



# Difficulties Implementing Big Data: A Big Data Implementation Study

Kyle Spraker<sup>(✉)</sup>

Florida Gulf Coast University, Fort Myers, USA  
krs.spraker@gmail.com

**Abstract.** The exponential increase in the volume of data, the velocity at which they are created and their vast and progressively expanding varieties can be derived from virtually all aspects of our everyday lives. This situation prompts the need for urgent change in the way data are stored, received, and analyzed. Today, organizations under appreciate the changes, growth and strategic development their Big Data are capable of providing. Organizations also overestimate their own ability to access and interrupt their data in order to derive benefit from it. Big Data problems in organizations have historically been approached with an isolated outlook, rather than viewing issues as co-dependent parts of one another. The purpose of this research is to identify and summarize the general challenges faced by an organizations ability to adequately utilize and capitalize on the opportunities presented by its Big Data. Through the research methods of data collection and multiple case study analysis, this paper proposes a three-step framework model, which focuses on definition, organization and value creation. The proposed framework serves to mitigate Big Data implementation challenges and their involved issues faced by an organization.

**Keywords:** Big data · Implementation · Difficulties

## 1 Introduction

Research from the International Data Corporation [1] shows that 1.8 ZB of data were created and replicated in 2011, and estimates that the figures will multiply 50 times by the year 2020. Currently, around 90% of the world's existing digital data were created over the past 2 years [2]. This exuberant amount of available data is widely referred to as "Big Data". The concepts and ideologies that support Big Data are rapidly evolving. The most frequently adopted definition of big data is one suggested from the 2001 research of Douglas Laney, in which he states that Big Data is based on its characteristics or the rule of 4 V: volume, variety, velocity, and value. However, more recent research suggests that veracity and visualization are also important characteristics when defining Big Data [3, 4].

Developing dimensions for what define big data is challenging due to the many existing definitions and lack of agreement as to what constitutes big data and how to recognize it [5]. Research that surveyed 400 companies from around the world identified that 77% of the companies surveyed did not have a clear vision or strategy for their

enterprise Big data processes. Thus, organizations of all kinds are striving to build up their Big Data capabilities, leverage their data competitively and hurdle the inevitable challenges that Big Data presents [6].

Developments and investigations in Big Data have proven that organizations aren't ready to embrace and leverage big data to improve their organization's performance [3]. A number of new organizational challenges and barriers arise such as developing new employee skills, upgrading IT infrastructure and instating new management practices within an organization [3].

The article presents a focused overview of the characteristics that make up Big Data, their challenges and shared difficulties of implementations and compares them against existing research. The study proposes an enterprise framework as a holistic approach to overcoming the organizational challenges of Big Data.

## **2 Literature Review**

### **2.1 Difficulties Implementing Big Data**

Whether it's a government protecting its citizens, a cell phone sending a tweet, an airline offering a sale, or a farmer harvesting his crops mass amounts of data are being created and collected every second. The ability to use data in order to execute decisions as intended, to make a decision faster and to create competitive advantage, play a large role across many different types of organizations. An increasing reliance on technology, the growth of the global marketplace and an overabundance of data have increased the need for business units of all types to recognize, understand and deal with the challenges of interpreting and deriving value from mass quantities of data. The following review of literature confirms that big data collection and analysis present problems that go well beyond mass volume and variety of data, this review discusses specific and general solutions and concludes that enterprise-wide initiatives and a holistic view on big data are needed for gripping the challenges and difficulties that the implementation of big data present.

### **2.2 Identifying Potential Benefits**

New technology and innovation in fields like big data are creating new and exciting application possibilities that have the potential to benefit many different industries. For example, the case study analysis research conducted by Huang [7] identifies multiple problems that big data can solve for health sciences. For example, healthcare recommendation systems, sensor-based health condition and food safety monitoring and internet-based epidemic surveillance [7]. Additionally, research performed by Wang [8] through a quantitative approach of multiple case analysis identified 179 potential benefits that big data analytics can derive for healthcare organizations. Most notably the immediate access to analyze clinical data and the ability to improve cross-functional communication and collaboration among administrative staffs [8]. The highly technical healthcare industry is not the only beneficiary of big data. Kamilaris's [9] research offers insights into big data opportunities and challenges in the agricultural field. The research

was conducted via a bibliographic analysis beginning with 1330 papers and then narrowed down to 34 that were considered. Kamilaris's [9] research opens by describing exciting potential areas of application for big data in the agricultural industry. A few of these applications being, remote sensing of land and crop mapping, self-operating agricultural robots and tools that help better predict harvest yields.

