and Privacy

Morris (2002) has reported that the top five sources for big data include social networks, social influencers, activity-generated data, software-as-a-service (SaaS) and cloud applications, and public data. Many individuals use these services as part of their daily routine without intending to provide their personal data for organizations to exploit. Social networking in particular contains information viewed by the user as both private and public. Wall posts are viewed as public information, while personal messages sent through these systems are deemed private and personal. However, both private and public data are used in big data for business intelligence and analytics to increase marketing capabilities, which many individuals view as invading their privacy. Without formal laws dictating how big data can be used, we must rely on ethical considerations to find a solution. It would be unethical for a store manager at a grocery store to listen to their customers' private conversations to improve marketing capabilities. Despite customers being in the grocery store, their conversation is still considered private and do not want to be eavesdropped on.

We can apply this same concept to social networking. There is an intended party that an individual wishes to receive private messages sent across the network. The intended party encompasses other users on the social network and not the social network itself. To use these private message for marketing and further developing digital dossiers is unethical and an invasion of privacy. Due to the recency of which the technology exists, privacy laws do not exist to protect social networking users. We must depend on the owners of big data to be responsible with the data they collect and how they use it.

4.4 Digital Divide 2.0

Since the commercialization of the Internet, there has been a digital divide between those who have access to Internet technology and information and those who do not. Great strides have been taken to bridge this gap and provide access to even the most rural areas. While there are areas of the world that still lack the ability to access information through the Internet, businesses worldwide have been able to overcome this gap and compete on equal ground. With the growth of big data due to social networking, search engines, Internet tracking, and other data sources, a new digital divide in information is being formed. Small to mid-sized companies do not have the luxury of collecting data from a half-billion users who are provided broad ranges of information on various topics. They also do not have the monetary assets needed to purchase this data to mine and develop their own business intelligence. Large companies are buying out the smaller ones (e.g., Facebook buying WhatsApp) and building industries that make it difficult for others to compete, which has developed a divide between those who have access to big data and those who do not. As this trend continues, a natural monopoly is being formed where smaller companies cannot hope to compete. To gain access to the data, one has to be willing to provide information that will contribute to the continued growth of big data by these companies.

5 Maximizing Return for Big Data Projects

5.1 Improving ROI for Big Data Projects

As more companies look to accelerate their growth and better engage with customers, big data has taken off as a set of common principals for leveraging data to better manage an organization. As with any new trend, the early adopters were very informal in how they began and executed projects: they knew that value was there but focused on execution rather than measurement. As big data has become more mainstream, organizations have looked for ways to measure the impact of investment in big data projects.

