2020

Exploring Strategies to Transition to Big Data Technologies from DW Technologies

Mbah Johnas Fortem
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral study by

Fortem Mbah

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Jon McKeeby, Committee Chairperson, Information Technology Faculty
Dr. Jodine Burchell, Committee Member, Information Technology Faculty
Dr. Steven Case, University Reviewer, Information Technology Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2020

Abstract

Exploring Strategies to Transition to Big Data Technologies from DW Technologies

by

Mbah Johnas Fortem

MS, Walden University, 2016

MS, University of Applied Science, Austria, 2014

BSc, National University of Ukraine, 2012

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Information Technology

Walden University

May 2020

Abstract

As a result of innovation and technological improvements, organizations are now capable of capturing and storing massive amounts of data from various sources and domains. This increase in the volume of data resulted in traditional tools used for processing, storing, and analyzing large amounts of data becoming increasingly inefficient. Grounded in the extended technology acceptance model, the purpose of this qualitative multiple case study was to explore the strategies data managers use to transition from traditional data warehousing technologies to big data technologies. The participants included data managers from 6 organizations (medium and large size) based in Munich, Germany, who transitioned from data warehousing technologies to big data technologies. Data collection included interviews with 10 data managers and a review of 15 organizational documents. Inductive coding was used to analyze the data. Four major themes identified included identify a business need or use case, identify data sources, executive support, and use a data lake. A key recommendation is that data managers can use these findings to implement best practices when transitioning to big data technologies, thereby improving successful big data transition implementations and adoption. An implication for positive social change from this study is that data managers and organizational leaders might use the research findings to improve products and services offered, leading to better products and or services offerings for consumers.

Exploring Strategies to Transition to Big Data Technologies from DW Technologies

by

Mbah Johnas Fortem

MS, Walden University, 2016

MS, University of Applied Science, Austria, 2014

BSc, National University of Ukraine, 2012

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Information Technology

Walden University

April 2020

Dedication

I dedicate this work to my mother Julie, and my siblings Walters and Ernestine, for their unconditional support, and motivation while I worked through this long but fulfilling process. I am very grateful that I will be completing this program while my mother remains alive, so she can celebrate with me and be proud of my achievement. I My roots are from an average working family in a rural area in Cameroon, so accomplishing this goal is truly a dream come true and unique achievement. I am the first in our family to achieve a doctoral-level degree, and I hope it will not be the last Doctoral degree of my family.

Acknowledgments

There are several people I wish to thank for their support and encouragement. First, I want to thank my committee chair, Dr. Jon McKeeby, for his unfailing support. Dr McKeeby always encouraged me to continue working hard even in the darkest of times. Without his mentorship, I have strong reasons to believe I might not have successfully completed this program. I would also like to thank Dr. Jodine Burchell. She was instrumental to me completing my proposal. She provided judicious advice and valuable input together with well-timed inspiration. I would also like to thank Dr. Steven Case for his speech during the second residency, which was the catalyst that gave the much needed direction to my study. I would also like to thank Dr. William Wood, and Dr. Jeff Miko for their contributions and time in reviewing my work. A big thank you goes to all my study participants. You gave of yourselves generously and provided valuable input to my study. I am hopeful you find the results worthy of your time investment. I would like to thank my peers for their continual support and motivation over these past few years. Lastly, I would like to thank my family for continual support and perseverance during this entire process.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study

Section 1 includes a discussion on the foundation of the project. This section introduces and emphasizes the background of the problem, purpose of the study, the problem as seen by current data managers, and the approach to evaluating the problem. Then the research design is described, which was focused primarily on an applicable qualitative design while using the extended technology acceptance model (TAM2) as the conceptual foundation for evaluating the problem. The remaining sections define terms needed to understand the context of this project, the research questions used in interviewing data managers in the data collection process, and assumptions, limitations, and delimited conditions for the project. Finally, the last components of the section address the significance of the study to current and existing research and the positive impact to social change as well as the contribution to data managers and current and future data initiatives and practices.

Background of the Problem

The amount of data generated by humans evolved with technologies like mainframes computers, client/server technologies, enterprise resource planning systems, Web 2.0, storage technologies, and the Internet-of-things (IoT) (Hormann & Campbell, 2014; Mohanty, Jagadeesh, & Srivatsa, 2013). Data, big or small, for many organizations has always been an integral part, as it allows them to gain insights into data for increased information technology (IT) performance, automation, and business profitability (Auffray et al., 2016; Gandomi & Haider, 2015). Organizations that employ a data-driven decision-making style are more successful than those that make decisions based on intuition and experience (Waller & Fawcett, 2013).

While there is no widely accepted definition of *"big data*," Gartner glossary (2013) defined big data as "Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." Further, with the advent of big data, businesses are faced with new challenges dealing with datasets from a variety of sources and size beyond the capacity and format of typical database applications like traditional DWs to capture, process, and store, integrate, manage, and analyze (Padgavankar, & Gupta, 2014). Big data include information assets that require forms of information processing that help with decision making ("Big Data," n.d.). Thus, limitations in data warehousing and business intelligence make it imperative for businesses to consider new technologies like big data if they want to maintain a competitive advantage (Hashem et al., 2015). This study was focused on the strategies that data managers use to transition from data warehouse (DW) technology to big data technologies. I also explored the characteristics of big data and explored emerging themes that helped answer the research question.

Problem Statement

Because of the limitations in DW technologies to process unstructured data and limited knowledge in big data technologies (Coleman et al., 2016), data managers are having problems transitioning to big data technologies from DW technologies (Bansal & Rana, 2014). Due to the limitations in DW technologies to handle and process high volumes of data (Win & Thein, 2015), coupled with the lack of support for evolving business needs like support of real-time and ad-hoc customer analytics, 40-50% of DW initiatives have ended in failure (Asrani & Jain, 2016; Kimpel & Morris, 2013). The general IT problem is that limited knowledge in big data technologies is negatively affecting IT performance and efficiency in organizations. The specific IT problem is that some data managers lack strategies to transition to big data technologies from DW technologies.

Purpose Statement

The purpose of this qualitative multiple case study was to explore the strategies data managers use to transition to big data technologies from DW technologies. The target population consisted of data managers from 5 mid-sized organizations located in Germany that have transitioned to big data technologies from DW technologies. The chosen population have had experience with DW technologies and were or still involved in big data initiatives. These data managers participated in semistructured interviews to identify the strategies they used to overcome the limitations in the DW as they transitioned to big data technologies. Transitioning to big data technologies could have a positive social impact by facilitating information exchange by providing new interactive platforms designed to connect citizens and service users with organizations that are conducting interventions or representing their needs.

Nature of the Study

Three different research methodologies were considered in this study: qualitative, quantitative, and mixed methods. In a qualitative method, the researcher is the most significant instrument in the data collection process as well as inductive and recursive analysis of data to build up a complex picture of the question under investigation and ruminate on her or his role and responsibility in the study (Yilmaz, 2013). In contrast, the researcher in quantitative research uses an inferential approach based on a theory related to the question under study, then develops one or more hypothesis grounded on this theory and assesses the hypothesis to prove or disprove it with data using statistical procedures (Barczak, 2015). Consequently, a quantitative method was not appropriate for this study because the study did not involve any theory or hypothesis testing, dependent or independent variables, and numerical data collection for statistical testing. Additionally, mixed method research requires the collection, analyses, and integration of quantitative and qualitative research (Heyvaert, Hannes, Maes, & Onghena, 2013). Given the lack of quantitative component in my study, a mixed method by extension was not appropriate for the study.

As a qualitative approach, I chose a multiple case study design to help bring readers to an understanding of the complex issues surrounding a transition to big data technologies from DW technologies and adding strength to what is already known through previous research. A qualitative case study improves and expands a researcher's perception of a complex phenomenon in a real-world setting by answering *how* and *why* questions (Abma & Stake, 2014; Cronin, 2014). A multiple case design enables the researcher to gain a holistic view of the phenomenon under investigation, providing a rounded picture given the various sources of evidence (Tsang, 2014). Because I wanted to develop a comprehensive interpretation of the phenomenon under investigation, a multiple case design was appropriate.

Other qualitative designs were considered but not chosen. A phenomenological research design is used to explore human experience from the perspective of those living and experiencing the phenomenon (Sohn, Greenberg, Thomas, & Pollio, 2017), but this was not the focus or the objective of this study. Additionally, an ethnographic study requires the researcher to focus on describing social behaviors and cultures (Furuta, Sandall, & Bick, 2014) of a group of people to gain insights of the participants` social conducts within their culture (Cruz & Higginbottom, 2013). However, the focus of my study was not on the social or cultural behaviors of participants, so an ethnographic design was not appropriate. Finally, narrative designs usually occur in the context of an informant sharing a story, followed by the researcher's analysis of that story (Campbell, 2014). But my research was focused on the strategies data managers use to transition to big data from data warehousing, and studying the strategies used by a single data manager would not yield the relevant data to answer my research question. DW

Research Question

The overreaching research question for the study was "What strategies do data managers use to transition to big data technologies from DW technologies?"

Demographic Questions

1. What role and position do you currently hold at your organization?
2. How long have you been in your current position?

3. How many years of experience do you have in data management and analytics?
4. What types of diplomas, degrees or industry certifications do you possess?

Interview Questions

1. How would you describe big data and how would you compare big data to relational database technologies like DW?
2. What are the tools and methods your team uses to for big data initiatives and How would you describe the usefulness of those tools and methods?
3. What challenges have you had if any, with relational database management systems like DW in your big data initiatives?
4. What strategies have you used to transition from data warehousing to big data technologies?
5. Has your business analytics methodology and tools changed as a result of the transition to big data technologies?
6. Why did you consider making changes to your data management and analytics strategies?
7. How has the transition to big data technologies affected the overall performance of your team in performing their day to day tasks?

Conceptual Framework

The conceptual framework that was used for this study is the TAM2, which is an extension of the TAM. TAM was originally conceived by Fred Davis in 1986 and considered in the literature as the most influential extensions of the theory of reasoned action (TRA; Fayolle & Liñán, 2014). The central principle of the TAM is that acceptance of technology by a person or groups of people is dependent on their perceived usefulness and perceived ease of use of that technology (Cheng, 2019). Due to TAM`s limited predictive and explanatory power (Yoon, 2018), Venkatesh and Davis (2000) extended the original TAM to TAM2 by adding cognitive instrumental and social influence processes as determinants of the perceived usefulness and perceived ease of use of technology.

TAM2 depicts the impact of three social influence processes:—subjective norm, voluntariness, and image—as influencing an individual's decision to adopt or reject a system (Ramírez-Correa, Mariano-Melo, & Alfaro-Pérez, 2019; Yoon, 2018). TAM2 also reflects the influences of four cognitive instrumental processes (job relevance, output quality, result demonstrability, and perceived ease of use) that technology users use for assessing the match between essential or critical work objective and their perception of the usefulness of the technology (Marangunić & Granić, 2015). The TAM2 was used to examine the specific IT problem by investigating how cognitive instrumental factors and social influence factors impact the acceptance of big data technologies in organizations in the Germany. Investigating the cognitive and social processes was used to examine the strategies used to transition to big data technologies from DW technologies by data managers. This approach allowed for a better understanding and interpretation of the strategies used by data managers as they transition to big data technologies from data warehousing technologies.

Definition of Terms

Big data: Big data is a general term used to describe the substantial amount (2.5 quintillion bytes of data added each day) of digital data produced from a variety of sources that is too large and lacks the structure to be analyzed using conventional relational database techniques (Kharat & Singhal, 2017).

Business intelligence: Business intelligence is a combination of strategies, technologies, and tools for the analysis of business data or information (Arnott, Lizama, & Song, 2017).

Data Warehouse (DW): A DW, also known as an enterprise DW in IT, is a central repository of data integrated from one or more incongruent sources (Jukic, Vrbsky, & Nestorov, 2016).

Semistructured data: Semistructured data that does not reside in a traditional database system like structured query language but have some form of structure or organization (Yan, Meng, Lu, & Li, 2017).

Structured data: Structured data are data that depend on the data model and reside in a fixed field within a record. Structured data are the type of data that are generally stored in traditional database systems like structured query language (Guha, Brickley, & Macbeth, 2016).

Unstructured data: Data that cannot be easily fit into a data model because the content are context-specific and or varying (Seng & Ang, 2019).

Assumptions, Limitations, and Delimitations

Assumptions

The researcher will make certain assumptions based on certain aspects of the study with regard to the research participants and methodology. Assumptions are views and beliefs considered by the researcher to be true and executes on the study (Rahi, 2017). Assumptions form an integral part of the research problem and model the study (Rahi, 2017). However, researchers are influenced by their beliefs, and a lack of examination or consideration of these assumptions or beliefs could lead to questionable findings (Rahi, 2017). I made the following assumptions for my study (a) my interview questions would be understood by the participants; (b) study participants would give honest answers to my questions by comprehending that their responses would remain confidential and private; (c) all chosen participants would be able to participate in the interview process including follow-up interviews; (d) criteria used for choosing the sample population were fitting, ensuring that each participant was knowledgeable and experienced with data warehousing, business intelligence, and big data; (e) interview participants would provide responses that are representative of my research population; and (f) qualitative methodology would provide the correct data and constructs for exploration.

Limitations

Limitations in a study represent the potential weaknesses, restrictions, and shortcomings in the study that are in most cases out of the control of the researcher, given the choice of research, research participants, available funding, and other factors (Busse, Kach, & Wagner, 2016). Limitations are also the restrictions on the interpretations and conclusions of a study because of the chosen research topic and methodology (Greener, 2018). The study was limited to data managers located in Germany, which may limit the representability of the study. Other limitations include internal company policies on data governance and information sharing policies that I might not be aware of that could restrict research participants from answering some questions in detail or bias the participants' answers.

Delimitations

Delimitations denote the boundaries in a study as identified by the researcher (Snelson, 2016). The researcher sets these delimitations, so the research goals do not become difficult to complete, and items out of the boundary range are irrelevant to the problem under investigation (Greener, 2018). A delimitation for this study is the criteria for selecting the research participants, which required that each participant be actively working with big data initiatives involving a transition from data warehousing technologies. My choice for the organizational structure, settings, and participants are also delimitations as characteristics of the study that can be influenced by the researcher (Shariati & Seyyedrezaei, 2015). This study was also limited geographically to mid-sized companies located in Munich, Germany, which enabled me to narrow the scope of the research based on the type of participants and the context of the study. Furthermore, the interview questions were limited to transition strategies from data warehousing technologies to big data technologies and nothing more.

Significance of the Study

I did not find any existing studies that address the strategies to transition to big data technologies from traditional data warehousing technologies, meaning that this study can increase the knowledge and practice in this area. Improvement in the practice can provide additional benefits in multiple areas other than IT. This study may contribute to the practice of IT by helping IT specialists, especially data managers, better understand big data technologies and improve the knowledge in the field of data science.

Contribution to Information Technology Practice

This study has the potential of filling a literature gap regarding big data. Additionally, the study may have a positive impact on business and the practice of IT by providing knowledge on strategies to transition to big data from DW technologies. This increase in knowledge can lead to an increase in profitability, as organizations can benefit from big data implementation. Big data can be a transformative, wealth-creating force like the 19th century focus on oil (Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015). The findings of this study can also be used by business leaders and data managers to evaluate the effectiveness of the currently employed strategies by organizations to transition to big data from DW. It is important for business leaders and data managers to understand the technologies associated with big data, its characteristics, and main sources, so they can understand and develop effective strategies that will increase profitability, maintain a competitive advantage, and boost overall business morale.

Implications for Social Change

Transitioning to big data technologies represents new approaches that bring people with varying skill sets, different perspectives, and diverse backgrounds together as they work to solve a common problem. New skills could help create and increase revenue streams for the company, reduce production cost and maintenance, and lead to a higher job and customer satisfaction. Furthermore, transitioning to big data technologies may help in facilitating information exchange by providing new interactive platforms designed to connect citizens and service users with organizations that are conducting interventions or representing their needs. These technologies can also help promote a continuing process of discourse and feedback between citizens and services users and information providers. By processing and analyzing large data sets stemming from a variety of sources (mobile devices, sensors, social media, etc.), organizations can evaluate citizens' information like financial product performance, patient health, and disease outbreak, among others. This information can then be used by these organizations to provide positive feedback and awareness to citizens that will have a long-lasting positive impact on their lives.

A Review of the Professional and Academic Literature

The purpose of this qualitative, multiple case study was to understand and explore the strategies data managers use to transition from DW technologies to big data technologies. The literature review was focused on the research question "What strategies do data managers use to transition from DW technologies to big data technologies?" The purpose of this literature review was to explore TAM2 (extended TAM), data warehousing technologies, big data technologies and transition strategies that data managers use in big data initiatives.

Primary libraries including databases that were used for the review include the Walden University Library, ProQuest Computing, ProQuest Dissertations and Theses Global, ACM Digital Library, EBSCO host Computers and Applied Sciences Complete, SAGE, IEEE Xplore Digital Library, Science Direct, and Google Scholar. Ulrich`s Global Serials Directory was used to determine the status (peer-reviewed or not peer-reviewed) of all articles. A total of 160 articles out of 200 articles initially selected were reviewed including four books and one dissertation, with 90 (90%) are peer-reviewed and 138 (88%) were published within the recommended 5 years of my projected graduation date (at the time of approval of the proposal).

The literature review is focused on the following key areas: (a) TAM2, (b) big data technologies, (c) DW technologies, and (d) transition strategies. Because TAM2 is an extension of TAM, the review of the TAM is focused on the perceived usefulness, perceived ease of use, social influence processes (subjective norm, voluntariness, and image) and cognitive instrumental processes (job relevance, output quality, result demonstrability, and perceived usefulness) relating to data warehousing and big data technologies. The investigation into DW included history, purpose, technologies, methodologies, and limitations of the DW. The investigation into big data included history, characteristics, tools, technologies, and use scenarios.

Technology Acceptance Model

After considering several theories to support my study, the TAM2 was the most appropriate for exploring the strategies to transition to big data technologies from DW technology. The TAM was developed to address determinants of technology acceptance and usage in the field of information systems and technology (Fathema, Shannon, & Ross, 2015; Marangunić & Granić, 2015). Because most studies carried out in the 1970s did not show reliable measures that could explain user acceptance or rejection of technology of systems, in 1985 Davis as part of his doctoral thesis proposed the TAM (Davis, 1985). Davis proposed that people's usage of a system is a phenomenon that can be explained or predicted by an individual's motivation to use the system, which is in turn influenced by some external stimulus which include among others the systems characteristics and features (Figure 1).