### 2.3 Defining the Problem

There are many potential benefits that can be derived from big data for all different types of industries, but prospective benefits don't come without their challenges. Kamilaris's [9] research goes on to suggest some of the current technical issues that the agricultural industry struggles with relating to big data, such as sensing systems that collect and deliver images in a timely manner, data accuracy, and other common issues like data reliability and variety of data. However, developing dimensions for what defines big data and data analytics are challenging due to the many existing definitions and lack of agreement as to what constitutes big data and how to recognize it [5]. Researched by Isitor and Stanier [5] through case study analysis they have identified big data in terms of data characteristics or the 3 V's which are velocity, variety, and volume. The speed at which data is received or processed, the format or structure in which the data is received and lastly, the mass volume of data that is received [5].

Each of the defining characteristics of big data come with their own unique challenges and implementation difficulties. For example, Friedberg [10] used surveys and voluntary data collection on health practices both large and small regarding their data which is required to participate in alternative payment models. The authors [10] reported that the data they received from external sources were not timely enough, therefore, limiting their usefulness towards the purpose of improving payment model performance. The authors [10] also found that performance data relating to one payer was continually begin corrected and coming in late, making hard for the operation to gauge its own performance and make corrections in a timely manner.

Products like cloud computing have paved the way for faster infrastructure set up and decreased pitfalls such as data delays [11]. However, cloud computing has also left security and privacy of user information in business-critical applications vulnerable [11]. Rastogi's [11] case study analysis research suggests that there are security and privacy concerns when the exchange of information occurs between a cloud server and a user. Unfortunately, the challenges don't end at the security and privacy of the data being collected. Porche [12] examined case studies and records to get a sense of the overwhelming amount of data being generated by the U.S. Navy. The Intelligence Science Board [13] said that the number of images and signal intercepts are far beyond the manpower and capacity available to the Navy. Huge backlogs for translators and image interpreters have amassed and most of what is collected has never been reviewed. The authors [12] present their view on big data as a data set so large that it passes the limits of traditional relational databases along the four parameters of volume, variety, velocity, and veracity. The authors [12] found that all of the information that the Navy collects is considered potentially useful, but processing and deriving value from the vast datasets is beyond the analytical capability of the Navy's human resources. As the Navy acquires

new equipment and technology evolves the amount of data will only grow and so will the difficulty to process it [12].

Additional research into understanding the implementation challenges of big data in the search of a solution was done by Jin [3] through case study analysis of 21 different research studies present the grand challenges of big data. Jin [3] described that the traditional features of big data can be characterized by volume, velocity, variety, veracity, and value. However, other research conducted by Chen and Zhang [4] suggests that variability and visualization should be included in the characterization list as well. From research conducted by Huang [7], it is suggested that the best way to intuitively obtain meaning from data is through a graphical presentation, also known as “visualization.” Ultimately, out of all of these characteristics presented, the main challenges of big data are suggested to be variety, velocity and veracity [3]. The authors [3] conclude by suggesting that the solution to solving these major big data challenges lie within an integrated data solution, rather than isolated successes of a few aspects of big data.

## 2.4 Finding Solutions

Ribarsky [14] conducted research about how value can be derived from the analysis of unstructured data, in this case, social media data, primarily focused on streaming Twitter data. The authors [14] collected a 1% random sample of these data for 1.75 years (approx. 20B Tweets). The authors [14] found that providing meaning and organization to unstructured data, with various technologies and naming strategies offers the ability to analyze competitor strategies and knowledge about the demographics of people that are generating messages during certain events. Another example of deriving value comes from a [15] study of sales data provided by a pharmaceutical distribution company. Other research methods included exploratory analysis, graph-based analysis, data sampling for model fitting and testing, model building or sales prediction and model evaluation. The authors [15] offered a suggested method for modeling and forecasting sales and how to control inventory levels in order to prevent costs of excessive inventory and to prevent losing customers due to a drug shortage.

Not all companies have answers to the hard-hitting questions of big data, Wang [8] points out research by Wegener and Sinha [16] who conducted a survey of 400 companies from around the world and found that 77% of companies surveyed did not have clear strategies for effectively utilizing their big data. Wegener and Sinha [16] also suggest that research is needed to formulate appropriate strategies that will enable organizations to leverage big data analytics efficiently and effectively. Additionally, Alharthi's [17] research through case study analysis and corporate benchmarks suggests that there are three main types of barriers to big data, which are, technological, human, and organizational. Under the barrier of organization, Alharthi [17] suggests that cultural barriers within an organization have a major impact on an organizations strategy, structure, and processes when it comes to their big data. In order to overcome these organizational barriers to big data, the authors [17] suggest that documenting, implementing and communicating a clear organizational vision in relation to big data is a step towards improving organizational performance.