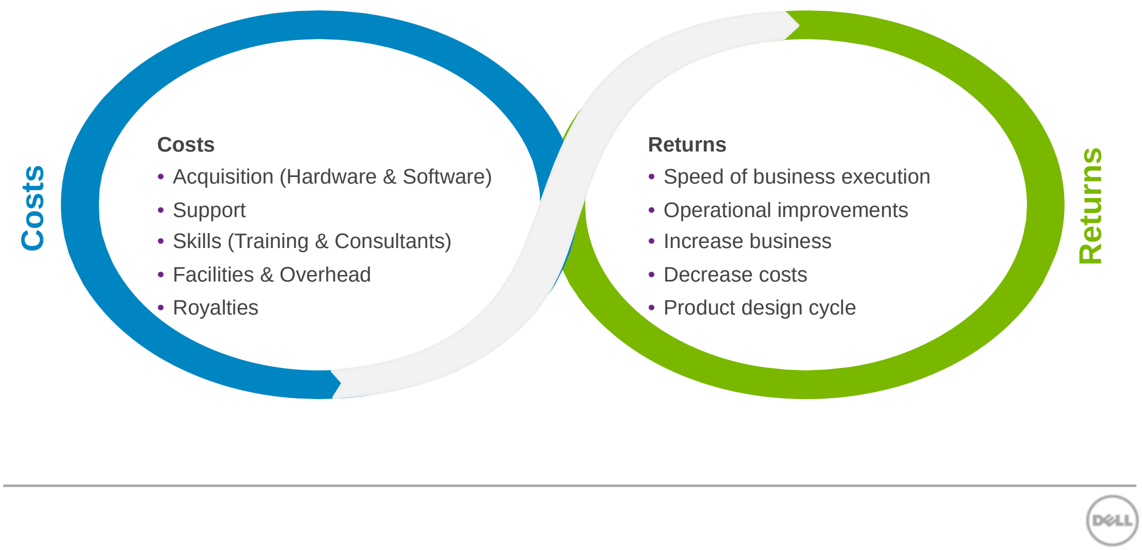

Figure 4. Big Data Return On Investment (Jablonski, 2014)

Senior leadership has said that greater than 70 percent (Bertolucci, 2013) of big data projects fail to live up to the initial hype because of unrealistic measurements or measurements that are not properly aligned with the expectations of senior leadership in an organization. Big data projects must both align with the organization's needs and have measurements that are unique to the rapidly changing technology and customer landscape. Some organizations have seen returns as high as ten times their investment (Tata Consultancy Services, 2013) for big data projects, so it is possible to successfully execute on these complex engagements.

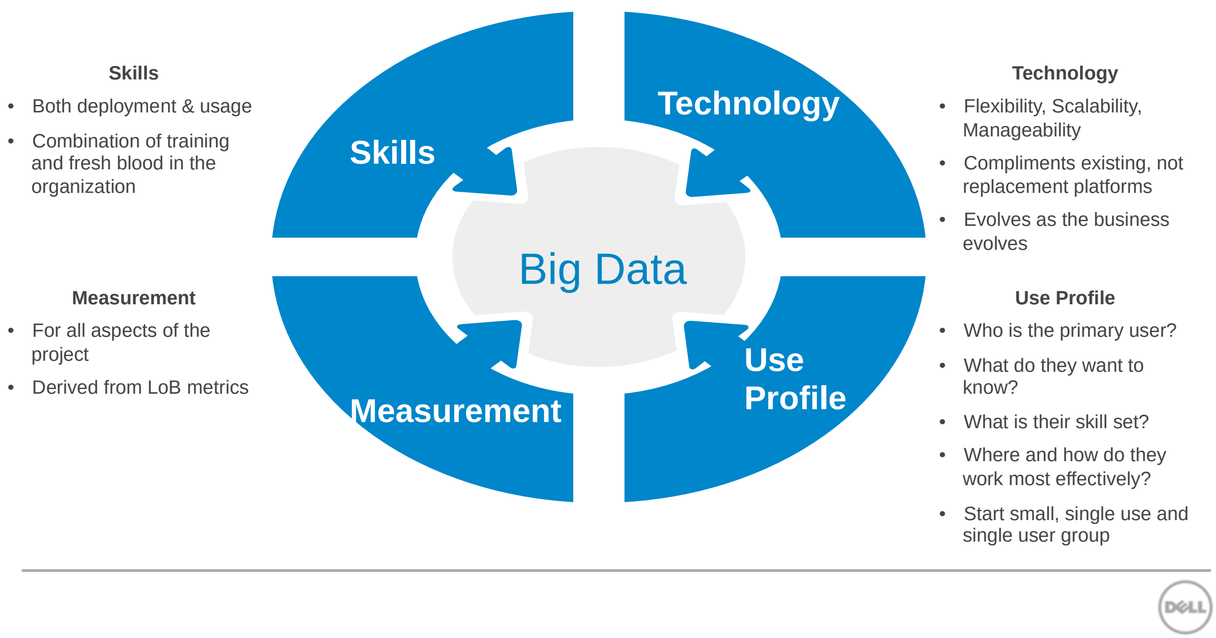

Figure 5. Big Data Project Fundamentals (Jablonski, 2014)

As organizations look to execute big data projects, they should consider four key areas of planning:

- Skill: skills are a key consideration for planning any big data project. Big data projects often introduce new technologies and methodologies into an organization, and any project should include resources to ensure proper training is provided for staff to minimize the learning curve and ensure maximum understanding of new technology.