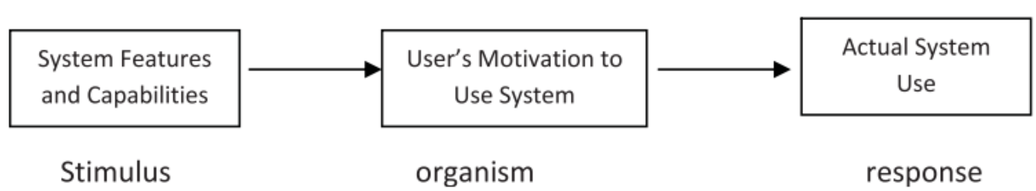

Figure 1. Conceptual model for technology acceptance. Adapted from *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results,* by F. D. Davis, 1985, p. 10. *Copyright permission to use for scholarly and educational purpose without additional approval.*

Davis (1985) suggested that users' motive or rationale to use a system can be explained by three factors: perceived ease of use, perceived usefulness, and attitude toward using the system. Davis posited that the attitude of a user toward a system is a determinant for whether the user will reject or use the system or technology. The user's attitude in turn is influenced by perceived usefulness and perceived ease of use, with the perceived usefulness having direct effect on perceived ease of use (Davis, 1989). Ultimately, both perceived usefulness and perceived ease are directly influenced by the system design characteristics (Davis, 1985).

Davis later refined his theory to propose the TAM as an extension of the TRA (Fayolle & Liñán, 2014) and the theory of planned behavior (Chu & Chen, 2016) to explain the factors influencing an individual's acceptance and use of technology. Davis (1985) also refined and enhanced his model to include additional variables to modify the relationships that he originally formulated. For example, Davis et al. (1989) added a new variable to TAM called behavioral intention that would be influenced directly by the perceived usefulness of a system. Davis et al. suggested that there would be instances where an individual could form a strong behavioral intention to utilize a system without the need to form an attitude even when the given system is perceived useful, resulting in a modified TAM (see Figure 2).

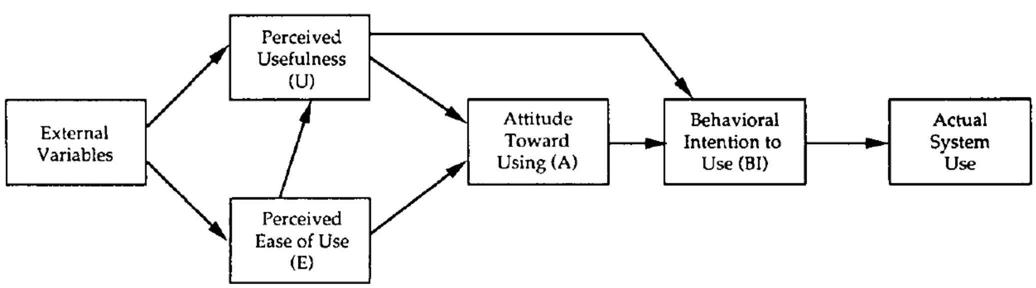

Figure 2. The technology acceptance model. Adapted from "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," by F. D. Davis, R. P.

Bagozzi, and P. R. Warshaw, 1989, *Management Science, 35*(8), p. 985. Copyright permission to use for scholarly and educational purpose without additional approval.

Now the TAM is considered as one the most influential information systems models as far as understanding people's intention to accept and use technology (Marangunić & Granić, 2015; Tarhini, Hone, & Liu, 2014). The TAM has been cited in most of the research that deals with user acceptance of technology (Rondan-Cataluña, Arenas-Gaitán, & Ramírez-Correa, 2015). Studies have shown the usefulness of the TAM as a theoretical basis or conceptual framework in explaining and predicting user behavior of IT (Gangwar, Date, & Ramaswamy, 2015; Peek et al., 2014; Tan, Ooi, Leong, & Lin, 2014). Other researchers proposed and applied multiple additions to the TAM (Venkatesh & Davis, 2000) until it developed into a leading model for predicting and explaining technology and system use.

Evolution of TAM to TAM2

With over 700 references to the TAM (Yurovich, 2015), Davis's examination has been adjusted from multiple points of view. Many researchers have even attempted to consolidate the results obtained from the earlier studies on the TAM (see Yousafzai, Foxall, & Pallister, 2007). But despite strong evidence from early research on the TAM, the model still had some limitations and could not go past the general elements that measured perceived usefulness and perceived ease of use (Venkatesh & Davis, 2000). Thus, it became a challenge to identify the reason(s) accounting for perceived ease of use or perceived usefulness variables utilized within the model. Furthermore, earlier research on the TAM was mainly focused on voluntary settings with little consideration for mandatory environments (Tom-Dieck & Jung, 2018). To address these issues, the TAM was extended. One of the most significant extensions added to the TAM is by Venkatesh and Davis (2000) who proposed an updated model called the TAM2 model (see Figure 3).

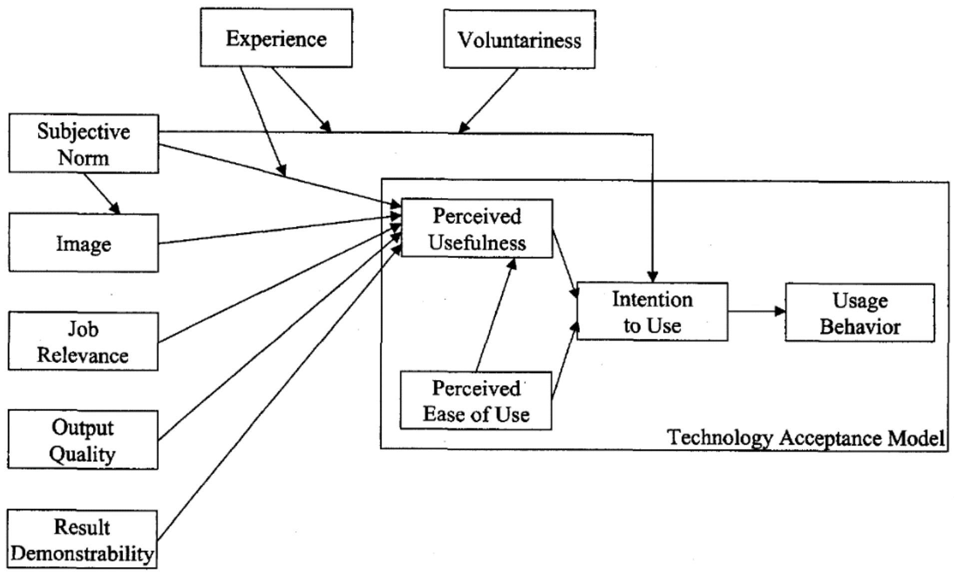

Figure 3. Extension of the technology acceptance model. Adapted from “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” by V. Venkatesh and F. D. Davis, 2000, *Management Science, 46*(2), p. 188. Copyright permission to use for scholarly and educational purpose without additional approval.

Venkatesh and Davis (2000) identified some of the limitations that the TAM had in explaining the reasons why a person would perceive the usefulness of a given system, so they suggested that additional variables be added as antecedents to the perceived usefulness variables in the original TAM (Miki, Kagiri, & Nganga, 2017). Venkatesh and Davis were also interested in other shortcomings like evaluating the performance of the new model, TAM2, in a mandatory environment or voluntary setting. A mandatory environment is one where the use of the technology in the organization is compulsory (Brown, Massey, Montoya-Weiss, and Burkman, 2002, p. 283). In most organizations, mandatory use of technology is an integral part of the workforce (Williamson, & Muckle, 2018). In a mandatory setting, ease of use is of greater importance, and adopting a technology in a mandatory setting could provide the following advantages to an organization: (a) show management’s commitment to employees, (b) promote positive change, (c) promote employee safety, and (d) for compliance and legal reasons. DW

TAM2 could now offer detailed explanations why participants found a given system or framework useful. By using the TAM2 model, Venkatesh and Davis (2000) could give itemized clarifications to the reasons members found a given framework helpful. Their outcomes additionally demonstrated that the TAM2 performed well in voluntary and mandatory with the exemption that subjective norm had no impact in voluntary settings yet did in mandatory settings. Hence, the current study was focused on organizations that transitioned to big data technology from DW technology in a mandatory setting. The primary focus was to explore where the perceived usefulness and usage intentions regarding the social influence of big data technologies ascertain their actual usage by data managers.

Social influence processes. Subjective norm is individuals’ perceptions that most people who are important to them believe that they should or should not perform a certain behavior, which is a determining factor of perceived usefulness and use intention (Zhou & Li, 2014). Research has shown that subjective norm significantly predicts people’s intention to use a specific technology or system (Venkatesh & Davis, 2000) Applied to software development, subjective norm explains how application developers are more likely to use a framework or programming methodology if they are convinced those who are important to them approve of their usage (Martinez, Cachero, & Melia, 2013). This confirms Venkatesh and Davis's (2000) research that social influence prompts the adoption of technology.

In addition to subjective norm, voluntariness is the extent to which a person believes that the acceptance of technology is not mandatory (Park, Rhoads, Hou, & Lee, 2014). For example, Martinez et al. (2013) found that developers are less likely to adopt new application development methods and framework when they perceive the adoption is voluntary due to the substantial amount of mental effort required in adopting new development methods and frameworks. However, image, according to Venkatesh, Morris, Davis, and Davis (2003), refers to a group's influence on an individual regarding a certain behavior or innovation that should be implemented, and implementation of this innovation or behavior by the individual could potentially improve quality and productivity within the organization in a persistent way.

Cognitive instrumental processes. Job relevance as explained in the TAM2 is an individual's perception regarding the degree to which the target system applies to his or her job (Venkatesh & Davis, 2000). It is an essential component where a likely user evaluates the impacts of using a technology or system on his or her job (Kieras & Polson, 1999). The potential user of a system or technology is more likely to accept a new technology or system if the technology is perceived to be pertinent to their job, and less likely to accept it when the technology is considered it to be less pertinent to their job.

Output quality is dependent on job relevance and measure the degree to which an individual evaluates the effect of a new technology or system (Venkatesh & Davis, 2000). In other words, it is the extent to which a user believes a new technology or system can perform the required tasks.

In addition to job relevance and output quality, result demonstrability is the visibility of the results when using a new technology or system and how it directly influences perceived usefulness (Wallace & Sheetz, 2014). Thus, result demonstrability suggests that the perceptions of the usefulness of a system or technology will be more positive if positive results are obvious and noticeable (Bhattacherjee & Lin, 2015). To put it in a different context, if the result demonstrability of a technology or system is low, the achievements attained might be attributed to work behavior rather than usage of the technology or system.

Before a technology is implemented or new system is developed, intended users should have a certain degree of understanding, beliefs, and knowledge about the technology or system. If knowledge, understanding, or beliefs are all vague, a wide range of thoughts or opinions for improvements will possibly be proposed (Zheng, Zhao, & Stylianou, 2013). Following the improvements, potential users must improve their knowledge about the technology or system. This is achieved through seminars and training sessions including major stakeholders, senior managers, technology designers, and vendors to train employees or potential users the best practices and procedures to deliver the required functionality. A review of the training follows to measure the progress and ensure it is in line with technology implementation goals. For example, Miyan, Nuruzzaman, and Sagar (2016) found that when an organization intends to introduce a new technology in the workplace, some employees in the organization might not react positively to the change or introduction, either because they are not happy with the change or because they will be required to learn new skills to use the technology.

Analysis of Related Theories

Technology acceptance models and theories have been applied in multiple domains to predict and understand users' behavior. Researchers developed a number of frameworks in IT to assess and explain user adoption of a specific or new technology. In the following paragraphs, I provide details on alternative theories used by multiple researchers to study technology adoption and acceptance in different context including but not limited to transitioning from one technology to another.

Theory of reasoned action. Ajzen and Fishbein (1980) developed the TRA using various theories and earlier studies on attitudes such as the theories of attribution, the theory of cognitive dissonance, learning theories, balance theory, and expectancy-value theories as the basis for developing TRA with the objective of predicting human behavior in diverse situation and environments. TRA was not appropriate for this study because it is limited in its capacity in predicting behavior (Ajzen, 2012), and it claims that all behaviors including behaviors related to the acceptance of technology are sets of salient beliefs. However, the TAM2 defines a set of predetermined constructs as forecasters of the intentions to accept or reject technology. Because this investigation involved examining the strategies that data managers use to transition to big data technologies from DW technology, TRA was not suitable for the study.

Theory of planned behavior. The theory of planned behavior was developed because of the limitations of TRA in addressing a person's behavior over which he or she has no full volitional control (Ajzen, 1991). As in the TRA, the behavioral intention in the theory of planned behavior remains a core construct of the theory. The theory of planned behavior was not selected because as TRA, it is limited in predicting behavior (Ajzen, 2012), and it is limited to the subjective norm and does not extend the social influences as is the case with the TAM2. The DWTAM2 was appropriate for the study because it extends to include social and cognitive processes unlike the theory of planned behavior.

Matching person and technology model. Another theory considered for this research was the matching person and technology (MPT) model. The MPT model contains a series of instruments used in assessing technology selection and decision-making as well as the results of the technology on technology users, non-users, avoiders and those reluctant (Vroman, Arthanat, & Lysack, 2015). This assessment process considers the environments in which the person uses the technology, the likely user`s characteristics and preferences, and the technology's functions and features. The MPT model was not suitable for this study because of the limitations of its theoretical foundation. The MPT model is a theory that is primarily used in developing a personalized approach to matching people with the most appropriate technologies based on their personal needs, unlike the TAM2 that has a set of predetermined constructs including social and cognitive processes to predict the reasons why a person will accept technology. Because this study's objective was to investigate the strategies used by data managers to transition to big data technologies from DW technology, the MPT model was not selected for the study.

Self-efficacy theory. Self-efficacy theory is another theory that was considered for my study. Self-efficacy theory according to Albert Bandura is the personal judgment of “how well one can execute courses of action required to deal with prospective situations (Bandura, 1982). In other words, self-efficacy is the belief that a person`s ability to influence events that affect one’s life and govern over the way these events are experienced (Bandura, 1982). Self-efficacy is a theory mostly used in psychology and has limited usage in determining behavior in technology. However, Jun, Lee, and Jeon (2014) established that self-efficacy theory is closer to the perceived usefulness, but not perceived ease of use. Self-efficacy theory was not selected for my study because it lacks the constructs to support studies that investigate technology acceptance unlike TAM2. Since the objective of this study was to examine the strategies used by data managers to transition to big data technology from DW technology, self-efficacy theory was not suitable for the study.

Diffusion of innovation theory. Another theory considered for this research is the diffusion of innovation theory, which originated in communication to explain an individual`s attitude and commitment to accept a technology based on its use in communication within an organization in a certain period (Cheng, 2019). Diffusion of innovation research seeks to elucidate the reasons behind the diffusion of some innovations through a social system at a faster rate than others (Rogers, 2003). Diffusion of innovation draws similarities to TAM2 in that both models strive to understand user credence and adoption of innovation. Diffusion of innovation was not selected for this study because unlike TAM2 which seeks to understand the reason why an individual accepts or rejects technology, diffusion of innovation is more focused on the adoption of innovation through time (Gupta, Bhaskar, & Singh, 2017). This study examined the strategies used by data managers to transition to big data technologies to DW technologies and not the time required to implement or accept these strategies.

Unified theory of acceptance and use of technology. Venkatesh et al. (2003) formulated a new model called the unified theory of acceptance and use of technology (UTAUT) to explain a user`s intentions to use an information system and the ensuing usage behavior. UTAUT holds that there are three key constructs that act as determinants of intention to use an information system or technology. The three constructs include performance expectancy, effort expectancy, and social influence (Naheb, Sukoharsono, & Baridwan, 2017). UTAUT is considered as another extension of TAM because some or all the constructs in UTAUT originate from the original TAM model (Baptista, & Oliveira, 2015). UTAUT model uses 41 independent variables for predicting intention of use and at a minimum eight independent variables for predicting behavior and includes demographic and experience as an important factor, making it more appropriate for a product or service-oriented research (MacLennan & Belle, 2014). TAM2 was more appropriate for this study as it explores the adoption of technology at a workplace where the perceived usefulness and perceived ease of the technology might affect a user`s acceptance or rejection of the technology, rather than a product or service (e.g., Bring Your Own Device).

Limitations of TAM2

Even though TAM2 has been widely used in research and remains popular for analyzing technology and information system acceptance, there has been considerable criticism on the heuristic value of TAM2, its limited explanatory power, lack of any practical value and its sense of triviality (Cheung, & Vogel, 2013). TAM was extended by Venkatesh and Davis (2000) to encompass some or most of the limitations observed in the original TAM model with close to 30 additional elements to explain additional sources of variance (Fletcher, Sarkani, & Mazzuchi, 2014).

Cheung and Vogel (2013) contended that TAM`s reliance on subjective, self-reported surveys instead of actual system usage, made the data collection approach used for TAM weak. Fletcher et al. (2014) claimed that the perceived usefulness of a technology or system might find a lot of support amongst users, but the lack of user support mechanism within the technology and poor reliability could lead the users to reject the technology or system. Certain personality traits such as extraversion and emotional stability could impact an individual`s perceived usefulness and behavioral intent to use a technology (Svendsen, Johnsen, Almas-Sorensen, & Vitterso, 2013). The accuracy and predictability of behavioral intent is lessened because of these additional determinants. Fletcher et al. (2014) reasoned that frequency and timing of the data collection process in TAM are more focused on the functioning decision-making process at a given point-in-time and not user-adoption after that initial decision. Since someone could initially accept a technology then reject it after a certain amount of time has elapsed, the focus on a given point in time is an important aspect to consider.

Usage of TAM2 in Research

TAM2 has been used widely used by various researchers in a variety of industry sectors and contexts as the conceptual framework. Park, Baek, Ohm, and Chang (2014) used TAM2 to investigate the psychological elements that may contribute to user behaviors regarding mobile-social network games. Tarhini et al. (2014) employed TAM2 to investigate the factors affecting the user acceptance of e-learning systems to enhance students` learning experience. Rana, Dwivedi, and Williams (2013) utilized constructs from TAM2 to study the adoption of e-government services. Magdaleno, de Oliveira Barros, Werner, de Araujo, and Batista, (2015) examined user acceptance of mobile applications in the banking industry to investigate mobile banking. Lee and Lehto (2013) applied TAM to identify determinants affecting behavioral intention to use YouTube for procedural learning. Park, Kim, and Ohm, (2015) employed TAM2 to investigate driver acceptance of technology (car navigation technology). Kucukusta, Law, Besbes, and Legohérel (2015) used two constructs in TAM2 namely perceived usefulness and perceived ease of use to examine online booking in Hong Kong. Nasri and Charfeddine (2012) used the TAM to investigate the factors affecting the adoption of Internet banking in Tunisia. Ovčjak, Heričko, and Polančič (2015) used the concepts of TAM and TAM2 to determine the factors impacting the acceptance of mobile data services. Chen and Chengalur-Smith (2015) used the extended acceptance model to study the factors influencing the continued use of university library Web portal. Key tenets in TAM2 were used by Alalwan, Dwivedi, Rana and Williams (2016) in their study as the foundation to study consumer adoption of mobile banking in Jordan. Kim (2006) used TAM2 to evaluate the understanding of web-based subscription database acceptance.