## **2.5 Growing Smarter Before Data Grows Larger**

The global marketplace will only continue to grow larger and data complexity and volume will grow alongside it. Research on difficulties in implementing big data are only just beginning to blossom and have much room to grow. In light of the adolescences of the topic, the literature review defines several issues involved in the characteristics of volume, velocity, variety, veracity, visualization, and value, which make up big data and states a few proposed solutions such as, defining a clear vision and strategy for big data. However, further research and innovation are going to be the fundamental factors in determining how companies and organizations choose to plan for and handle their big data operations.

## **3 Research Methodology**

To reach the goals of this study we analyzed previous literature to gain a fundamental understanding of Big Data challenges and potential benefits.

### **3.1 Case Collection**

Cases were collected from past Big Data research papers, articles, and reports from companies. Beginning with a keyword search from the scientific databases ScienceDirect, ABI/Inform Collection, Academic Search Complete, ACM Digital Library and JSTOR (Ebsco). The following keywords were used in order to query the previously mentioned databases:

“Big Data” AND [“Enterprise” OR “Challenges” OR “Benefits” OR “Definition” OR “Opportunities” OR “Framework”]

Out of the large variety of Big Data research available, we initially identified 61 papers that were fitting to the research. Papers were then refined based on the case’s presentation of Big Data implantation and categorized based on industry. The initial 61 papers were refined down to 32 that specifically met the themes and topics required for the study.

### **3.2 Research Approach and Process**

A content analysis was then conducted on the reviewed research in order to extract topics and themes and understand the research on a more refined level. The process consisted of breaking down articles into relevant themes, organizing those themes based on their nature (problem, solution, challenges, and insight) and finally reporting on the research in a meaningful manner.

Themes were identified from the analysis in a structure that moved from potential benefits to challenges and difficulties and finally to potential solutions.

## 4 Proposed Framework

### 4.1 Research Themes

Three distinctive themes emerged from this study. One theme is that any contribution to Big Data analysis in an organization, mostly fall completely on an individual data analysis department. The second theme being that Big Data issues are being examined and researched as isolated challenges, rather than being examined as co-dependent events of the other Big Data characteristics. The last theme identified in this research is the lack of organizational structure and strategy in regards to Big Data.

**Theme 1.** One of the major barriers to successfully implementing and utilizing Big Data in an organization has to do with that organization's culture, specifically towards Big Data [17]. Alharthi [17] suggests that documenting and communicating a clear organizational vision in relation to Big Data is one of the methods in which to lift organizational barriers around Big Data. Successful Big Data operations require cross-functional communication and collaboration [8].

**Theme 2.** Developing dimensions that make up and define big data are challenging due to the many existing definitions and lack of agreeance on which is correct [5]. Big Data solutions lie within integrated solutions, rather than the isolated success of a few aspects of Big Data [3].

**Theme 3.** Organizing a controlled effort for implementation and utilization of Big Data methodologies can be difficult without a formal plan or order of operations. During a survey of 400 companies, 77% of companies self-reported that they did not have strategies for utilizing their Big Data effectively [8].

### 4.2 Open Issues in Implementing Big Data

Big Data is still largely an emerging field of study, therefore many risks and challenges still exist and have yet to be researched. However, from the literature review, we found a few of the outstanding problem areas in Big Data:

- Value creation
- Security and privacy
- Data format variety
- Organizational, technological and human barriers
- Clear vision and Big Data strategy [3, 11, 12, 17].

### 4.3 Enterprise Big Data Framework

According to prior studies, the following actions were typically done by organizations who were implementing Big Data:

- Relying solely on a Big Data team for all input, organization, and contribution to the organizations Big Data efforts. As opposed to a holistic organizational approach in which each business unit has some contribution, increasing the organizational overall chances of success with Big Data [17].
- Organizations are viewing big data issues as isolated events rather than viewing them as co-dependent challenges, which require an integrated solution [3].
- Large percentages of organizations have a lack of strategy and organizational structure when it comes to Big Data. The ability to express a clear vision of the organizations Big Data strategy can greatly improve its chances of success [8].

In light of these identified issues from the reviewed literature, we propose a framework which consists of three sections to consider when implementing Big Data processes in an organization, which are definition, organization and value creation (Table 1).