- Measurement: all big data projects should have clear project measurements defined that are aligned with the needs and daily metrics of the sponsoring line-of-business. All senior leadership in specific lines-of-business in an organization leverage key performance indicators (KPIs) to measure the performance of the organization; the most successful big data projects will align with those KPIs and work to improve them. Since it is very hard to quantify the benefits of a big data project, this measurement area is critically important to consider.
- Technology: big data projects will combine the deployment of new technologies and integration with existing technology and work flows. Big data projects should consider all necessary requirements and plan for phased deployments of new technologies to both gather experience over time and to ensure upfront plans and designs are feasible and can be executed. Ensure the focus establishes clear lines of sight to technology requirements (Marchand & Peppard, 2013).
- Use profile: big data projects affect a variety of staff in an organization including system administrators, system architects, program managers, and business analysts. Each staff member has a different skill set and job description that should be planned for during a big data project. All big data projects should inventory the various types of users that will interact with the platform and ensure the project accounts for the various needs around interfaces, presentation, and usability.

Organizations also need to understand how each category affects designs and planning for other categories:

- Technology influences training: the training plan will be heavily influenced by the technology strategy for a big data project. Any new technologies will impact training that will need to be provided to staff, both operations and users.
- Use profile influences technology: big data projects vary on usage; some are deployed with developer-centric users in mind, while others are deployed with leadership users as the primary design target. This use profile influences the technologies that will be used to present and analyze the results of any data in the big data environment.
- Measurement impacts technology: many KPIs used by business are components of time and cost. These KPIs influence technology choices around the company's environment's sizing, the scalability of the technologies involved, and integration with existing business systems.
- Skills impact use profile: the relative skill set of users in an organization will impact the use profile because of how users will use new technology provided to them and the preferences they will have when adopting new technologies.

Properly executing big data projects requires a balance between skills, measurement, technology, and use profile to ensure maximum return. Insufficiently considering or investing in any one of them will negatively affect the final platform and staff's ability to leverage the big data investment.

6 Conclusion

In this paper, we present several distinct but interesting perspectives on current issues of big data and analytics and, specifically, on maximizing return from big data projects. As mentioned earlier, big data and analytics are an important part of today's business (i.e., analyzing customer information, categorizing customer responses, and tracking trends and patterns). Further research should explore new and growing subsets of big data and analytics, such as speech and call analytics, biometrics data analytics, and sensor data analytics.

While the term big data is relatively new, the concept of big data has existed since the late 1800s when the U.S. Government conducted the country's first census. As the capabilities of processing data continues to increase, the amounts of data being collected also increases. The term big data gained significant traction with the explosion of social media and user-generated data. The speed at which data is generated has exceeded the capabilities of current technology to process it. Text mining has created significant strides in our ability to process and understand big data, but there are still mountains of data in the form of images and video that can be explored as technology continues to evolve. However, many of this user-generated data was not intended for corporate use and raises ethical issues about whether or not it should be processed and mined for monetary purposes.

Some have argued that Facebook owns the messages sent from one user to another through the social networking platform because the company owns the platform on which those conversations occur. However, conversations taken place in a retail store do not belong to the organization despite two customers having a discussion while in the store. This raises the ethical debate as to who owns the data, how it should be used, and how it should be secured. However, many technologies are too new to answer many questions. We need further research to explore ethical issues related to big data and how the levels of detail being captured should be used. Furthermore, data scientists who have expertise in quantitative skills (i.e., how to use big data tools for managing huge sets of data like Hadoop and analytics) and effective at communicating their findings in a manner that functional managers can understand are rare (Tata Consultancy Services, 2013). Since some industries have invested heavily in IT over several decades, they have different levels of data intensity.