In their study on big data automation in strategic communication, Wiesenberg, Zerfass, and Moreno (2017) used the tenets in TAM2 (perceived usefulness and perceived ease of use) to investigate the personal experiences of communication professionals with big data, algorithms and automation, their familiarity with these new concepts as well as their competencies in the field as the relevant indicators for the usage and acceptance of these new technologies in strategic communication. Based on a quantitative survey amongst communication practitioners in Europe, of which there was a special section about big data automation, a pre-test was held with 40 practitioners in 15 European countries. They used key tenets of TAM2 investigated the implementation of practices for automated communication and represented multiple ways in which algorithms influence agencies and communication departments and demonstrated how algorithmic tools might be used in strategic communications using big data technologies. The study revealed that there is a major correlation between the day-to-day usage of big data analytics to guide the daily activities, actions and the adaptation of external algorithms and the application of algorithms for personal activities. This shows, according to the authors that big data technologies and algorithms are vital for strategic communication.

Hart and Porter (2004) used key tenets of TAM2 to study the impact of cognitive and other factors on the perceived usefulness of online analytical processing (OLAP) databases. Decision support systems range from elementary (query and retrieve) to high end (manipulation, modeling, and analysis) systems. The term “OLAP” was first used in 1993 by Dr. E.F. Codd (Sale, Patil, Thube, & Student, 2018), who in a somewhat questionable manner defined a set of rules. There now exists several definitions, some of which are quite simple like OLAP being described as computer-based enhanced multidimensional analysis (Sale et al., 2018). FASMI’s (Fast Analysis of Shared Multidimensional Information) definition is perhaps the most widely accepted definition (Loudcher, Jakawat, Morales, & Favre, 2015) that influences a users’ built-in cognitive processes and is of vital importance if the perception of technology that can supplement natural decision-making processes is to be understood. By using TAM2 to highlight the effect of user perception on perceived usefulness, and ultimately actual usage, Hart and Porter (2004) hoped that companies, as well as vendor community, may attain a more holistic understanding of OLAP user requirements, and of how they could encourage and sustain its usage. Usage of TAM and TAM2 could also provide additional insight and gains into patterns of OLAP usage in developing countries according to Hart and Porter (2004). Hart and Porter used the perceived usefulness construct of TAM2 as the dependent variable on the foundation that it can be interpreted as the antecedent of Usage intent and actual use. Hart and Porter (2004) examined perceived ease of use, result demonstrability, output quality, and job relevance as part of the research since each is positively associated with perceived usefulness. They also tested the effects of any of the demographic variables measured.

Rahim (2008) used key constructs of TAM2, namely perceived usefulness, perceived ease of use, and subjective norms in their exploratory case study to investigate the factors that affecting the acceptance of e-procurement systems – a database system used to manage an organization`s spending, by employees at an Australian City Council. Rahim used semi-structured interviews, internal documents, and the Australian city council website to gather data. Rahim (2008) in his investigation relied strongly on interviews because it provided rich insights and contextual information from the research participants (procurement manager, IT manager and several executives from the council).

Rahim`s (2008) findings suggested that the perceived ease of use, perceived, usefulness, performance expectancy, effort expectancy, social influence (which are key constructs of TAM2), facilitating conditions, vendor support, senior management support, user training and user involvement were responsible for the satisfactory acceptance of the e-procurement system by the council employees. Among the factors responsible for employee acceptance, Rahim (2008) found that system usefulness, ease of use, system reliability, vendor support and employee involvement influenced employee acceptance of the e-procurement systems the most. Vendor support provided mechanisms to maintain the stability of the system as well as providing customized training programs for council employees under the leadership of the procurement manager.

The conclusions from the study suggest that a mere introduction of a new technology or system in an organization is not sufficient for its automatic acceptance by employees. The acceptance of the technology or system depends on the usefulness of the technology or system, the ease of use, technology or system reliability, employee involvement or acceptance, senior management support and employee training. These factors are directly related to key constructs in TAM2 showing there is a major correlation between technology acceptance and TAM2. Usage of TAM2 in different research studies involving such an extensive and diverse collection of organizations and businesses supports the use of TAM2 in evaluating the acceptance and adoption of big data technologies.

Data Warehousing

Data are an asset that has always been part of every organization. The advent of relational databases during the last 30 years has enabled organizations of various sectors to gain business insight by storing and processing their data out from a central repository known as a DW (Berg, Seymour, & Goel, 2013; Jukić, Sharma, Nestorov, & Jukić, 2015). Since then, Data Warehousing has been providing business value to organizations. Data Warehousing and Business Intelligence became more popular as the technology matured during the 1990`s and 2000`s (Chen, Chiang, & Storey, 2012). The primary purpose of DW is to integrate data from many different applications and systems in a formal and well-architected manner rather than having to pull data together haphazardly as was the case before the advent of Data Warehousing (Jukić et al., 2015).

DW architecture. A DW allows data managers and other senior managers to have access to huge quantities of an organization's operational data for diverse business needs like making strategic decisions, determine business trends, promotion of anticipation and for planning purposes (Jaber, Ghani, Suryana, Mohammed, & Abbas, 2015). A DW is a system that facilitates the extraction of object-oriented, integrated, time-varying, and non-volatile business or company data stored in relational database management systems (Jaber et al., 2015). Relational database management systems are used to store data needed for online transaction processing, unlike DW, which is exclusively used for OLAP, decision making and report generation (Kour, 2015). Theoretically, data in traditional relational databases are organized in one-dimensional view. Each record within the database has information organized in fields (Jaber et al., 2015; Kour, 2015).

Data managers and business analyst get information from data stored in a DW and analyse this data to make critical business decisions and measure business performance (Jaber et al., 2015). Having a DW in an organization gives the following advantages:

- A DW enhances business productivity since it can gather information quickly and efficiently
- A DW helps manage business information and customers by providing a consistent view of customers and items.
- A DW helps in prize reduction by tracking trends, and patterns over a long period of time in a reliable and consistent manner.

An effective and efficient DW design requires a thorough understanding and analytics of business needs. Every DW designer has its own view with respect to the design of a DW (Renu, Ashish, & Nitin, 2013). There are four different views:

- The top-down view – allows for the extraction of relevant information required for the DW (Renu et al., 2013).
- The data source view – represents information that is being captured, stored and managed by the operational system (Renu et al., 2013).

- The DW view – includes the fact and dimension tables. This view denotes information stored inside the DW (Renu et al., 2013).
- The business query view – the view of the data from the end-user's viewpoint (Renu et al., 2013).

In general, a DW adopts a three-tier architecture (See Fig 5). The three tiers of DW architecture include:

- Bottom Tier – The bottom tier of a DW architecture is the DW database server which is a relational database system. Data is fed into the bottom tier using back-end tools and utilities. These backend tools and utilities are responsible for the extraction, cleaning, loading, and refreshing of the data (Renu et al., 2013).
- The Middle Tier – The middle tier contains the OLAP server that can be implemented using one of the following methods (Renu et al., 2013):
    - By relational OLAP (ROLAP) which is simply an extended relational database management system. The relational OLAP has as role to map the operations on multidimensional data to standard relational operations
    - By multidimensional OLAP (MOLAP) model, which is responsible for the direct implementation of the multidimensional data and operations.
- Top Tier – This tier represents the front-end client layer. This layer holds the tools used for querying, reporting, analysing and data mining processes.

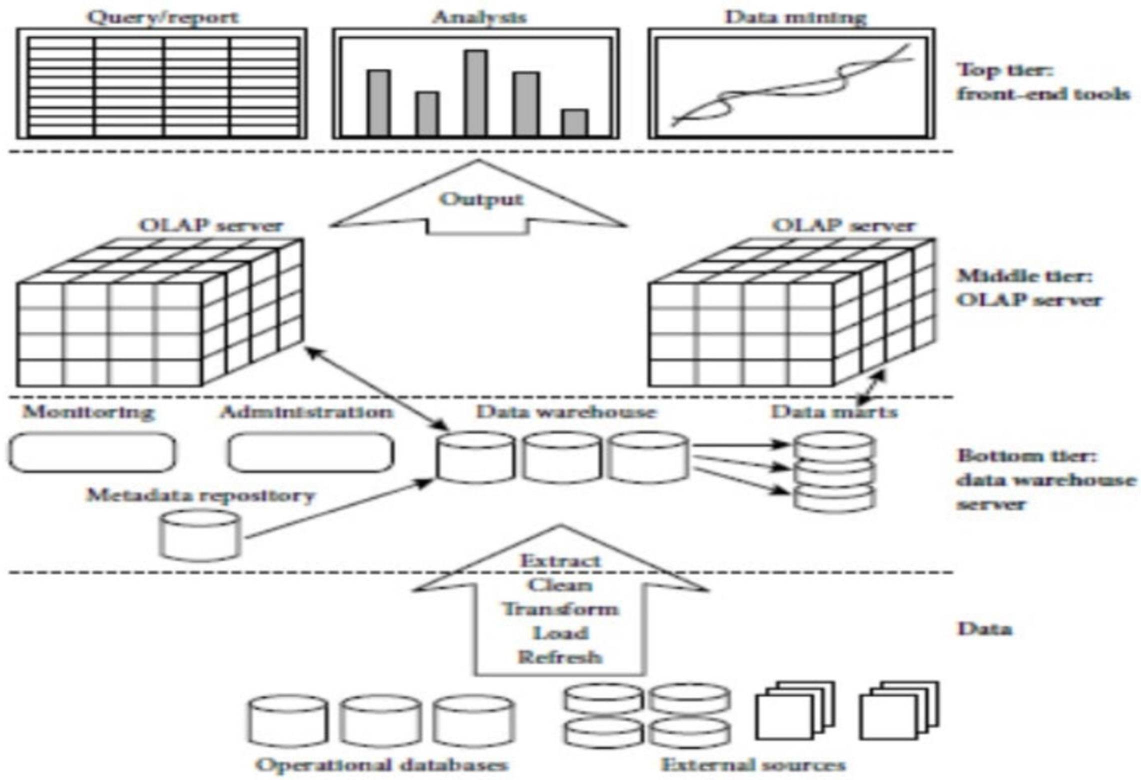

Figure 4. Architecture of a DW—Three tier architecture. Adapted from "Three-tier architecture of DW," by B. Renu, A. Ashish, and G. Nitin, 2013, *International Journal of Latest Technology in Engineering, Management & Applied Science*. *Copyright permission to use for scholarly and educational purpose without additional approval.*

Big Data Analytics

The advent of cloud computing and the Internet-of-things has led to an exponential increase in interest in Big Data and Big Data analytics (Kim, Trimi, & Chung, 2014), One of the major catalysts of this exponential growth was Google`s adoption of MapReduce, which led to further developments in the big data arena (Gopalani, & Arora, 2015). Additionally, the development and subsequent deployment of technologies like Apache Hadoop from the Apache foundation has also contributed greatly to increase interest in big data, as well as given the opportunity to organizations of various sizes and from different sectors to capture and analyze an unprecedented amount of datasets they simply could not afford to analyze due to the restrictions and limitations on traditional relational database management systems capacities (Chen, 2015; Elgendy & Elragal, 2014). Big data is slowly but surely paving its way into various sectors including but not limited to governments, health, e-commerce, finance, retail, insurance, etc. (Kim et al., 2014). Supporting evidence of this permeation is the overwhelming amount of data available from a variety of sources like, web applications, emails, streaming data, radio-frequency identification, and data generated by sensors (mobile devices, GPS, etc.), which build up at an increasingly upward scale prompting organizations to develop plans for the exponential growth of their data (Waller, & Fawcett, 2013). As technology becomes more accessible and data becoming more available in massive amounts, most organizations understand that if they are able of capturing all the data that streams into their businesses, they can get significant value from the data by applying analytics to it.

Big Data Analytics, is the use of advanced techniques, mostly data mining and statistical formulas to operate on big data sets to find hidden patterns in (big) data (Wang, Gunasekaran, Ngai, & Papadopoulos, 2016). The size of data (big data) is constantly increasing, ranging from a few dozen terabytes to many petabytes; zetta-bytes; yottabytes in a single data set (Hormann, & Campbell, 2014). Consequently, the most apparent difficulties related to big data initiatives include the capturing, storage, analyzing and visualizing (Mahrt, & Scharkow, 2013). Analytics performed on large data sets uncovers trends and leverages business change and hence profitability, however, the larger the data set, the more problematic and challenging it becomes to manage (Sumbal, Tsui, & See-To, 2017). A thorough understanding of big data analytics framework, process, and the technologies involved is necessary for a successful transition from data warehousing.

Characteristics of Big Data

This section covers the characteristics of Big Data as well as its importance. There are dozens of definitions available on the Internet for Big Data. Some of the definitions have referred to Big Data as the 3,4 or 5 V`s (Kim et al., 2014) for Volume, Variety, Velocity, Veracity, and Value. Other researchers and analysts have defined big data in a more technical context like the definition from Edd Dumbill (2012), an author and analyst at O'Reilly Media: "Big Data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn`t fit the structures of your database architectures. To gain value from this data, you must choose an alternative way to process it." While such definitions might be accurate, they miss the true value of Big Data. Big Data should not be measured by the amount of processing power or storage space it consumes but rather be measured by the size of its impact (Provost & Fawcett, 2013; Waller & Fawcett, 2013). Volume refers to the size of data, which comprises all types of data produced from different sources (Berman, 2013). Variety represents the different types of data collected through sensors, social networks, and portable devices like smartphones. This type of data includes images, video, text, audio, and data logs, most of which are in an unstructured format (O'leary, 2013). Velocity refers to the speed or the rate at which data is changing or how often the data are created (Berman, 2013). Value and Veracity are considered the most important aspects of big data; they signify the truth-worthiness of the data and the process of discovering hidden values from it through various types of algorithms and analytical tools (Chen, Mao, & Liu, 2014; Özköse, Arı, & Gencer, 2015).

The primary attribute of big data is data volume (Jee & Kim, 2013). Most organizations and businesses quantify big data in terabytes or petabytes, but big data could also be quantified by the number of records, transactions, tables, or files (Hu, Wen, Chua, & Li, 2014). Additionally, the one feature of big data that makes it “big” is the fact that it is produced from a variety of sources, including but not limited to logs, clickstreams, sensors, social media, etc. Using these many sources for analytics purposes means common structured data stored in relational databases is now amalgamated with unstructured data, such as text, audio, video, and human language, and semi-structured data, such rich site summary (RSS) feeds and eXtensible markup language (XML). Furthermore, multi-dimensional data from DWs can be added to supplement historical context to big data. Thus, variety is much as important as a volume in big data.

Big data is also described using velocity and speed, which is the frequency of data delivery or frequency of data generation. Streaming data is considered as the leading edge of big data since it is collected in real-time from social media platforms, online web portals, and other connected devices (Chen et al., 2014; Raghupathi, & Raghupathi, 2014;). These unique characteristics of big data discussed in this section are what makes big data unique but understanding how these features relate to data stored in DWs is important when transitioning to big data from data warehousing.

Big Data Architectures

Big Data and Data Warehousing based on architectural diagrams look quite similar, but there exist some significant differences between them (Assunção, Calheiros, Bianchi, Netto, & Buyya, 2015; Salinas & Lemus, A. C. N., 2017). (see fig 3). With traditional Data Warehousing, data is brought in from an organization`s source systems into the DW, typically in a batch manner from a small number of applications (Storey, & Song, 2017). Big Data; however, is built around the paradigm called the 3,4, or 5 V`s depending on the definition and context (Gandomi, & Haider, 2015; Storey, & Song, 2017). The first V is Volume where there now exists an infinite number of sources and capacity because of the new generation of technology in use (Gandomi, & Haider, 2015; Storey, & Song, 2017). The second V is Velocity: Unlike DW where data is brought in batch manner, big data which consists of massive amounts of data from various sources can be incorporated both in real time and batch mode not only from existing sources, but also from new data sources extremely quickly (Gandomi, & Haider, 2015; Storey, & Song, 2017). The third V is Variety (Gandomi, & Haider, 2015; Storey, & Song, 2017)., DWs are built around structured data (numbers, text fields, and dates), whereas with Big Data, data is structured in many ways (Assunção et al., 2015). The fourth V is Veracity – the truth worthiness of data and the fifth V is the Value of the data (Bello-Orgaz, Jung, & Camacho, 2016).

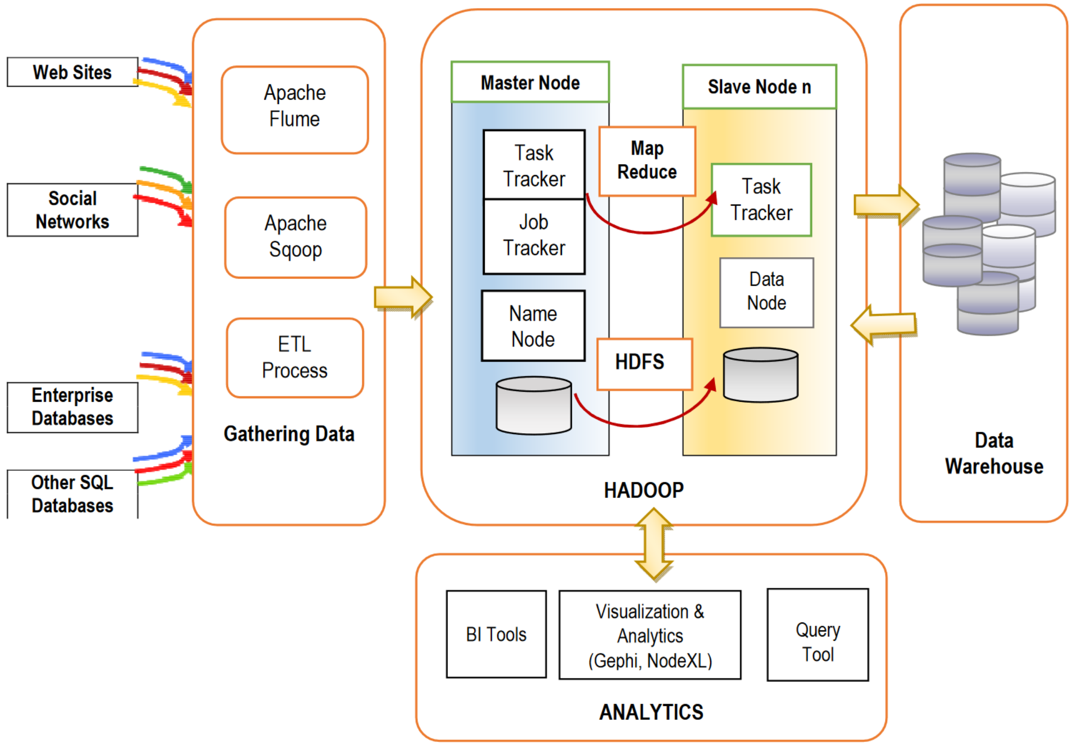

Figure 5. Big data architecture. Adapted from "Big Data in Business Environment," by L. Banica and A. Hagiu, *Scientific Bulletin-Economic Sciences*, 2015. *Copyright permission to use for scholarly and educational purpose without additional approval.*

The analytics at the disposal of organizations and businesses nowadays are incredibly valuable for giving insights into not only the past (as it is typically with Business Intelligence and Data Warehousing), but as well as the present, future, and the unknown (Bedeley, Ghoshal, Iyer, & Bhadury, 2016; Sun, Strang, & Firmin, 2017). Analytics has developed into becoming a catch-all term for a variety of data-driven insights. According to Zakir, Seymour, and Berg, (2015), data analytics is a means by which value is extracted from massive volumes of data, and it drives new market opportunities and at the same time maximizes customer retention. There are distinct categories of analytics, each for different purposes. Most seem to have settled on four various categories of analytics; descriptive analytics, predictive analytics, discovery analytics and prescriptive analytics (Adnan, 2015; Evans, 2015). Business Intelligence as practiced since the 1980s has mainly focused on descriptive analytics (Chen et al., 2012). However, modern analytics with Big Data needs all these categories.