**Table 1.** Themes for Big Data implementation

Definition	Define the organizations Big Data vision <ul style="list-style-type: none"> <li>• Vision: Define what the different characteristics of Big Data mean to your organization:</li> <li>• Volume</li> <li>• Velocity</li> <li>• Variety</li> <li>• Veracity</li> </ul>
Organization	Identify a dedicated data owner for each business unit: <ul style="list-style-type: none"> <li>• Sales:</li> <li>• Marketing:</li> <li>• Technology:</li> <li>• Engineering:</li> <li>• Human Resources:</li> </ul> Establish single data sources to avoid overlapping/redundant data. Establish data standards: <ul style="list-style-type: none"> <li>• Format</li> <li>• Timeliness/Time of delivery</li> <li>• Data Types</li> <li>• Data change procedures</li> </ul>
Value creation	The last step is to interrupt and utilize your data to create value. <ul style="list-style-type: none"> <li>• Visualization:</li> <li>• Implementation:</li> </ul>

## 5 Conclusion

It is no secret that Big Data has become an aspect of everyday life and is gradually becoming more and more important in most business functions. Systems, data, and

collection methods are becoming more complex and collecting more data than ever before. The IDC [1] reported that 1.8 ZB of data were created in 2011 and is expected to be 50 times that by the year 2020. As a result, Big Data opportunities and challenges are changing and expanding every single day. New technology will be needed to handle larger, more complex data, new organizational structures will be needed in order to develop company cultures conducive to a Big Data mindset, and new talent will be needed in order to keep up with the growing demand for data professionals [8, 17].

Successfully hurdling the challenges of implementing Big Data requires, not only technical abilities but structural and organizational abilities as well. The conclusion drawn from discussion and literature reviews suggests that Big Data developments are only as strong as the foundation and organization they're built on. In addition, multiple outstanding issues for future research have been identified and discussed. Finally, the challenges of Big Data implementation should be faced with a holistic ideology and framework, rather than a modular outlook that focuses on the isolated limitations to Big Data.

## References

1. IDC. 2011 Digital universe study: extracting value from chaos (2011). <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>
2. Gobble, M.: Big Data: the next big thing in innovation. *Res.-Technol. Manage.* **56**(1), 64–67 (2013)
3. Jin, X., Wah, B., Cheng, X., Wang, Y.: Significance and challenges of big data research. *Big Data Res.* **2**(2), 59–64 (2015)
4. Philip Chen, C., Zhang, C.: Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. *Inf. Sci.* **275**, 314–347 (2014)
5. Isitor, E., Stanier, C.: *Defining Big Data* (2016)
6. IDC: New IDC forecast sees worldwide big data technology and services market growing to \$48.6 billion in 2019, driven by wide adoption across industries [Press Release] (2016). <http://www.idc.com/getdoc.jsp?containerId=prUS40560115>
7. Huang, T., Lan, L., Fang, X., An, P., Min, J., Wang, F.: Promises and challenges of big data computing in health sciences. *Big Data Res.* **2**(1), 2–11 (2015)
8. Wang, Y., Kung, L., Byrd, T.: Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Chang.* **126**, 3–13 (2017)
9. Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.: A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* **143**, 23–37 (2017)
10. Friedberg, M., Chen, P., White, C., Jung, O., Raaen, L., Hirshman, S., Lipinski, L.: Increased importance of data and data analysis. In: *Effects of Health Care Payment Models on Physician Practice in the United States*, pp. 53–62 (2015)
11. Rastogi, N., Gloria, M., Hendler, J.: Security and privacy of performing data analytics in the cloud: a three-way handshake of technology, policy, and management. *J. Inf. Policy* **5**, 129–154 (2015)
12. Porche, I., Wilson, B., Johnson, E., Tierney, S., Saltzman, E.: Big data: challenges and opportunities. In: *Data Flood: Helping the Navy Address the Rising Tide of Sensor Information*, pp. 1–6 (2014)



13. Intelligence Science Board, Integrating Sensor-Collected Intelligence. Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, Washington, D.C., November 2008
14. Ribarsky, W., Xiaoyu Wang, D., Dou, W.: Social media analytics for competitive advantage. *Comput. Graph.* **38**, 328–331 (2014)
15. Zadeh, N.K., Sepehri, M.M., Farvaresh, H.: Intelligent sales prediction for pharmaceutical distribution companies: a data mining based approach. *Math. Prob. Eng.*, 1–15 (2014)
16. Wegener, R., Sinha, V.: The value of big data: how analytics differentiates winners
17. Alharthi, A., Krotov, V., Bowman, M.: Addressing barriers to big data. *Bus. Horiz.* **60**(3), 285–292 (2017). Bain and Company (2013)