In this paper, we discuss trends and raise awareness of these issues to stimulate further research into big data to address big data's capabilities, privacy issues and ethical concerns, and security issues. This opens to door to a plethora of research possibilities ranging from new analytical techniques to ethics and privacy. From the organizational perspective, we need to improve the ability to generate information from images and videos beyond the meta-data provided by the content developer. Images can contain a lot of information relating to who, what, when, where, why, and how things took place. User comments provide feedback to the events taking place in the image that provide additional information. From the user perspective, we should evaluate ethics and privacy concerns when using this information for organizational gains. Most users provide content to share with their friends and other social contacts and not organizations to evaluate and determine the best way to manipulate them into purchasing products or services.

Another area of research could revolve around mashup research, which involves combining data from multiple sources to research phenomenon that we were previously unable to evaluate. Many data sources contain personally identifiable information (PII) that one can use to group individuals based on various factors such demographic and geographic data. As an example, a researcher could combine social networking data from Twitter and Facebook with Google Maps to conduct textual analysis across multiple networks to determine how perceptions in various regions may differ. Organizations continue collected large amounts of data to answer questions that have yet to be asked or test hypotheses that have yet to be developed (Agarwal & Dhar, 2014). This is where the creativity of researchers will come in to start discovering the questions and hypotheses that can take advantage of these data sets as we continue to advance business and the IS field. While we identify some areas that are in need of immediate attention, the opportunities for continued research will grow as expeditiously as big data itself.

References

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

Bawab, H. (2014). Privacy concerns raised around facial recognition technology. *LinkedIn.* Retrieved from http://www.linkedin.com/today/post/article/20140407061644-14091619-privacy-concerns-raised-around-facial-recognition-technology

Bertolucci, J. (2013). Big data ROI still though to measure. *InformationWeek*. Retrieved from http://www.informationweek.com/big-data/big-data-analytics/big-data-roi-still-tough-to-measure/d/d-id/1110150?

Beyer, M. (2011). Gartner says solving "big data" challenge involves more than just managing volumes of data. *Gartner*. Retrieved from http://www.gartner.com/newsroom/id/1731916

Chen, H., Chiang, R., & Storey, V. (2012). Business Intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188.

Compeau, D., Haggerty, N., & Fraiha, S. (2011). *Privacy issues and monetizing Twitter*. Ivey Publishing.

Davenport, T., & Dyche, J. (2013). Big data in big companies. *International Institute for Analytics.* Retrieved from http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/bigdata-bigcompanies-106461.pdf

Ferrera, P., De Prado, I., Palacios, E., Fernandez-Marquez, J.-L., & Serugendo, G. D. M. (2013). Tuple MapReduce and Pangool: An associated implementation. *Knowledge and Information Systems.*

Fulgoni, G. (2013). Big data: Friend or foe of digital advertising? *Journal of Advertising Research*, *53*(4), 372-376.

Gartner. (2013). Drive value from big data through six emerging best practices. Retrieved from https://www.gartner.com/doc/2600415?ref=SiteSearch&sthkw=2013%20big%20data%20volume%20velocity%20variety&fnl=search&srcId=1-3478922254

Gillon, K., Aral, S., Lin, C. Y., Mithas, S., & Zozulia, M. (2014). Business analytics: Radical shift or incremental change? *Communications of the Association for Information Systems*, *34*(13), 287-296.

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

Hashem, I., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98-115.

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

Jablonski, J. (2014). Maximizing return for big data projects. Presentation presented at 2014 AMCIS Panel Session.

Kabir, N., & Carayannis, E. (2013). Big data, tacit knowledge and organizational competitiveness. In *Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning* (pp. 220-227).

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

Marchand, D., & Peppard, J. (2013). Why IT fumbles analytics. *Harvard Business Review, 91*, 104-112.

Martin, K. (2010). Facebook (A): Beacon and privacy. *Institute for Corporate Ethics*. Retrieved from http://www.corporate-ethics.org/pdf/Facebook%20_A_business_ethics-case_bri-1006a.pdf

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*, 60-68.