Business Intelligence

Business intelligence provides organizations the ability to "slice and dice" their data, i.e. organizations can look at their data in many different ways where certain key metrics can be identified then look at the data by various dimensions like time, date, by product, by customer, etc. as long as the DW has been set up that way (Gonzales, Wareham, & Serida, 2015).

Because DWs are not architected to support other forms of analytics other than descriptive analytics, many organizations stop their investigations into data with descriptive analytics, because the descriptive analytics turns out to be the reports that have run the organizations for many years. Therefore, the traditional business intelligence only delivers some of the critical insights needed, and hence, organizations need to look at different techniques to get access to the predictive and descriptive analytics (Gonzales et al., 2015). Big data technologies offer the capabilities to perform predictive and descriptive analytics but not without challenges.

Olshannikova, Ometov, Koucheryavy, and Olsson, (2015) outlined the list of existing big data analytics methods, that include (in alphabetical order): A/B testing, Association rule learning, Classification, Cluster analysis, Data fusion and data integration, Ensemble learning, Genetic algorithms, Machine learning, Natural Language

Processing, Neural networks, Network analysis, Pattern recognition, Predictive modelling, Regression, Sentiment Analysis, Signal Processing, Spatial analysis, Statistics, Supervised and Unsupervised learning, Simulation, Time series analysis and Visualization. Understanding all or some of these analytical methods is a daunting task, that`s why Dubey and Gunasekaran, (2015) stressed that there is a need for organizations and institutions to revamp teaching methodologies and invest in Business intelligence and Analytics education that would be interdisciplinary and would cover critical analytical and Information Technology skills, and communication skills required in big data initiatives.

The Cultural change that would span across the entire organizations and urge employees (especially those working with big data) to be more pragmatic and efficient in data management as they incorporate them into the decision-making process, should accompany any investment made to increase the analytical knowledge (De Mauro, Greco, & Grimaldi, 2015). The rise of new specialized professional entities called algorithmists that would have a mastery in mathematics, statistics and computer science and can act as unbiassed auditors to review and approve the validity and accuracy of big data predictions will foster development in big data research (Johnson, & Weeks, 2016; Loui, 2016).

Moreover, Power (2016) and Schoenherr, and Speier-Pero, (2015) describe a data scientist as an individual with fundamental skills capable of analyzing and interpreting complex data sets in digital format to extract value and assists businesses in the decision-making processes. These skills for data scientist (Enterprise business and process decision making, data management, analytical and modelling tools) are not available in sufficient amounts to meet the current and future demands (Lyon, & Brenner, 2015; Schoenherr, & Speier-Pero, 2015). According to the McKinsey Global Report (Lyon, & Brenner, 2015) the United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million data managers and analysts to analyze big data and make informed decisions based on their findings. By analyzing the competency gaps, effective teaching methods can be created to fill them both for existing and future data managers as well as data practitioners (Schoenherr, & Speier-Pero, 2015).

Furthermore, the ability by businesses to make informed decisions because of big data analysis keeps improving and changing as big data technologies expand as the latter infer that there is a shift from logical, causality-based thinking to the acknowledgement of existing correlation between data sets (De Mauro et al., 2015). The usage and applications of insights resulting from the analysis of big data by businesses, organizations, institutions and government agencies provides for an acclimatization to a new culture of decision-making based on big data analysis (Ur-Rehman, Chang, Batool, & Wah, 2016; Wang et al., 2016) and the development of algorithms and methods used in data analytics (Dubey, & Gunasekaran, 2015), both of which continue to be improved thereby providing opportunities for future research in the field of big data.

Understanding the limitations of methods involved in big data initiatives and plausible methodological issues is an essential resource for businesses as well as organizations that are keen to use big data for decision-making based on data: for instance, every prediction based on data should be backed by valid confidence intervals in data so as to avoid any false sense of precision or exactitude that the apparent intricacy that some big data applications can suggest (De Mauro et al., 2015). Data scientists and analysts should also endeavor to avoid model' overfitting that could potential pave the way for apophenia i.e. the penchant of humans to "see patterns where none actually exist simply because enormous quantities of data can offer connections that radiate in all directions" (Prinsloo, Archer, Barnes, Chetty, & Van Zyl, 2015, p. 287).

In summary, making sense out of huge quantities of data i.e. big data requires proficiency in specific data analysis techniques, knowledge of their strengths and limitations and the adjustments in cultural behaviors and tendencies to informed decision-making processes in most cases still have to be built.

What Fuels Big Data: Information Generation and Utilization

This section covers one of the fundamental or central motives for the existence of the big data phenomenon. One of the main reasons that has fueled the existence of big data is the current magnitude to which information can be generated from a variety of systems including gadgets and things and made available to individuals, businesses, government agencies and non-governmental organizations (Abbasi, Sarker, & Chiang, 2016; Bello-Orgaz et al., 2016; Gandomi, & Haider, 2015). The process of converting continuous, analog data into distinct, digital and machine-readable format, popularly known digitization, gained prevalence and acceptance with the first "mass digitization" (Murrell, 2017). Mass digitization is the undertaking and pursuit to convert or transform complete printed book libraries into digital collections by using optical character recognition software so as to minimize to the least human intervention (Christy et al., 2017). The Google Print Library Project3 is considered as one of the most popular attempts of mass digitization, which started in 2004 and was aimed at digitizing more than 15 million volumes of printed books stored in multiple university libraries including but not limited to the libraries of Harvard, Stanford and Oxford (Murrell, 2017). More recently, De Mauro et al. (2015) proposed a sharp differentiation between digitization and its subsequent phase, datafication, i.e. placing information in quantifying format in a way that it can be organized meaningfully and analyzed. The fundamental difference between digitization and datafication is that digitization allows the transformation and storage of analog information in a more convenient and accessible digital format while datafication is focused on organizing the digitized version of the analog information so that meaningful insights could be generated from the data which would not have been possible if the data remained in its analog form (De Mauro et al., 2015). The value of datafication was proven in the previous cited Google mass digitization project, where researchers where demonstrated that the efforts allowed them to provide insights on lexicography, collective memory, technology adoption, censorship, historical epidemiology, the evolution of grammar, and the pursuit of fame and etc. (De Mauro et al., 2015; Pechenick, Danforth, & Dodds, 2015; Pettit, 2016).

Thanks to the extensive availability of devices in all shapes and forms both mobile and stationary and connected to each other and digital sensors seamlessly, digitization and datafication have become prevalent or omnipresent phenomena. Digitization is enabled through digital sensors while seamless connections between the sensors and sensors and other devices enable the aggregation of data, and, hence, allows data datafication (De Mauro et al., 2015). Cisco estimated that between the year 2008 and 2009, the number of existing connected devices overtook the number of people living on earth (De Mauro, Greco, & Grimaldi, 2016; Rathore, Ahmad, Paul, & Rho, 2016). Similarly, De Mauro et al. (2015) claimed that by the year 2020 there will be 26 billion devices on earth, which will be equal to more than 3 devices on average per person. The emergence of technologies like the internet of thing – the ubiquitous presence of a variety of objects or ‘things’ (including sensors, smart phones, radio-frequency identification, tags, Ipads, actuators, etc.), which connect seamlessly with each other, have equipped businesses agencies with massive information assets from which they can improve IT practices, create new business models, and reduce associated costs and consumer risks (De Mauro et al., 2016).

The seamless connection of different devices and ‘things’ means data generated from these devices and things will have different characteristics mainly in type. Structured data which is comprises mainly of numeric and traditional text data (Guha et al., 2016) is combined with unstructured data which is mainly image, video, audio text and human language and semi-structured data mainly RSS feeds and SML data (Gandomi, & Haider, 2015) and semi-structured data such as RSS feeds and XML data (Yan et al., 2017). The diversity of data types is considered as one the major challenges faced by organizations and government agencies as they explore the massive amount of data in their position in search of meaning (Hashem et al., 2015).

Working with Big Data: Technology

This section addresses common technologies associated with big data and how they enable big data implementations. Even though the term big data is used to describe a wide range of concepts, it is commonly associated with the technology that facilitates or enables its utilization (De Mauro et al., 2015). The magnitude of the size of data and the complexity associated with operations required for handling and processing the data entail rigorous memory, computational performance and storage prerequisites (De Mauro et al., 2015). The most related query associated to “Big Data” is “Hadoop” according to Google Trends, indicating that Hadoop is the most prominent technology associated with big data (De Mauro et al., 2015). The Hadoop distributed file system (HDFS) is an open source framework that facilitates the distributed processing of substantial amount of data stored across multiple nodes of machines using definite computer programming models (Landset, Khoshgoftaar, Richter, & Hasanin, 2015; Zaharia et al., 2016). HDFS has the following component; (1) the file system HDFS, that enables access to data distributed across multiple machines without having to deal with the complexity associated with their distributed nature, (2) MapReduce, a data processing engine designed for the implementation of distributed and parallel algorithms in a well-organized and proficient manner (Landset et al., 2015). Both HDFS (De Maio, Fenza, Loia, & Orciuoli, 2017) and MapReduce (Pulgar-Rubio et al., 2017) are the results of concepts that were originally initiated by Google in 2003 when Google published its paper on the Google File System and MapReduce Framework (Casado, & Younas, 2015), and were later developed as opensource projects within the Apache`s framework. This supports the argument that

Google is central in the initiation and development of the current technologies around big data.

The Hadoop framework through the Apache foundation has been extended extensively from its original implementation and now contains multiple modules and libraries that are compatible with the HDFS and MapReduce (Landset et al., 2015). This allows IT professionals especially data managers to extend its applicability to the various needs of analysis, performance management, coordination and workflow and activity design that are typical in big data initiatives.

Due to the distributed nature of information, specific technological efforts are required for transmitting huge quantities of data as well as for monitoring purposes of the overall system performance and health status using special techniques designed for big data benchmarking (Aydin, Hallac, & Karakus, 2015). Another important element of technology is the fact that bigger quantity of data can be stored on smaller physical devices (Wong, & Salahuddin, 2015). Although Moore`s law suggests that storage capacity increases in exponential manner over time (Lucas, Kelton, Sanchez, Sanchez, & Anderson, 2015), continuous research and development efforts are required to keep up the rate at which the size of data increases (De Mauro et al., 2015; Wijndaele et al., 2015) especially with the increasing share of byte-hungry data types like videos, audio and images (De Mauro et al., 2015).

Methods Involved in Big Data Transformation

This section discusses the methods used in the transformation of huge quantities of data. The continuous desire to fully comprehend and grasp value out of the extensive quantities of data available, explicit processing methods that go beyond ordinary statistical techniques and backed by industry standards are required (De Mauro et al., 2015; Ur-Rehman et al., 2016). The knowledge required to comprehend such methods, their potential as well as their limitation requires specific skills sets that are scarce in today`s job market and the lack of trained talent represents a major obstacle for realizing the full potential of big data analysis (Schoenherr, & Speier-Pero, 2015).

The quantity of data generated from websites, sensors, web applications, emails, streaming data, radio-frequency identification, satellites and data generated by sensors (mobile devices, GPS, etc.) overpowers or incapacitates traditional data analysis approaches, which as a result has paved the way for the development of new types of programming models such as MapReduce (Begam, & Sandhya, 2017; Gopalani, & Arora, 2015). MapReduce is a new type of distributed programming model that makes it extremely easy to tun high-performance parallel programs on huge quantities of data (big data) using commodity hardware extremely easy (Aydin et al., 2015). MapReduce applications are made of two main modules, the mapper(s) and the reducer(s), which are mainly user-defined programs implemented using the MapReduce API (Aydin et al., 2015; Heintz, Chandra, Sitaraman, & Weissman, 2016). Hence a MapReduce job is made up numerous processes like mapping and reducing codes, splitting and distributing data, and saving the results to a distributed file system. Depending on the volume and type of the data, analyzing and transforming data using MapReduce may require running more than one job (Heintz et al., 2016). The jobs can be chained in a more complex scenario or can be independent of each other. Figure 6 shows a visual representation of how

MapReduce paradigm works: a master node controls MapReduce jobs which are split into two separate functions called Map and Reduce. The Map function splits the input data into a group of key-value pairs and the output generated from each map function is sorted by their key. The values are then merged into the final result by the Reduce function (Aydin et al., 2015).

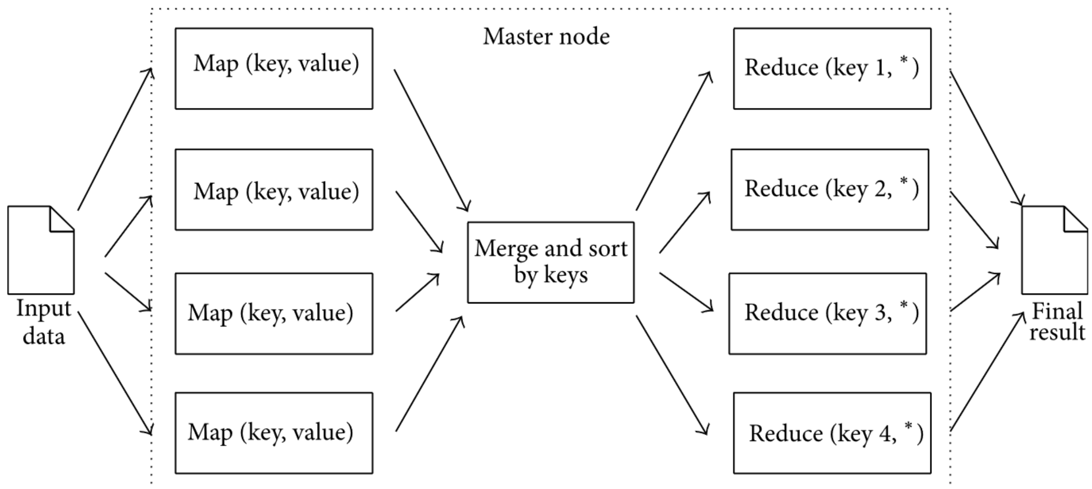

Figure 6. An overview of the map and reduce steps. Adapted from "Architecture and Implementation of a Scalable Sensor Data Storage and Analysis System Using Cloud Computing and Big Data Technologies," by G. Aydin, I. R. Hallac, and B. Karakus, 2015, *Journal of Sensors*. *Copyright permission to use for scholarly and educational purpose without additional approval.*

MapReduce has since evolved and has been implemented in open source projects like Hadoop (Landset et al., 2015). Hadoop is considered the most popular implementation of MapReduce and is used in different projects involving all areas of big data (Aydin et al., 2015; Landset et al., 2015). The Hadoop ecosystem also provides additional big data tools like the Hadoop Distributed File Systems, used for data storage in clusters, HBASE (Lee, Shao, & Kang, 2015), Google's Big Table non-relational database system, Pig (Sahoo et al., 2016), and engine for the execution of parallel data flow in Hadoop, Hive (Yu, Dou, Zhu, & Wang, 2015) a DW-like software on Hadoop, and Mahout (Bagchi, 2015), a data analysis and recommendation tool on Hadoop. Main advantages of the Hadoop MapReduce framework are scalability, flexibility, cost-effectiveness, resilience to failures and speed (Aydin et al., 2015).

The Hadoop uses the Hadoop Distributed File System (HDFS), and stores data on a block-by-block basis. The files are first split into blocks and then distributed to all nodes in the Hadoop cluster (Landset et al., 2015). Each block in the HDFS has a default size of 64 MB unless the user modifies the block size (Landset et al., 2015). Hadoop splits any file that is larger than 64MB are moved to a new block from a line where the file size does not exceed the maximum block size and rest of the lines are not exceeded (Aydin et al., 2015). Hadoop utilizes a master-slave architecture in the cluster where the master nodes are called the Name Node and Job Tracker, while slaves nodes are the Data Nodes and Task Tracker (Aydin et al., 2015). Input data is partitioned into blocks and are placed in the Name Node which also stores the metadata of the blocks so that the HDFS can retrieve block information and know which block is stored on which Data Node (Aydin et al., 2015). And if one for any reason one of the nodes fails, it does not affect job completion because HDFS stores the replicas of those blocks (Aydin et al., 2015). The execution of the processes is tracked by the Job Tracker and Task Tracker. The Job

Tracker has an analogous relation with Name Node and Data Node. The Task Trackers has the responsibility for running the tasks and sending the resulting messages to the Job Tracker. The Job Tracker communicates with the Task Tracker and keeps a record of all running jobs or processes. If the Job Tracker notices that the Task Tracker has stopped or is unable to complete its assigned tasks, it schedules the missing executions on another functional Task Tracker (Aydin et al., 2015).

Other big data transformation tools that have MapReduce-like characteristics are Apache Spark and Shark, HaLoop, and Twister. These systems are more suited for iterative statistical and complex algorithms inside a MapReduce-like programming model but are still lack the data management features available in relational data systems (Aydin et al., 2015).

The Impact of Big Data on Society and Businesses

The degree to which a business, organization, government, and society, in general, is impacted by big data is often depicted through the success stories and anecdotes of technology implementations and methods (Jordan, & Mitchell, 2015). These success stories when proceeded by proposals of new methodologies and technical improvements represent a reasonable contribution to the creation of knowledge in big data research. The persistent or extensive nature at which digital information is currently produced and made readily available leads to the development of many applications in several business sectors and scientific fields that can be sometimes very distant from each other. Most often than not, the same data and techniques have been applied to problem-solving in different and distant domains. For instance, scientists used correlation analysis on logs from Google searches to forecast influenza epidemics (Yang, Santillana, & Kou, 2015) as well as unemployment levels (Baker, & Fradkin, 2017) and inflation (Li, Shang, Wang, & Ma, 2015). Existing big data technologies and applications are expected to improve and grow as technology advances: therefore, their systematic depiction constitutes development area for researchers and IT professionals willing to contribute in the scientific evolution in the field of big data.

Big data can also have a negative impact on society. In fact, some researchers have raised some concerns arising from the rapid technological advancement of big data, the first of which is privacy concerns (Abbasi et al., 2016; Eastin, Brinson, Doorey, & Wilcox, 2016). Although the mining, cleaning and analytical process involved in any big data initiative will normally include multiple people assigned to different tasks, it is difficult to envisage that the consequences of working with and benefiting from that data will not impact a single individual in an unexpected manner. Although it has been proven that an individual's identifiable information from the data set can be circumvented through the process of anonymization, it is a process that is extremely hard to be fully guaranteed as the it can be reversed through the process of de-anonymization (Ji, Mittal, & Beyah, 2016; Panackal, & Pillai, 2015). The ability to predict future events and actions from individuals all made possible through the analysis of behavioral patterns, also poses an ethical issue especially when it comes to the protection of free speech or free will in the future (Kosinski, Wang, Lakkaraju, & Leskovec, 2016).

Other concerns that should be taken under consideration is the accessibility to information: if a business or organization have exclusive control over the sources of data, it can become an abuse of dominant position and restrict competition by exercising the power of monopoly. For instance, access to the vast amount of social media data is limited only to social media companies, and they have full control over the access rights to that data (Abelson et al., 2015; Batrinca, & Treleaven, 2015). The divide between data-lacking and data-rich companies can result in the creation of a new digital dominance (Hilbert, 2016) that can slow the pace of innovation in big data research. New policies will have to be created to address some of these issues as data likely to transcend to become a new dimension to consider within antitrust regulations (De Mauro et al., 2015).

The impact of big data is not only limited to society but to businesses as well: the need for acquiring new skills and talent as well as technology in order to be competitive in a data-driven market infers that organizations need to reconsider their business processes as well as the overall organization of the business (Barrett, Davidson, Prabhu, & Vargo, 2015). The transformation of data into a competitive advantage (Prescott, 2016) is the reason big data has become such an impactful revolution for most businesses today.

Trends and Challenges in Big Data Analytics

Technological developments and evolution have led to the advent of different types of devices of different sizes connecting seamlessly to each other to produce a network of connected devices (Li, Da Xu, & Zhao, 2015). These interactions between devices mean an increased wealth of data flowing in and out of an organization on a daily basis. As such, there has been an increased urgency for not only a faster and efficient technology, but as well as for more efficient ways of analyzing such data (Bello-Orgaz et al., 2016). Just having huge amounts of data on hand in various formats and structures is no longer sufficient to make proficient business decisions at the right time (Kitchin, 2014).

The trends, opportunities, and challenges that big data analytics offer to businesses, organizations as well as government agencies have been well addressed in the current literature. With more connected devices (Internet-of-things), the evolvement of social networks, technological developments in processing power, and the increasing speed in networks, businesses are more equipped than ever to exploit big data in many ways (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016). Several researchers have explored the theoretical and engineering trends in the domain of big data analytics (Chen, 2015; Hashem et al., 2016; Hazen, Boone, Ezell, & Jones-Farmer, 2014). These studies concluded that technological advancements are directly responsible for the increased research and utilization of big data. For instance, the advent of powerful computer systems that have enabled video analytics, software, defined networks (SDNs) and ads placements have all contributed to research and exploitation of big data (Hashem et al., 2016; Ma et al., 2015). Additionally, the development of systems capable of supporting multi-connectors like Hadoop has played a big role in analytics on large datasets (Chen, 2015). Moreover, IT practitioners and researchers continue to work to enhance several algorithms used for data mining, that will help in large-scale multifaceted or complex analytics, like video and real-time analytics (Chen, 2015). A similar analysis of current trends in big data analytics and cloud computing from a database perspective has been achieved by Fernández et al., (2014).

Singh and Singla (2015) in their study investigated the current adoption trends, best practices and leading tools and technologies of big data storage and processing. They argued in their study that customer activity and insights obtained from online contents are key to the success of big data initiatives. Also, Dutta and Bose (2015) investigated the application of big data technologies in the Ramco-Cements-Limited Big Data project to understand and evaluate the challenges that come with big data projects. They concluded that for a successful big data implementation the following conditions need to be met; 1) a clear understanding of the business problem, 2) a well-planned and detailed project map with a step-by-step implementation, 3) a cross functional project team, 4) the adoption of the company of innovative visualization techniques, 4) active involvement and patronage from senior management, and 5) a data driven decision making culture.

Transition and Summary

I took the following steps in Section 1. Section started with a background of the problem, the purpose statement, the research question, conceptual framework and the significance of the study. I reviewed the literature next to provide the reader with exhaustive information about the current and past research on big data technologies. The literature review in this study contains a comprehensive description of the conceptual framework TAM2 that I used. TAM2 provided the perspective from which the perceived usefulness and perceived ease of use of big data technologies can be explored. TAM2 presented the framework and could provide an insight into the strategies that some data managers used to transition to big data technologies from DW technology as well as the reasons why some data managers lack the strategies to transition to big data technologies from DW technology. Section 1 ended with the trends and challenges associated with big data initiatives.

In section 2, I described the role of the researcher, the participants, as well as the research method and design. The description of the study populace, the sampling method, the data collection and data analysis technique follow. I conclude section 2 with a discussion on the validity and reliability of the study and transition into section 3 which will include the results of the study following the analysis of the data collected.

Section 2: The Project

Section 2 contains a detailed discussion on the approach of the project. The next sections reemphasize the purpose and the role of the researcher. The Research Design section is focused on applicable methods and design, and the population and sampling is focused on the area of senior healthcare IT leaders. Ethical research was used to identify the boundaries essential for participant protection and adherence to Walden University Institutional Review Board (IRB) requirements. Finally, data collection, analytics, and validation are discussed on the error types and compensation method.

Purpose Statement

The purpose of this qualitative multiple case study was to explore the strategies data managers use to transition to big data technologies from DW technologies. The target population consisted of data managers from 5 mid-sized organizations located in Germany that have transitioned to big data technologies from DW technologies. Transitioning to big data technologies could have a positive social impact by facilitating information exchange by providing new interactive platforms designed to connect citizens and service users with organizations that are conducting interventions or representing their needs.

Role of the Researcher

The role of the researcher in a qualitative study is to attempt to access the thoughts and lived experiences of the participants by initiating and welcoming them to a participative and process-oriented evaluation, focusing on the best method to convey research validity to participants and minimizing the likelihood of autonomy infringement

(Galdas, 2017). It is also the responsibility of the researcher to create interview questions that will allow creativity and flexibility during the interview process to ensure that each participant's lived experiences or story is uncovered entirely (Dikko, 2016; Sorsa, Kiikkala, & Astedt-Kurki, 2015). The researcher while probing individual participants must remain flexible and acknowledge participants responses without any preconceived biases or notions (Roulston & Shelton, 2015). Before the interview starts, the researcher can establish interview expectations for both researcher and participants by using initial rapport-building discussions (Arsel, 2017). The initial rapport-building makes the researcher and participants more at ease with one another, fostering a smooth flow of the interview (Arsel, 2017). Therefore, I initiated an initial rapport-building with participants to establish interview expectations.

Additionally, as the researcher, my personal experience with the topic under investigation is my rationale for undertaking this study. I have over 10 years of professional experience working in various roles in IT across different organizations. I started as a Windows Server Administrator, then Database Administrator and Linux Administrator, Security Analyst, DevOps Engineer, Cloud Architect and Data and Artificial Intelligence professional. Because honesty when conducting qualitative research is important for integrity and ethical discipline (Rolbiecki et al., 2017), I had neither a professional nor personal relationship with the chosen participants.

I also reviewed the Belmont Report (United States Department of Health and Human Services, 1979) to familiarize and understand the basic principles and guidelines required for the protection of human subjects in research studies. The Belmont Report identified three core principles:—the respect for human subjects, beneficence, and justice—as the fundamental principles significant to ethical research (Hammer, 2016). The Belmond Report also identified three primary areas of application: informed consent, risks assessment and benefits, and selection of subjects (Hammer, 2016). It is important for researchers to comprehend ethical standards to appropriately handle human subjects in research (Cugini, 2015). A clear understanding of the principles outlined in the Belmont Report helps r participants avoid any form of pressure and improper influences and guide the participants in understanding any risks involved in the research (Botkin et al., 2015). I completed the Protecting Human Research Participants training offered by the National Institutes of Health (see Appendix A), ensuring that I adhered to the principles outlined in the Belmont Report. However, the possibility of bias exists in all studies (Roulston & Shelton, 2015). The researcher in a qualitative study needs to be subjective to understand and interpret data (Botkin et al., 2015; Roulston & Shelton, 2015). But I followed the three fundamental ethical principles of the 1979 Belmont Report, and my personal views were not incorporated into the data collection process, as I was mindful of my values, susceptibilities, and subjectivity to avoid bias.

I also recorded and transcribed the interviews with the participants to mitigate bias in this study. Further, I asked open-ended, nonleading questions to avoid any form of influence on the participants during the interview process to effectively avoid bias in the study (Brubacher, Powell, Skouteris, & Guadagno, 2015). Recording thoughts before and after interviews also allows researchers to identify and reflect on any would-be bias that possibly influenced a participant’s response and biased the data (Fusch & Ness, 2015). A journaling approach was used to document my thoughts before and after each uninterrupted interview to avoid bias in my data.

Additionally, qualitative researchers are encouraged to use interview protocol, member checking, and data saturation to mitigate bias by using a personal lens to view data throughout the data collection process (Fusch, Fusch, & Ness, 2017). Thus, I implemented an interview protocol, which also provided the participants the flexibility to elaborate further on each question that was important to them, thereby pushing the conversation even deeper. Furthermore, the interview protocol helps in controlling precision in data saturation, protecting participants in a study by ensuring their details are kept secret, providing guidelines in ordering questions, and offering a strategy for meeting procedures (Castillo-Montoya, 2016). I used an interview protocol (see Appendix B) that required interview participants to sign a statement of informed consent, explained confidentiality, built affinity with participants, and explained expectations after the interview. The interview protocol ensured consistency with participants as well as focus on the task.

Participants

The participants consisted of data managers located in Germany who have successfully transitioned to big data technologies from relational data warehousing technologies. One of the most fundamental aspects of qualitative research is defining participant criteria and selecting the most appropriate participants to answer the research question (Lewis, 2015). Inclusion criteria, together with exclusion criteria, provide the eligibility criteria used to select or reject the target population for a research inquiry

(Koslowski et al., 2016). Proper selection of inclusion criteria will augment or improve the internal and external validity of the study (Kennedy-Martin, Curtis, Faries, Robinson, & Johnston, 2015). I defined the eligibility criteria for my study to align with my research question and the phenomenon I was investigating. Research participants were selected on the basis that they met my criteria for participation. Initial contact was made with the human resource departments of IT of the selected organizations to review the participants, eligibility, logistics, and the criteria for participating in the study. Participants were at least 18 years of age and had a minimum of 2 years working specifically on big data or DW technologies to be eligible to participate in the study. DWAdditionally, all participants had a practical knowledge and experience with technologies and tools used to transition from big data technologies to DW technologies.

It is important for researchers to develop a strategy to review and for gaining access to the participants (Hoyland, Hollund, & Olsen, 2015). Some of these strategies include sending the participants that match the selection criteria a brief introduction, potential study outcomes and benefits, data confidentiality, and the suitability and expediency of the interview process. In this study, I used a nonlinear process divided into four parts: (a) study design and planning, (b) identifying study contributors, (c) contacting contributors, and (d) interacting with the study contributors during the data collection process (Peticca-Harris, deGama, & Elias, 2016). Obtaining approval from key stakeholders during Stage 1 of this strategy can determine the pace and success of the data collection process. My strategy included planning and identifying participants as well as making the necessary contacts (phone and e-mail) with the human resource departments of the organizations whose data managers met my eligibility criteria to ensure I was in line with the organization internal policies regarding their right to participate in the research. I also ensured that I had the consent and approval of my committee members as well as permission from the Walden University IRB before the start of the data collection process.

Along with my strategy for accessing participants, gatekeepers are individuals who can help in accessing research participants and promote the research by validating or endorsing the phenomenon under investigation (Harris, MacSween, & Atkinson, 2017). Gatekeepers influence from within an organization and can use their knowledge and understanding of the study to provide suggestions and best practices to gain access to the study participants (Harriss et al., 2017; Hoyland et al., 2015). For my study, I established trust with at least one gatekeeper—a senior member of the human resource department from the organizations—to ensure that participants met the eligibility criteria. I contacted the participants identified by the gatekeeper primarily through e-mail containing information about my study before any involvement and contribution from their part in my study.

Researchers can also take a series of measures to guarantee a smooth process as they engage and exchange information with the participants (Peticca-Harris et al., 2016). Some of the steps include obtaining informed consent, making arrangements for interview times and locations, setting interview boundaries, interviewing the participants, and maintaining some form of flexibility to accommodate participants and avoid surprises. Ethical responsibility, trust, disclosure, and time consciousness are important aspects of the research that can help the researcher to gain access to the participants (Hoover & Morrow, 2015). The ethical standards and research quality of a study can be sustained by using the informed consent form acknowledged by the participants (Smith, & Fogarty, 2015). I sent a consent form to participants via e-mail, which included the study background, interview procedures, study benefits, privacy measures and risks. Participants acknowledged their willingness to participate in the study by replying to my e-mail using the words “I Consent” in their reply. Following their approval of the inform consent, I followed-up by setting a 1-hour schedule interview with each participant at their convenience with respect to the time and place.

Additionally, autonomy is important in maintaining a professional relationship between researcher and participants with regard to information disclosure, voluntary participation, and informed consent (Grady, & Fauci, 2016). Autonomy entails treating study participants as autonomous and protecting any participants with diminished autonomy (Grady & Fauci, 2016). A well-defined interview process and protocols and unambiguous questions help avoid any confusion, exploitation, and possible harm to participants (Hewitt, 2007). Further, by protecting the privacy of the participants, the researcher and research participants maintain a professional relationship, which promotes confidentiality and facilitates information sharing (Berger, 2015; Bromley, Mikesell, Jones, & Khodyakov, 2015).

Selecting the right location with a professional set-up for the interview is also important for avoiding bias and fostering relationships between the researcher and participants. I selected the most appropriate location for the interviews. I started by welcoming the participants and explained the interview process and the objectives of my study. I followed with a review of confidentiality and the steps I would be taking to preserve participants’ privacy. I also verified that the participants were participating voluntarily and had all read and consented as written in the informed consent e-mail that was sent to them. I encouraged participants to ask any questions pertaining to the study and my background before starting the interview to preclude any exploitation, bias, and confusion.

Research Method and Design

In this section, I expand on the research methods I adopted, focusing on the appropriateness of the qualitative practice in the study methods. Additionally, I explain why both quantitative and mixed methods were not appropriate for this study.

Method

Qualitative, quantitative, and mixed method methodologies were considered for this study. Qualitative research is a naturalistic inquiry process used for in-depth analysis and understanding of a phenomenon in its natural setting (Noble & Smith, 2015). A qualitative research method is more focused on the *why* rather than the *what* of the phenomenon under investigation and is dependent on the lived experiences of study participants to understand the phenomenon (Noble & Smith, 2015; Sarma, 2015). The researcher in a qualitative study explores the phenomenon under investigation as observed and experienced by study participants without any predetermined or preconceived beliefs (Lewis, 2015; Sarma, 2015). The perceptions and opinions of the participants are considered as one of the most important attributes of qualitative research

(Lewis, 2015; Noble & Smith, 2015). Rather than using logical and statistical methods, qualitative researchers learn from study participants to understand and analyze their lived experiences (Fusch & Ness, 2015). The researcher maintains a position of neutrality as he engages with research participants to ensure rigor and trustworthiness (Morse, 2015b; Smith & McGannon, 2018). I chose a qualitative research method for this study, as it was best suited to capture participants' perceptions, opinions, and interpretations on transitioning to big data technologies from data warehousing technologies in a real-world setting. DW

A quantitative research method was also considered for my study. However, the researcher in quantitative research uses an inferential approach based on a theory related to the question under study, then develops one or more hypothesis grounded on this theory and assesses the hypothesis with data using statistical procedures (Barczak, 2015). Additionally, quantitative researchers use measurements as the main data collection mechanism and use the reliability and validity of measures to deal with validation issues. But the current study did not involve any theory or hypothesis testing, variables, or numerical data collection for validating measures or statistical testing.

A mixed method research method was also considered for my study because it combines qualitative and quantitative research in the same study through the collection, analysis, and interpretation of data including both the numerical and narrative types (McCusker & Gunaydin, 2015). Researchers in mixed method research do not only design, build, and test theories, but they also perform inductive and deductive analysis within the same study (Hussein, 2015; McCusker & Gunaydin, 2015). However, my research was focused on the real-life experiences of the participants and not in generating theories, studying relationships between variables, or testing hypothesis. Given the lack of quantitative component in my study, a mixed method by extension was not appropriate for my study.

Research Design

The key design types for a qualitative study include phenomenology, ethnography, case study, and narrative analysis (Percy, Kostere, & Kostere, 2015). Case study research is suitable for bringing an understanding to a complex phenomenon and can broaden experience to what is already known from previous research (Fusch, & Ness, 2015; Lewis, 2015). Case studies emphasize detailed contextual analysis by developing detailed interpretations of a specific case or multiple cases of a limited number of conditions and events and their relationships (Lewis, 2015; Palinkas et al., 2015). A case study design enriches a researcher's understanding of a phenomenon through a broad examination to investigate and explore the phenomenon in a real-world setting by asking mainly how-type questions and to a lesser extent what-type of questions (Yazan, 2015). I selected a case study design to perform a comprehensive inquiry into the complex phenomenon of transitioning from data warehousing to big data technologies. I chose a multiple case study design to investigate the choices made by data managers to utilize a particular big data transition strategy, the implementation of these strategies, and the influence these strategies on big data technologies.

A phenomenological research design examines the human experiences from the viewpoint of the participants living through the phenomenon (Zhang, Yan, Barriball, While, & Liu, 2015). The purpose of phenomenological research is to depict the lived experiences of the participants (Zhang et al., 2015). The researcher in a phenomenological study sets aside all preconceived assumptions and biases about the human experiences, including the feelings and responses to the phenomena under investigation. This neutral nature allows the researcher to dig deep into the perspectives, understandings and feeling as well as the political, cultural and social perceptions of the people who have lived or experienced the phenomenon under study (Mayoh, & Onwuegbuzie, 2015). Participants interpretation or version of their life experiences would be significant in answering my research question, but the political, cultural, and social standpoints of their reality are certainly not. While some properties of a phenomenological design would make it appropriate for this study, I did not consider it as the most fitting design for my study.