Merrick, A. (2014). A death in the database. *The New Yorker*. Retrieved from http://www.newyorker.com/business/currency/a-death-in-the-database

Morris, J. (2012). Top 10 categories for big data sources and mining technologies. *ZDNet*. Retrieved from http://www.zdnet.com/top-10-categories-for-big-data-sources-and-mining-technologies-7000000926/

Negash, S. (2004). Business intelligence. *Communications of the Association for Information Systems*, *13*, 177-195.

O'Toole, J. (2014). Facebook's new face recognition knows you from the side. *CNN Money.* Retrieved from http://money.cnn.com/2014/04/04/technology/innovation/facebook-facial-recognition/

Stephens, C. (2013). The power of big data and high performance analytics. *Finweek*, 4-5.

Tata Consultancy Services. (2013). *The emerging big returns on big data*.

Vriens, M., & Brazell, F. (2013). The competitive advantage. *Marketing Insights, 25*(3), 32-38.

Wixom, T., Ariyachandra, T., Douglas, D., Goul, M., & Gupta, B. (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems*, *34*, 1-13.

About the Authors

J. P. Shim is a faculty of Computer Information Systems and Executive Director of Korean-American Business Center at Robinson College of Business at Georgia State University and professor emeritus at Mississippi State University (MSU). Before joining at GSU in 2011, he was faculty of BIS and Larry and Tonya Favreau Notable Scholar at MSU. During the past 27 years at MSU, he was a seventeen-time recipient of outstanding faculty awards, including 1994 John Grisham Faculty Excellence awardee, 2006 Powe Research Excellence award winner, and winner of 2011 MSU Diversity award. He has published several books and seventy journal papers. He serves on Wireless Telecommunication Symposium as Program chair and served on 2013 AMCIS Program co-chair. He has received awards, grants, and distinctions, including NSF, Microsoft, U.S. Small Business Administration, Japan Foundation, and Korea Foundation. He has been interviewed by the media (CBS TV, AP, AJC, Global Atlanta) and worked as a consultant for Booz Allen, U.S. EPA, Rehabilitation Associates, Inc. (in Wisconsin), CYR International, and Kia Motors. His current research interests are cross-cultural study of BYOD, big data, speech analytics, and wireless telecommunications.

Aaron M. French is an Assistant Professor of Management Information Systems in the College of Business at the University of New Mexico in Albuquerque, New Mexico. He received his PhD in Business Information Systems at Mississippi State University. He is a three-time recipient of the Outstanding Teacher of the Year Award. His research has been published in the *Journal of Information Technology*, *Behaviour & Information Technology*, *Journal of Internet Banking and Commerce*, and *The Journal of Internet Electronic Commerce Research*. His research interests include social networking, e-commerce, cross-cultural studies, and technology acceptance.

Chengqi (John) Guo is an Assistant Professor and Madison Research Fellow of Computer Information Systems and Management Science in the College of Business at James Madison University. He received his PhD in Business Information Systems from Mississippi State University. He received a Masters of Operations Management and Information Systems from Northern Illinois University and a BS in International Marketing from Guangdong University of Foreign Studies, Guangzhou, China. He is a senior consultant and Director of International Business Development at JDArray Co. Ltd, Beijing China. His research interests are information systems security, social media, mobile computing, technical innovation, human-computer interaction (adoption, trust, privacy, and communication), innovative technology in education, and cross-cultural studies.

Joey Jablonski is currently a VP/Principal Architect, leading the Big Data Practice at Cloud Technology Partners, a professional services firm focused on strategy, architecture and development of cloud applications and platforms. Before joining Cloud Technology Partners, he was an Enterprise Technologist at Dell working as a member of the Office of the CTO. He focused on the architecture and strategy for the deployment of complex analytic technologies and big data platforms. His technical interests include big data, high performance computing, low-latency networking technologies and information security. He has previously held technical and organizational leadership roles at DataDirect Networks, HP, and Sun Microsystems.