Ethnographical research centers around the connections between a group of people and cultures and requires the researcher to fully immerse himself in the lives, and cultures of the situation under investigation (Blake, & Gallimore, 2018). Furthermore, an ethnographical research design aims at describing the cultural behaviors and social conduct through participant observation where the researcher is fully immersed and is fundamental to the understanding of the phenomenon (Blake, & Gallimore, 2018). An ethnographical researcher investigates participants practices and behaviors as opposed to their sensitivities or views (Marion, Eddleston, Friar, & Deeds, 2015). An ethnographical design was not appropriate for my study because it focuses mainly on cultures, practices, and behaviors within a group of people. Additionally, ethnographic designs rely immensely on observations that I believed would not be able to provide enough information for a comprehensive analysis of the data to answer my research question.

A narrative design was considered for m study but not used. A narrative design typically occurs in the context of an informant sharing a story, followed by researcher's analysis of that story (Lewis, 2015; Wang, & Geale, 2015). The researcher in a narrative study explores the life experiences of participants who describe or narrate their encounter with the phenomenon to the researcher (Wang, & Geale, 2015). Data collection in a narrative design is centered around how participants view themselves and their life experience with the phenomenon or event (Kostov, Rees, Gormley, & Monrouxe, 2018). Participants recount their version of the story as they remember it. My research focused on the strategies data managers use to transition to big data from data warehousing, and studying the strategies used by a single data manager will not yield the relevant data to answer my research question.

Data saturation in a study involves the continual adding of new participants until the research reaches a point in the analysis that sampling or testing more data will not lead to more or new information connected to the research (Fusch, & Ness, 2015; Gentles, Charles, Ploeg, & McKibbon, 2015). Data was collected to the point where no new information was generated from the data, indicating the research had attained data saturation. Data Saturation was achieved by collecting data from all participants and adding new participants if necessary, until the point where adding new participants brought in no new data. Getting contributions through interviews with participants with considerable experience and knowledge of the phenomenon under investigation aids in achieving data saturation (Malterud, Siersma, & Guassora, 2015). I interviewed participants who had considerable knowledge and experience in transitioning to big data technologies from data warehousing technologies to support the study reaching data saturation. By using face-to-face interviews, a study can easily reach data saturation by asking the same probing questions to multiple or all participants to ensure the depth and richness of data collected (Fusch, & Ness, 2015). I asked the same probing question to all study participants to ensure depth and richness in data I collected to help reach data saturation.

Population and Sampling

The population of my study consisted of data managers from 5 mid-sized companies located in Germany. Some organizations depending on their size could have more than one data manager. The choice of 5 organizations is to provide a reasonable sample given that some organizations might have more than one data manager or data lead. Sampling was used as the sampling method, meaning that all chosen participants were used in the data collection process. The chosen population had the experience and knowledge surrounding the transition process from data warehousing technologies to big data technologies. The population for the one-to-one in-depth interviews consisted of all selected data managers or data leads with over 2 years of experience working primarily with the data management department of the organization and were involved or actively contributed to the transition process to big data technologies. A recommended initial step in the data collection process is to define the study population using inclusion and exclusion criteria (Kennedy-Martin et al., 2015). I used a gatekeeper to assist in coordinating efforts between myself and the participants.

Census sampling is a method used in a study to select samples and all eligible employees (Asadollahi et al., 2015). Census sampling design provides accountability in qualitative research (Tobin, Nugroho, & Lietz, 2016), a fact supported by Kennedy-Martin et al. (2015) who added that the inclusion criteria used for participant selection should be able to capture all participants of interest. Therefore, the inclusion criteria for my study included data managers or data leads with at least 2 years of experience working with big data and DW technologies and actively involved in the transition process to big data technologies from data warehousing technologies. Census sampling was incorporated into the chosen participant selection method. When used, census sampling establishes a sampling skeleton for conduction studies of numerous establishments (Charman, Petersen, Piper, Liedeman, & Legg, 2017). I used census sampling from within the populace of the participants who met the eligibility criteria to participate in the study.

Census-based surveys method of sampling was inappropriate for this study due to the following reasons: (1) including a greater number of participants than it is effectively required could lead to ethical issues in research, (2) census survey could be impacted by budgetary limitations due to the high number of people required limiting its use as a strategy to select participants, (3) logistics and available resources could limit the effectiveness of a census survey as staff, and equipment might be required to conduct the study,(4) census surveys could also have time restriction issues as the amount of time needed for planning and conducting a census-based survey can be excessive, and (5) the population for a census-based survey is often unknown making it a difficult to conduct a census-based study (Martínez-Mesa, González-Chica, Duquia, Bonamigo, & Bastos, 2016).

The chosen location and setting of the interview are vital to the success of the interview and may have a direct impact on the content of the interview and influence data collection (Parker, & Kingori, 2016). Arrangements must be made to guarantee that the interview occurs in a place that is free from interruptions to avoid any potential distractions (Parker, & Kingori, 2016).). Additionally, the place and time of the interview must be convenient to the participants and allow for adequate time for the interview (Dikko, 2016; Parker, & Kingori, 2016). I had 30 – 45 minutes interview with selected participants either through phone, Skype or in-person in a closed room void of any distractions and noise. Moreover, I followed the interview protocol (see Appendix B), which includes the same interview questions for all participants. Lastly, it was required that all participants sign the consent form before the interview, which outlines the measures to protect and guarantee the non-disclosure of their personally identifiable information

Member checking and triangulation was used to verify the accuracy and reliability of the study in assuring credibility of the results. Member checking which involves a back and forth of return and review of results of the study with research participants is vital in assuring credibility in research (Gunawan, 2015; Hadi, & Closs, 2016). Additionally, the association of member checking and triangulation are essential in ensuring the authenticity of the data collected (James, 2017). I will be performing member checking and data analysis within a few days of the interview, to ensure that data saturation is reached in the research.

Ethical Research

I maintained ethical responsibility in the research by protecting every participant's privacy and making sure every participant had signed the consent form before the start of the interview. The consent form is a mandatory ethical requirement that describes the researcher's obligation to counsel the study participants of the benefits and risks associated with the research, as well as their rights as a participant (Koonrungsesomboon, Laothavorn, & Karbwang, 2015). More so, research participants should be able to exercise their right of free will in deciding to participate in a research and participants selection should be objectively obtained by the researcher (Cho et al., 2015; Lewis, 2015) Lastly, Koonrungsesomboon et al., (2015) also mentioned that the use of an informed consent form vital in enforcing and protecting the rights the participant. Consequently, language-specific terms were included in the consent form to stress the fact that participant protection and privacy as well as confidentiality shall be guaranteed. The consent form was then emailed to all participants for approval before the commencement of any data collection. Once consent form signed, the participants all agreed on a voluntary basis to participate in the study with the option of opting out of the study at any given time.

After receiving the signed copy of the informed consent, participants were given a copy, and another copy was saved along with other documents associated with the research for five years. Additionally, participants were given a choice to either deliver the informed copy in-person, for those who opted for in-person interviews or by email for those who opted for phone or skype interviews. The right to withdraw from research as a participant is essential ethical protection, and researchers should always inform participants of their rights to do so at any given time (Harvey, 2015). I ensured that participation is voluntary by informing participants of their right to withdraw from the study in the informed consent document as well as at the start of the interview. I also reminded the participants that withdrawal could be made either through verbal or a written notice with no repercussions.

All information recorded as well hard-copy data from this multiple case study was stored on an encrypted hard drive, which is secured in a safe place for a five years period to safeguard the rights of the participants. The prerequisite for security in a study depends on the context and the setting (Deuter & Jaworski, 2016). As a result, the name of the research participants as well the participating organizations will be kept secret by using code words to replace their real names. These code names which include words like participantA, companyA, participantB, companyB and so on used to mask the real names of the participants and organizations are known only by the researcher. Moreover, I obtained approval from the Walden University's Centre for Research Quality (approval number will go here) before contacting and recruiting the study participants. The need to seek and obtain approval from the IRB before recruiting study participants was stressed during the National Institutes of Health training course for using human subjects in research that I completed (see Appendix A). Finally, Brière, Proulx, Flores, and Laporte

(2015) noted that the use of incentives such as gifts and money could lead to unethical conduct such as intimidation and duress. Consequently, all study participants were made notice of the fact that no incentives – monetary or gifts would be made available for participating in the study.

Data Collection

In this section, I discussed the data collection methodology and application throughout the study. Semi-structured interviews with open-ended questions were the primary data collection method (Appendix C). that were used to collect data from the selected participants based on the on the interview protocol introduced in an earlier section. More so, the interview protocol helped in capturing valuable information like the participant background, and position occupied in the organization. I relied on the use of semi-structured interviews techniques to create a general picture of how data managers transition to big data technologies from DW technology

Instruments

The principal data collection methods in qualitative research include participant observation, interviews, and document analysis (Percy et al., 2015). The researcher in qualitative research should be an informant centered instrument whose main role as the primary data collection is to view and engage with participants as experts, allowing them to share any information they deem most relevant to the research (Santiago-Delefosse, Gavin, Bruchez, Roux, & Stephen, 2016). By maintaining a journal of self-examination and self-reflection, the researcher will be able to reflect the degree to which his/her biases and subjectivity may have influenced the different sections or components of the study, and this can effectively help avoid bias (Berger, 2015). I recorded and updated a reflective journal throughout the research process consciously noting any beliefs, personal values subjectivity or biases that may have impacted my study. I reflected and followed recommended procedures to examine my involvement throughout the data collection process, searching for any undue influence, inadequacies, subjective interpretations, and any other forms of bias.

Interviews. Interviews are one of the most significant methods for data collection in qualitative research (Percy et al., 2015). The purpose of using semi-structured interviews in my research is to discern subjective answers from the participants with regards to the phenomenon under investigation following a well-planned and detailed interview protocol (McIntosh & Morse, 2015). Open-ended questions asked in an organized manner will help the researcher generate detailed narratives and probe for deeper comprehension or understanding to improve study validity (Bengtsson, 2016; Rosenthal, 2016). Semi-structured interview questions are organized by the researcher around the main topic under investigation as highlighted in the research question, but the researcher should remain flexible to allow participants to inject new themes or change direction as long as it remains within the bounds of the research (O'Keeffe, Buytaert, Mijic, Brozović, & Sinha, 2016).

I used an interview protocol (see Appendix B) that consisted of interview questions, interview reminders and pre and post-interview activities. A set of predetermined, open-ended questions were used in the interviews with the participants. I asked follow-up questions as required during the interview process to get any necessary clarifications on participant's responses. I remained flexible during the interviews to allow participants to inject new topics or change direction.

Document analysis. Document analysis in qualitative research involves the selection of recorded or written information that corroborates or provides an explanation and description of an event (Sommerhoff et al., 2018). Additional data sources can enable the researcher to check for consistency and triangulation of data to improve the reliability of research (Yazan, 2015). It is essential that researchers ensure that the quality and trustworthiness of the artifacts and documents they use in research are genuine (Islam, 2014). The factors that determine the quality and trustworthiness of the documents collected include but not limited to authenticity, credibility, accuracy, and reliability (Donaldson, 2016).

I used document analysis in my research to collect data on big data and DW technologies. Selecting and analyzing documents is a process that must be undertaking by the researcher to assess the quality and trustworthiness of the documents (Dunne, Pettigrew, & Robinson, 2016). Qualitative researchers are recommended to select a wide range of sources for documents to ensure they have a comprehensive understanding of the data or area of focus (Dunne et al., 2016). I collected and analyzed procedural documentation related to big data technology standards, architecture design best practices, methodologies, and processes.

The goal of collecting and analyzing documents was to measure the perceptions of the study participants to their respective organization's strategies regarding big data and DW technologies. I was the primary data collection instrument and the documents collected will provide an additional source of data to substantiate and check the constituency of other data sources.

Triangulation. Hussein (2015) asserted that to develop a thorough understanding of the phenomenon under investigation, methodological triangulation is recommended, which necessitates the use of multiple data collection and analysis methods thereby contributing to the accuracy, validity, and reliability of the study. Methodological triangulation provides rich and accurate data as well as correlation in the data, thereby fostering the understanding of the research phenomena as well as the validity of the research (Fusch & Ness, 2015).

I used the within-method type of methodological triangulation to analyze data collected during the semi-structured interviews, interview observations as well as document analysis. Data from the semi-structured interviews was crosschecked with those from the interview observations and document analysis. The aim was to find consistencies within the data that corroborated my interpretations.

Data Collection Technique

Semi-structured interviews and document analysis were the primary data collection methods in my study. The data collection technique included amongst others gaining access to study participants, obtaining informed consent from participants, selecting appropriate locations for the interviews, scheduling the interviews, creating and following an interview protocol, and interview follow-up or member checking. I ensured that I have the approval and authorization of the IRB before recruiting and engaging with participants as well engaging in any other data collection activities. Gatekeepers through the trust, affinity, and camaraderie that they have with the participants can facilitate contact between the researcher and participants by highlighting the importance and benefits of the study (Peticca-Harris et al., 2016). I solicited the help of one to two gatekeepers within the organizations that met my participant criteria to help identify and ease access to potential study participants. The participant selection criteria were emailed to the gatekeepers to enable easily identify potential participants. Following correspondence with the gatekeepers, all identified potential participants were sent an email with detailed information about my study.

The use of an informed consent form in qualitative research is an important ethical requirement that outlines the researcher's responsibility to inform study participants of the changing aspects of the study (Sanjari, Bahramnezhad, Fomani, Shoghi, & Ali Cheraghi, 2014). Participants are advised to sign an informed consent to confirm their willingness to participate in the study (Cridland, Jones, Caputi, & Magee, 2015). I emailed the informed consent form to the selected participants with information on the nature of the study, background, voluntary participatory note, risks, benefits and privacy considerations in the study. It was required of all participants to acknowledge their willingness to contribute to the research by replying to my email which contained the informed consent form with the words "I consent".

To avoid disruptions researchers are advised to inform participants of the time, place, scheduling and duration of interviews so participants could adjust their schedules as required to avoid any potential disruptions (Peticca-Harris et al., 2016). The researcher should make sure that the place and time of the interview must be convenient to research participants and allow adequate time for the completion of the interview (Dikko, 2016). If multiple interviews are conducted on the same day, qualitative researchers are advised to allow between 30 to 60 minutes between interviews for debriefing (Rimando et al., 2015). I scheduled a 30 to 45 minutes interview with each participant at a convenient location to them. I emailed the participants to agree on time, place, and duration of the interview. I remained flexible and gave the availability of the participants, I conducted at most two interviews per day with a minimum of 30 to 60 minutes between the interviews for debriefing.

The interview should be held in an area that is free from interruptions that might distract the participants and affect data collection (Seitz, 2016). The interview location and setting should minimize background noises that disturb or distract participants and interfere with the audio recordings potentially impacting the data collection process (Dikko, 2016; Seitz, 2016). I reserved a location that will ensure the privacy of the participants and is close enough to the participants to minimize travel time. The doors of the interview room shall remain closed during the interview with a note at the door requesting no disturbances to avoid any potential disturbances and distractions from individuals not participating in the interview. I ensured that any pictures, signs, and decorations in the room that might be an object of distraction were covered and that the room was free from any repulsive odors or smells. I ensured that there were no background noises within the interview room or anywhere close to the interview room that might interfere with the interview process or be an object of distraction to the participants.

Taking down notes during the interview process can be distracting and interfere with the process as it can limit the research's ability to listen actively to participants and their responses and hence impact his ability to ask follow-up questions (Oltmann, 2016). An audio recording of an interview gives the researcher the opportunity to re-listen to an interview in part or its entirety allowing a better interpretation of the interview as well as familiarity with the interview (Zolnikov, & Blodgett Salafia, 2016). It is advisable to use good quality recording equipment to avoid any nuances and be also well familiar with the operation of the equipment (Krogh, Bearman, & Nestel, 2015). Researchers can also use multiple devices to record audios during the interviews to provide redundancy in the advent that one device fails (Oltmann, 2016). Following the interviews and follow-ups, the researcher can transcribe the audio to familiarize with an entire interview (Zolnikov, & Blodgett Salafia, 2016). Despite its recommended usage, audio recordings of interviews have the disadvantage in their inability to capture impressions, behaviors, environmental contexts, and other-nonverbal cues (Sutton & Austin, 2015). I used at least two recording devices to record the interviews for accuracy with the consent and authorization of the participants. I took notes during the interview of my observation of the participants' expressions, and body gestures that may warrant further investigation using probing or follow-up questions. I listened to the audio recording multiple times after the interview to increase my familiarity with the interviews. A transcribing software was used to transcribe the audio recordings. Any personally identifiable information was removed from the transcription and participants were identified using codes.

Maintaining a reflective journal before and after the interview can help the researcher examine, evaluate, identify and reflect on any potential bias that may have influenced or swayed participant's response and skewed the data (Berger, 2015). I used a journaling approach to document my observations, experiences, methods, thoughts, and judgments throughout the investigation. My reflective journal will be maintained using Microsoft Word. Microsoft Word is equipped with review-tracking and commenting features that are all critical to maintaining and updating my journal. By using Microsoft Word, I was able to keep my initial thoughts and update them as they evolve. This process allowed me to periodically reflect on my journaling notes evaluating my initial perceptions and expanded my understanding of the research.

I used an interview protocol (Appendix B) to facilitate the interview process. I started the interview by introducing myself to the participants and thanking them for accepting to participate. I followed by reminding participants that the interview was recorded and reminded them that the interview will remain confidential while responding to any concerns they might have. I turned on the recording device(s) and followed-up with the announcement of the date, time, and identifying code assigned to the participant. Every question was asked according to my interview protocol starting with the first question on the list and following up on the last question. Participants were allowed to answer each question, and I asked follow-up, probing questions as appropriate. After going through all the questions, participants were asked if they had additional information to share that was not discussed as we went through the questions. Participants were also be asked if there is any documentation that they believe might be relevant to the topic. The concept of member checking was explained to the participants, and a schedule was agreed with the participants for a follow-up interview to review my interpretations of the interview as appropriate. The recording device were switched off and recording discontinued, and I thanked the participants for participating in the study. I made sure before parting with the participants that they had my contact information for any concerns or follow-up queries.

Follow-up meetings were organized with the participants for member checking after the conclusion of the interviews. Before the follow-up meetings, I transcribed and analyzed the data collected from the participants to interpret the data for any emerging themes. A copy of the transcription as well as my interpretation were emailed to the participants a few days before the follow-up interview. The participants were asked to corroborate my interpretation during the follow-up interview by comparing my interpretations, descriptions, and understandings of the interview data to their lived experiences and viewpoints. I also used this opportunity to ask follow-up questions to the participants as necessary to seek any necessary clarification of my understanding of their views and ask them if there wass any new data they wanted to share with me. If any additional information is received from the participants that warrants new follow-up meeting, I agreed on the schedule with the participant(s) and repeated the member-checking process until no additional information was received.

Data Organization Techniques

Data organization in qualitative studies is an indispensable process for analyzing and interpreting data collected enhancing the trustworthiness of the research (Kornbluh, 2015). Researchers rely on codes, concepts, and categories to ease the labeling, comparisons, and sorting of the data collected for analyses (Vaismoradi, Jones, Turunen, & Snelgrove, 2016). Strategies for categorizing data retained by researchers often involve identifying similarities in data during analyses and grouping the data by likeness (Plamondon, Bottorff, & Cole, 2015). I assigned a code to each participant to not only maintain confidentiality but also to track their data. Cloud storage with encrypted folders will be used as storage for participant information, interview data and recordings, member checking and follow-up interview data, my research journals and organization journal to develop and record artifacts.

A reflective journal was used to record my thoughts, methods, observations, thoughts, new understandings, and draw conclusions from the research. Reflexivity, journaling, and reflection all combine to improve the quality and validity of research data (Vicary, Young, & Hicks, 2016). A research journal is not limited to improving the research validity and quality but could also provide a means for other researchers to judge the transferability of the research (Lamb, 2013a). The reflexive journal enabled me to periodically reflect on my journal as the research progresses to review my initial perceptions and expand my understanding of the research. It is common for qualitative researchers to use a categorization matrix to identify, track and group data that correspond to categories and concepts related to the data collected to improve data and research validity (Neale, 2016). I used an Excel spreadsheet for my categorization matrix purposes to associate information, journal notes, forms and interview data with participant assigned codes. The spreadsheet was used to assign categories and concepts to collected data to enable sorting and grouping of the data to better my understanding of the data

All collected data in electronic format was encrypted and stored in encrypted folders in the cloud while non-electronic data was stored in a securely locked safe to safeguard the privacy and confidentiality of the participants. I will keep the secured data for a five-year period from the publication of my research. After the fifth-year elapses, I will permanently delete and destroy all electronic data and media as well as hard-copied documents.

Data Analysis Technique

For my study, the within-method type of methodological triangulation was used to analyze the data collected through the semi-structured interviews, documents about big data and data warehousing, methodologies and best practices. Methodological triangulation necessitates the use of multiple data collection and analysis methods about the phenomenon under investigation for a thorough understanding of the phenomenon improving the reliability and validity of the study (Hussein, 2015). Methodological triangulation is beneficial in many ways to a study including but not limited to data correlation, increasing validity and reliability and increase understanding of the phenomenon under research (Fusch & Ness, 2015). Palinkas et al. (2015) contended that a single method of data analysis is most certainly not adequate to describe a research phenomenon. The within-method of triangulation is recommended because it requires the use of complementary data collection and analysis methods to increase credibility and accuracy of a study (Hussein, 2015).

I analyzed the data collected from interviews and documents repeatedly until I found a substantive answer to my research question on the strategies used by data managers to transition from DW technologies to big data technologies. Data analysis process must be accurate, transparent, and have a clear methodology depicting an accurate or near accurate picture of the phenomena from the participants' perspectives and views (Noble & Smith, 2015). I tried to understand participants views expressed during our interactions and interpret their views to my understanding in a reliable manner. Qualitative data analysis is inductive and focuses on the meaning of the data to foster the emergence of concepts from the data (Noble & Smith, 2015). An inductive approach of data analysis requires the use of techniques like coding of data using categories, data interpretation, and verification of the representativeness and trustworthiness of the data (Neale, 2016).

I started the coding process by submerging myself in the data through repeated reading to familiarize myself with the deepness and breadth of the data. I took down notes and jotted important points throughout the coding process to track my thoughts and reflect on my analysis. Different coding methods were used to create an initial list of codes that best represent the data collected and are in line with a research question, which were followed by the execution of multiple coding cycles of coding to search for patterns, relationships, explanations and underlying meanings of the data. All findings including the codes were documented in my journal. I categorized and grouped the codes after generating them to develop themes that were organized later into various knowledge areas or domains. This process was repeated as required until a meaningful and conclusive elucidations of the phenomena in coherence with my research question was reached.

Computer-assisted qualitative data analysis software is a tool commonly used by qualitative researchers to perform different types of data analysis to foster the emergence of underlying relationships and hence provide better and quicker results than manual analysis (Moylan, Derr, & Lindhorst, 2015). Another great tool for data analysis is the Qiqqa and NVivo software - the Nvivo software can be used for word count, classical content-analysis, perform constant comparison analysis, put keywords in contexts, perform domain and taxonomic analysis, as well as componential analysis (Maher, Hadfield, Hutchings, & de Eyto, 2018). I used the latest version of the NVivo software to analyze all data collected from semi-structured interviews, audio-transcripts, documents, and notes by utilizing the different coding methods available in NVivo to code, categorize and group the data. (Maher et al., 2018). The memo and notes functionality is one of the features I used as a research journal for making assumptions, tracking my ideas and reflecting on the research.

To catalog data into themes, a systematic coding process is needed to foster a better understanding and subjective interpretation of data validty and reliably (Maher et al., 2018; Nicod, & Kanavos, 2016). Advanced functionalities in NVivo can also be used by researchers to visualize data by using graphs, models, maps, reports, and cluster analyses to monitor the emergence of themes (Maher et al., 2018). The advanced functionality in NVivo were utilized to generate mind maps, word trees, word clouds, cluster maps, cluster analyses, and graphs to have a complete visual representation of my data for a better interpretation of the relationships, context, frequencies, and the emergence of themes. I arranged, filtered, sorted, assembled and analyzed the data repeatedly until all major themes and trends that emerge remained consistent with my research question. I looked for recurring themes and familiar patterns that could be an indication of an existing correlation between big data technologies, DW technologies and the key tenets of TAM2.

My data analysis process included all relevant data from my literature review, semi-structured interviews and documents that are consistent with my research question, and conceptual framework. I also searched across scholarly and peer-reviewed libraries for any newly published studies and articles that could be relevant in finding answers to my research question. Any new articles found with relevancy to my study was considered in my data analysis.

Reliability and Validity

Reliability

Reliability and validity are two concepts used by researchers to reveal the quality of the research. To assess the reliability of research findings, it is required of the qualitative researcher to make a judgment about the 'soundness' of the research concerning the application and appropriateness of the methods used as well as the integrity of the concluding remarks (Noble, & Smith, 2015). The nature of reliability extends the consistency of the data collection process (Leung, 2015). In other words, each research participant will be asked the same questions multiple times as required. Reliability and validity are required conditions to make a qualitative inquiry trustworthy

(Morse, 2015b). To ensure reliability, I assured each participant that none of their information will the shared with any other participant or third-party. The trustworthiness and rigor of qualitative research is assessed by determining the credibility, dependability, transferability, and confirmability of the study (Gaus, 2017; Morse, 2015b). The concepts of dependability, transferability, and confirmability are addressed in my study. A qualitative study is also deemed reliable if the methods, techniques, and phenomena used in the research produced the same results when used by other researchers (Bengtsson, 2016). The prerequisite for the validity of research data in a qualitative study is the reliability of the data (Bengtsson, 2016). I respected and maintained the privacy and confidentiality of the participants to make certain the integrity of my study.

Validity

Validity is defined as the "degree to which inferences made in a study are accurate and well-founded" (Polit & Beck, 2012, p. 745). Miller (2008b) took the definition a step further stating that the validity of a study is measured by the "goodness" and "soundness" of the study (p. 909). The validity of a qualitative inquiry requires that the research and study participants through the findings, to demonstrate a credible, trustworthy, and authentic understanding of the phenomena, and this is usually determined by how well the actual phenomena is represented by the research (Morse, 2015b; Yilmaz, 2013). Strategies for ensuring validity in a qualitative study include prolonged engagement, persistent observation, peer review or debriefing, rich description, member checking, clarifying and avoiding researcher bias, external audits, triangulation and negative case analysis (Morse, 2015b). I used these strategies to ensure validity in my study by engaging with research participants on multiple times as required, will requests participants to corroborate my data analysis, avoid bias and perform triangulation.

Dependability. Dependability is achieved in a study if the study is repeatable with a same or an equivalent number of participants in the same context (Amankwaa, 2016). The dependability of a qualitative inquiry can be increased by establishing arduous sampling, performing member checking, using recommended and verifiable data collection and data analysis methods, and employing other recommended procedures (Zadvinskis, Smith, & Yen, 2018). For my research inquiry, census sampling was used to ensure that I collected data from all willing participants identified as study participants. Member checking is a critical process to researchers in establishing the dependability of a study as it allows researchers to discuss the development of tentative themes as well as data interpretations with the participants (Birt, Scott, Cavers, Campbell, & Walter, 2016). I used member checking to improve the reliability of the study by ensuring with research participants that my interpretations of the data collected match their experience. I asked participants to comment on my interpretations, and themes and make appropriate comments depending on the degree to which my interpretations reflect their viewpoints.

Credibility. The credibility (also known as internal validity) of a qualitative study stems from participant response. Credibility demonstrates and establishes the truth behind the findings (Amankwaa, 2016). I explained with as much clarity as I can all the processes and stages of my research, providing details where needed in every aspect of my study. I used triangulation to collect and analyze data from multiple sources to increase the reliability of the results. I describe the details of the data collection and analysis process, themes development process and interpretation of results as part an audit trail. The audit will serve the purpose of simplifying and facilitating the replication of my research by other researchers hence contributing to the credibility and dependability of the study.

Transferability. I documented every aspect and detail of my study to facilitate transferability. A critical factor in determining the transferability of a research study is the scope to which the research findings are applicable in other contexts (Amankwaa, 2016; Guba, & Lincoln, 1985). A determining factor in the transferability of a research study is the scope to which the findings are reproducible outside of the immediate research study or are generalizable to other contexts or settings (Guba, & Lincoln, 1985). The transferability (or external validity) of a study requires a careful and detailed description of study background, participants, population sampling and results or findings of the research so that readers of the study can determine the transferability of the research based on the findings (Bengtsson, 2016; Noble, & Smith, 2015). A researcher can improve the transferability of a study by providing a detailed and thick description of the study context as well as the procedures used, to allow readers make decisions as to the transferability of the research (Guba, & Lincoln, 1985). I provided a detail description of the study background, eligibility criteria for selecting participants, sampling methodology, the population size, and data analysis method so other researchers may ascertain the transferability of my study.

Confirmability. The ability by the researcher to keep records in an orderly manner of every single methodological choice they took throughout the research (like a record of data sources, sampling decisions as well as the informative systems and their execution) is referred to as confirmability or construct validity (Amankwaa, 2016; Tong & Dew, 2016). The confirmability of a study lies in the ability to corroborate the effective occurrence of research by other researchers using the principles and methods as outlined by the researcher (Guba, & Lincoln, 1985). A research experiences confirmability when there is clarity in data to substantiate its findings and the interpretations are logical to ensure trustworthiness since the result of a qualitative study is not shaped by the bias of the researcher, but rather through the responses provided by the participants (Amankwaa, 2016). I kept records in my reflexive journal of all procedures, data collection, analysis and interpretation methods to enable other researchers with interest, the possibility to review my reasoning. To establish confirmability in my study, I performed member checking as much as needed until all participants validated my interpretation of their data.

Data Saturation. Data saturation increases the quality of the research content validity, and without data saturation, the quality of the research content validity could be questioned. Data saturation occurs in research when the researcher has collected enough information until it becomes repetitive and no additional information available would improve the research (Fusch & Ness, 2015; Gentles et al., 2015; Morse, 2015a). Hence the importance of interviewing multiple participants. I collected data from all participants using census sampling and conducted member checking until information became repetitive indicating that I had reached saturation in data.

Transition and Summary

In this section, I extended the discussion of my prospective study, reemphasizing the purpose of the study, research design, discussing sampling, data collection, and analysis techniques, and addressing the role of the researcher and participants. I discussed my role as the researcher as the primary data collection instrument and how I followed ethical guidelines throughout my study while avoiding bias. Additional concepts like data reliability, validity, data saturation, and triangulation were also discussed. In Section 3, I presented my research findings, highlighted the implication for social change, the implication for the practice of IT, and made recommendations for future research.

Section 3: Application to Professional Practice and Implications for Change

This section includes information and presentation of the findings from this qualitative multiple case study and how these findings applied to professional practice. I follow with discussion on how the findings might provide positive social change. I conclude with recommendations for actions as well as provide recommendations for further study including my personal reflections with regard to the study.

Overview of Study

The purpose of this qualitative multiple case study was to explore the strategies data managers use to transition to big data technologies from DW technologies. The data were obtained from semistructured, phone/Skype interviews conducted with participants based in Munich, Germany. Data were collected from 10 data managers, leads, or senior data scientists from six separate organizations that had transitioned from DW technology and were actively using big data solutions. I also collected and analyzed data from publicly available sources on big data and DW technologies (n = 15) that were either provided directly by study participants or referenced (when organizational policies prevented them from sharing them with me directly). All participants had between 5 and 15 years of overall IT experience and between 2 and 7 years of experience working with big data and DW technologies. I used member checking and data triangulation to increase the validity of findings from the data analysis. The findings from the data analysis identified four strategies for transitioning to big data technologies from DW technologies.

Presentation of the Findings

The research question I sought to address was “What strategies do data managers use to transition from data warehousing technology to big data technologies?” All 10 participants indicated that there are many factors to consider when transitioning from DW technologies to big data. From these many factors there were four main themes that are essential for a successful transition from DW technology to big data technologies: (a) identifying a business need or use case, (b) using multiple data sources, (c) guaranteeing executive-level support, and (d) creating or using a data lake in the cloud. All participants also mentioned that the size of the organization mattered in the transition process but did not see it as an important factor because big data technologies can be implemented successfully regardless of the size of the organization.

A total of 10 participants from six organizations (cases) contributed to the research. Fifteen organizational documents were also provided and reviewed to increase study validity. To ease readability, I utilized the following naming convention: PxOy—P for participant number and O for organization number. For example, P1O1 represents Participant 1 from Organization 1 and P1&2O4 represent Participants 1 and 2 from Organization 4. The following sections contain four main themes mentioned by participants, which I link to the conceptual framework.

Theme 1: Identify Business Needs—Use case

The first theme that emerged from the data analysis was that organizations need to identify their business needs (use case) before embarking on the big data transition process. Theme 1 stresses the need for organizations to start by identifying the business needs or a business problem that transitioning to big data technologies from DW technologies will solve rather than investing resources in transitioning to big data for technology's sake. Big data comprises a multitude of technologies and understanding the use case or business need will determine which big data technology is the best fit. To effectively identify business needs, key stakeholders including technology implementers and users as well as senior management need to be involved in the process at an early stage, which sets the foundation to address the business need by transitioning to big data technologies.

All participants asserted that understanding the business needs before starting a big data transition process is essential for a successful transition to big data technologies (see Table 1). P1O1 stated that "understanding the specific business goal to achieve is imperative for a successful transition to big data technologies." P101 further stated that "There are plenty of big data technologies and the most appropriate big data technology will depend on the specific business vertical." P1O6 also insisted that for businesses to overcome the common pitfall of starting with the transition process without a business need, breaking down existing silos in the organization by getting leaders, business users, and technologists involved early in the transition process is important to define the business need for transitioning to big data technologies. P1&2O5 also highlighted that many considerations must be made in terms of design and technology before starting a big data transition project. With so many technologies, tools, and big data solutions available, the success of a big data transition project will depend on the big data technology stack chosen. This is another reason why identifying and defining a business use case that will determine which big data technology stack to use is important.

Table 1

Theme 1 Frequency of Occurrence

Source of data collection	t	*n*
Interviewees	10	18
Documents	15	32

Note. Theme 1 identify business need or use case; t = total number; *n* = frequency of occurrence.

A similar emphasis was found on this theme in the organizational documents I collected and reviewed for my research (see Table 1). Reviewing the documents corroborated the depth and importance of identifying a business need before starting a big data transition project. Methodological triangulation was achieved, as all 15 documents analyzed supported and referenced this theme. According to one of the documents, all big data projects should reference a specific use case. This document also highlighted the fact that selecting the appropriate big data technology is the foundation for every successful big data implementation. This document is consistent with the reports from six participants who asserted that defining the business needs or use case was the most important aspect of the transition process. P1O5 mentioned that agreeing to a well-defined business use case was the catalyst that led to a successful transition to big data technologies in their organization.

Other documents reviewed also supported this theme. According to the "*Top Five High Impact Use Cases for Big Data*" document, "The real business value of big data is always unlocked through specific use cases and applications. The document in another section stated that “organizations should first take the time to identify and crystallize the right use case or use cases for their business needs before starting any big data project. ” Describing this as a critical first step, the document highlighted that tThis will enable the organization to identify key business insights for their big data projects. This claim is also supported by other organizational documents reviewed like one that the “*Top 22 Use Cases for Big Data*” document which highlighted 22 different use cases where different big data technologies can be applied. The use cases highlighted in the document match some of the use cases described by participants. For example, P1O4 described how determining the business use case for financial fraud detection was vital to choosing the right big data technology for implementation. Given the vast amount of data to be processed, the easiest and most cost-effective way to achieve success was to transition to big data technology.

Scholarly literature also supported having a well-defined business use case or need before starting a big data transition project, which aligns with data collected from my interviews and organizational documents. For example, Ohno-Machado et al. (2017) stipulated that use cases help define appropriate boundaries and the level of granularity that will determine the success of any big data implementation project. Additionally, Klievink, Romijn, Cunningham, and de Bruijn (2017) remarked that although organizations can technically use or implement big data technologies, they will not significantly gain from these implementations if the use cases do not fit their organization’s business model and core business objectives. Klievink et al. (2017)

proposed a framework that will not only provide guidance on how to become big data ready but also help in identifying the right use case for the big data initiative or transition process. Furthermore, Urbinati, Bogers, Chiesa, and Frattini (2019) outlined the benefits of using a use case-driven strategy that organizations can adopt to determine an executable big data use case that will bring high business value. A case-driven strategy will enable organizations to focus on the most fundamental aspects of a big data transition process, which includes technology selection, resource management, and training. Therefore, researchers have supported the need for defining a use case before starting a big data project or transitioning to big data technologies from DW technologies.

Further, the participants in this study indicated that having a well-defined business need was important to all stakeholders and improves relationships between employees, which aligns with the conceptual framework of this study relating to social influences having a positive effect on perceived usefulness and perceived ease of use and through these, the intention to adopt a technology. P1O6 explained that collaboration among employees to identify a use case is a requirement and is important for success. Likewise, P1&2O5 explained that a big data transition project would not start without a clearly defined business use case, making it a mandatory step to follow. Employees working together to agree on a business use case in big data a project has a positive effect on the importance of subjective norm as a determinant (Weisenberg et al., 2017). Employees who perceive useful the implementation of big data technologies in a mandatory setting are more likely to adopt the technologies (Bai et al., 2017; Verma, Bhattacharyya, & Kumar, 2018). P1O6 also mentioned that employee involvement in defining a big data use case has a social component that encourages everyone and gives them a sense of importance to the team and respect for and by other people. Based on the TAM2, social influence factors like an individual's social group and importance in team or organization or personal prestige significantly influences acceptance of a new technology (Venkatesh & Davis, 2000).

Participants also suggested that identifying a business use case before starting the transition to big data was a necessary and relevant addition for their job and performance. P1&2O5 and P1O1 highlighted that identifying a business use case is an essential first step to start working with varying types of data effectively. Moreover, P1O4 explained that developers are more willing and find it easier to develop features using big data technologies when they have a well-defined use case than using difficult to implement DW tools. The willingness of developers to develop new features when working with big data aligns with the job relevance construct in TAM2 described by Venkatesh and Davis (2000) as the perceived importance of a technology or system to an individual's job. Employees at organizations that have transitioned to big data technologies to solve a business use case have been more effective and could achieve more with big data technologies rather than relying on traditional methods (Caesarius & Hohenthal, 2018).

Participants also stated that employee involvement to identify a big data transition use case was important in achieving their business goal of relying on data for data-driven business decisions. P1O4 indicated that mandating employees to participate in defining a business use case is important in choosing the right big data technology and has a positive impact on the result or output. This relates to the output construct of the TAM2, which is the degree to which a given task corresponds to an individual’s goals and how well that given task performs (Venkatesh & Davis, 2000). For instance, data scientists are more likely to implement big data technologies to enable data-driven business decisions within an organization when they have a clear understanding of the use case (Kim, Zimmermann, DeLine, & Begel, 2017). Similarly, organizations are more likely to adopt big data technologies to enable data-driven business decisions when employees are involved in the overall process and have a proper understanding of the use case or problem to be solved with big data technologies (Wimmer & Aasheim, 2019). This idea supports participants’ views that employee involvement in defining the use case for big data transition supports the organization’s goal of relying on data for data-driven business decisions.

Based on Theme 1 from the data, organizational leaders and data managers should start by identifying a business goal or use case as they plan to transition to big data technologies. Defining a business goal will help identify specific big data technology suitable for the use case to maximize the chances for success. The process of defining a business goal should involve multiple stakeholders from different departments of the organization to encourage collaboration and teamwork. Defining the business use case will also enable technical teams responsible for the implementation to focus on the specific aspects of the chosen big data technologies and determine the most suitable way to implement the technology rather than testing each big data technology to see if it fits. Employees will be more willing and motivated if they get involved at an early stage and understand that their views are important and have higher management support. After identifying and defining a use case, senior management can create a transition methodology that will include implementation guidelines and reviews, status meetings, lessons learned, and next steps or retrospective project meetings to define the next steps.

Theme 2: Identify Data Sources

The second theme that emerged from the data analysis is the need to identify all potential data sources for a successful transition to big data technologies from traditional DW technologies. Identifying all existing data sources will not only help in choosing the right big data technology but will also enable data managers to understand where and how to properly distribute available resources as the organization transitions to big data technologies. By correctly identifying potential data sources, organizations may address some of the unknown facts associated with data that the organization possesses but have never been used or explored: (a) which departments produce what type of data (b) which key elements of data from which source have value, and (c) what types of metrics can be efficiently generated from each data source. Determining these details early in the transition process is essential for a successful transition.

Participants stressed that it was significant for their organizations to identify all potential data sources before starting a big data transition process and explained that data could typically emanate from different both internal and external depending on the business type or organizational structure as well as on the application requirements sources (see Table 2). P1O1 explained that though it is important that organizations properly identify their data sources, it is also important that they determine whether data coming from those data sources are structured, semistructured, or unstructured. P1O4’s response when asked how organizations can ensure proper usage of a big data technology was that it is important to first identify the data source because identifying the data source will determine the big data technology to use. P1O4 stated that “The type of data you have to deal with will impact how you store and how long you will need to process it before it can be analyzed.”

P2O4 in support of properly identifying the data sources provided an example. P2O4 explained that most or some web pages almost exclusively contain images and text as found on most e-commerce websites. These data, he continued, as presented on such websites are either semistructured or unstructured, making it difficult to parse and process. P2O4 also asserted that customer data information such as name and contact information are different from text data as found on websites because it is effortless to parse and is structured by design. Moreover, P2O4 stated that “understanding the type of data you are working with is critical to choosing the right big data technology and the overall success of a big data implementation project.” Other participants like P1O6 mentioned that data can equally originate from internal sources like enterprise resource planning systems or from internal company websites, which can be processed and parsed easily and efficiently with big data technologies and presented in a compelling way to increase internal processes and improve on the effectiveness of employees.

Table 2

Theme 2 Frequency of Occurrence

Source of data collection	*t*	*n*
Interviewees	10	14

Documents	15	22

Note. Theme 2 identify data sources; *t* = total number; *n* = frequency of occurrence.

Organizational documents reviewed also provided similar evidence of the importance of identifying data sources before starting a big data transition project. My review of the documents further highlighted the importance of first identifying all possible data sources for a successful transition to big data technologies. "*The Five Essential Components of Data Strategy*" document outlines the five components for a data strategy starting with the identification of data sources. The document contended that identifying data sources and understanding its meaning regardless of structure, origin or location is vital to the success of a big data strategy. Other documents reviewed also supported this theme, providing guidelines in some cases like the "*Data Discovery*" document which provides guidelines that organizations can implement discover various data sources within and out of the organization, or highlighting its importance as in "*The Big Data Challenge and Success Factors*" document which listed as third most important factor required for a big data success project, the need to determine all data sources before starting a big data project. These documents support the views shared by all participants that starting a big data project without understanding or knowing the data sources might be a recipe for failure. P2O5 stated "the type of data determines the big data technology." A view shared by P1O6 and P1O2 who both explained that streaming data from sensors for example (which is unstructured data) would require a different big data technology than say data from web sites that are also unstructured data.

Understanding these multiple sources will help determine not only the big data technology but the skill level required, they asserted.

Similar evidence on the importance of discovering all data sources for a successful big data transition project was also found in existing scholarly literature. Blazquez and Domenech (2018) claimed that identifying not only the data sources but the type of data they provide and how these data should be processed is fundamental to a successful big data transformational project in a company. Additionally, Blazquez and Domenech (2018) proposed architecture for integrating and processing data stemming from various sources for big data projects or systems. Wolfert, Ge, Verdouw, and Bogaardt (2017) in their review of Big data application in smart farming insisted that identifying and integrating data from a variety of sources, both traditional and new, with multiple big data tools, was the first prerequisite for success. These studies support the statements uttered by P1O3 as well as P1O2 who explained that the number one prerequisite for a successful transition to and using big data technologies is to discover and recognize the type of data, and this can only be properly done if the data sources are identified beforehand. Nathali Silva, Khan and Han (2017) found that the unification of different data sources for any big data project represents a necessary step to ensure success. This is a view equally shared by P1O4 who explained that the success of a big data transition project from DW technology will primarily depend on your understanding of the multiple sources that fed data into the traditional warehouse and beyond.

Technological developments, digitization, and social networking introduced new data types (semi-structured and unstructured) to the scope of enterprises that need to be stored, processed and analyzed. A traditional DW's primary purpose is to integrate, store and consolidate data in a rigid multidimensional structure (Sebaa, Chick, Nouicer, & Tari, 2017). This principled architectural design of DWs makes it difficult to include heterogeneous data types like streams of data from sensors or unstructured data from a company's website. Salinas and Lemus (2017) found that the limitation in traditional DW technology to integrate data from multiple sources is a reason why many organizations will transition to using big data technology which provides more flexibility in working with data from numerous sources. A consistent view with most participants who explained that lack of scalability and the inability of traditional DW to integrate data from multiple sources including semi-structured and unstructured data was one of the multiple reasons their organizations decided to transition to big data technologies. Sebaa et al. (2017) elucidated that for organizations to comprehensively explore data from multiple sources, they need to transition from traditional DW technologies to big data technologies like Hadoop and Hive. These studies support the views reported by participants as well as the organizational documents reviewed.

The conceptual framework provided insights into the participants' perceived usefulness of identifying all potential sources before a big data transition project. Participants believed that identifying all potential data sources is important for their job and a mandatory process of their big data transition process. P1O4 indicated identifying the data sources is mandatory and follows guidelines set by senior management. Organizational documents reviewed also highlighted the fact the identifying all potential data sources should be an absolute requirement implying it is mandatory. This statement from P1O4 aligns with Venkatesh and Davis (2000) who found in TAM2 that subjective norm has a positive impact on perceived usefulness. Moreover, Kim et al. (2017) found that data scientists perceiving using big data technologies to work with multiple data sources to enable data-driven business decisions are more likely to adopt such technologies. P1O5 reported that identifying data sources involved different stakeholders and required management review to make sure the process is properly executed. Furthermore, P2O4 indicated that before starting the transitions process, a list of all potential data sources was created and each stakeholder had to understand their implication. This statement aligns with Venkatesh and Davis (2000) who found in TAM2 that the status or reputation of an individuals' social group significantly increases perceived usefulness. Moreover, Vidgen, Shaw, and Grant (2017) found that data scientists would likely use big data technologies after identifying data sources to enable data-driven business decisions if this is important to others and the organization. Subjective norm had a positive impact on participants' perceived usefulness of identifying data sources in their big data transition project and their intention to use big data technologies.

Study participants believed that mandatory use of all potential data sources for big data transition improves success rate, is effective, provides quantifiable results and is relevant to their job. These outcomes are consistent with Venkatesh and Davis (2000) who posited in TAM2 that output quality, job relevance, and result demonstrability influence perceived usefulness and use intention. P1O6 explained that making it mandatory to have all stakeholders from different organization departments participate in identifying data sources improved the success rate in their big data transition process. Del Vecchio, Mele, Ndou and Secundo (2018) asserted that data managers that require identification of data sources as a mandatory step in a big transition project are more likely to achieve expectations. Additionally, Verma et al. (2018) found that organizations are more likely to adopt big data technologies if their usage is mandatory and data managers or scientists think it is important for their job and will improve their efficiency and lead to firm-level benefits.

Study participants also indicated that identifying data sources before a big data transition project helps in reducing the time to perform their tasks. P1O4 reported that identifying all potential data sources enables them to develop appropriate tools to work with different data types effectively, quicker and with more efficiency. Venkatesh and Davis (2000) found in TAM2 that perceived ease of use significantly influences perceived usefulness which is consistent with what P1O4 reported. Employees are more likely to adopt big data technology if it can perform useful functions for them. A consistent view with the findings from the study conducted by Wang, Li, and Zhao (2017) who found that data analysts would be motivated to use big data-processing techniques or tools if they perceived that these tools or techniques are easy to use and can perform useful functions for them. These studies support the view shared by participants that getting all stakeholders involved in identifying data sources is not only an essential step in a big data transition project but also improves efficiency and are key to success.

Data managers and senior organizational leaders should make it a mandatory step to identify all potential data sources including sources that feed data into DW and beyond as part of their big data transition process from DW technologies. These sources should include structured data (like those that feed data into DW), semi-structured data (like data from internal systems stored as XML files), and unstructured data (like data from company websites, sensors, and Internet-of-things). The process of identifying potential data sources for a big data transition project from DW technologies should involve all stakeholders including sponsors (both internal and external) involved in the transition process. Data managers should set formal processes to verify and approve the data sources for a successful big data transition.

Theme 3: Top Management or Executive Support

Embarking on a big data transition project from DW not only requires technical skills and commitment from all stakeholders but most importantly executive involvement to provide strategic leadership in terms of communication and project execution. The third theme highlights the need for executive leadership championship for the transition from DW technology to big data technologies. in an organization. The top management support theme embodies the qualities (e.g. executive leadership) required for empowering and facilitating a big data transition project from DW technologies in an organization. Santos-Vijande, López-Sánchez, and Pascual-Fernandez (2018) mentioned that executive management support is one of the critical factors for project success. Santos-Vijande et al. (2018) also explained the different reasons why executive management support is required including but not limited to clarifying strategic objectives, facilitating project funding, project governance, providing project resources, project buy-in, promoting cultural change, decision making and managing risks. Study participants while all agreeing to the need for executive support for a big data transition process, viewed executive-level involvement differently.

Most participants shared the views (Table 3) that for the success of a big data transitions process, executive management support is required. P2O5 and P1O6 stressed that without executive management to set the tone and facilitate the adoption of big data technologies, the transition would represent a failure. P1O5 stated. “the primary objective of senior management involvement is to set and clarify strategic objectives as well as project expectations in a well-defined manner so that the transition is executed and completed in line with an objective that serves the overall business purpose.” Another significant area where executive management is essential is securing project resources in terms of training and management of functional teams he continued. P1O2 explained that executive support is mostly required in the form of pushing for organizational buy-in or internal support for big data. He contended that given that most senior management leaders are not technically savvy, you would not want them to be involved in the technical decisions, rather encouraging people within the organization to sponsor the project and limit their role to facilitating decision processes. A position that P1O3 did not fully agree with. He stated. “Top management was key to our transition to big data in the decision-making process”. Several decisions must be evaluated when an organization decides to transition to big data, most of which are not always related to the technical skill level, he continued. Funding approvals and allocation, scope changes authorizations, schedule extensions are areas where he explained top management support was most impactful.

Table 3

Theme 3 Frequency of Occurrence

Source of data collection	*t*	*n*
Interviewees	10	12
Documents	15	19

Note. Theme 3 top management or executive support; *t* = total number; *n* = frequency of occurrence.

The organizational documents reviewed for this study also provided information that aligned with or supported this theme. According to the "*Big Data Maturity: An Action Plan for Policymakers and Executives*" business executives must prioritize the following for a successful big data transition depending on the competitive environment of the organization, business model, and existing internal capabilities: (1) develop a clear big data strategy (2) formally establish a "Chief Data Scientist" position and make him/her owner of big data initiatives in the organization, (3) position big data as integral element of the organization's operating model, and (4) launch a communication campaign to promote a data-driven decision culture in the organization. Similarly, the "*Realizing the Promise of Big Data*" document asserted that before starting a big data transition project, executive leaders (e.g. Chief Information Officers or CIOs) must start by building a coalition around big data that includes reaching out to peers, define the broader opportunity and securing sponsorship. The assertions in the "*Realizing the Promise of Big Data*" document aligned with what participants reported. P1O6 mentioned that a big data transition project must start with the foundation of a strong coalition in the organization that includes various stakeholders with strong top management support.

The “*Building Analytics Driven Organization*” document highlights four major points that need to be guaranteed before an organization embarks on its big data journey: (1) Sponsorship: who is responsible for funding and governance of big data projects? (2) Leadership: who oversees realizing the big data transition? (3) Funding: how is the big data funded and who is responsible? and (4) Governance: who is responsible for governance? These documents support views shared by participants that top management support is an absolute requirement for a successful pre and post-transition to big data technologies from DW technologies.

Similar evidence on the importance of executive support for big data transition project was equally found in the existing literature. Sun, Hall, and Cegielski (2019) argued that top management support in big data projects is a strategic decision and not only helps in fostering a positive environment, ensure adequate resources for adopting big data, and coordinate the process of accepting big data among organization members but also to secure critical resources needed to support the implementation of the big data transition process. Sun, Cegielski, Jia, and Hall (2018) found that top management support ranked fourth (out of 24) in terms of factors indispensable to the success of big data adoption in an organization. Sun et al. (2018) also found that top management support is most associated with the decision-making process to secure resources for big data adoption. These results align with views from P1O1 who stated that “sometimes the success of a big data transition project comes down to who makes the decisions.” Dremel, Wulf, Herterich, Waizmann, and Brenner (2017) contended that the fact that the requirements are not always very explicit from day one when starting big data analytics project is strong reason why a Head of Big Data IT Competence Center needs to be created with top management support to manage the entire big data adoption process. These studies align with the organizational documents reviewed as well as the views shared by study participants on the importance of top management support when transitioning to big data technologies from DW technologies.

Participants believe that top management support in big data transition projects improves their job performance and is critical to success. P1O4 and P1O5 reported that top management support improves their work quality and efficiency. This statement aligns with Venkatesh and Davis (2000) who found in TAM2 that support from someone with powers in the organization will have a positive impact on output quality and increase the likelihood to accept a technology if it is perceived to improve a users' task. Cabrera-Sánchez and Villarejo-Ramos (2020) further highlighted the importance of management support, concluding that top management support in big data projects is an indication of the perceived usefulness of big data technologies in improving one's job performance. P1&2O4 and P1&2O5 explained that top management support improves big data awareness, fosters integration and increases big data perception across the organization. This aspect of top management support aligns with Venkatesh and Davis (2000) who found that result demonstrability had a positive effect on perceived usefulness. Ooi, Lee, Tan, Hew and Hew (2018) found that top management's support for improving and integrating technology into business processes, creating big data awareness and visibility had a positive effect on perceived usefulness. P1O2 reported that senior management support is a catalyst for team communication and comes with a social component that improves team cohesion. A thought equally articulated by Dubey et al. (2019) who asserted that the top management support encourages team cohesion which has a positive effect on image and output quality as described in TAM2.

P1O3 and P1O5 reported that top management support is required to secure proper training programs and adequate infrastructure to support a big data transition project. Similarly, P1O2 reported that top management helped secure customized training that helped employees develop confidence in using big data technologies. Securing essential training via top management is consistent with the job relevance construct in TAM2 described as the individual perception of matching the correct technology with one's job. Views shared by Dwivedi, Rana, Jeyaraj, Clement and Williams (2019) who found that individuals found facilitating conditions such as training programs, experience sharing forums as a motivation to adopt a technology and recommended that organizations through their leadership should provide such facilitating conditions so that employees can be positively inclined to use the technology.

Top management, according to Siddique and Ganguly (2019) is responsible for charting a company's vision and steering the organization in the appropriate direction, instilling a sense of identity while nurturing organizational values. Therefore, for a big data transition project from DW technologies, top management support's role cannot be underrated. Top management should provide the vision, provide the necessary support, encourage communication amongst peers and knowledge sharing, and commit to establishing a healthy working environment. Additionally, top management should provide support in terms of facilitating big data acceptance within other departments of the organization.

Theme 4: Create a Data Lake—Big Data in the Cloud

Most organizations use DW as a repository to aggregate and store data from multiple sources from where they perform different types of analytics on the data. Given the architectural and design limitations in DWs, especially when dealing with new data types like semi-structured and unstructured data, an organization planning to make data-driven decisions will be limited merely to a limited subset of data and could lose its competitive advantage. The fourth theme that emerged from the data analysis (Table 4) is the recommendation from most participants that an organization aspiring to transition to big data technologies from DW technologies to make data-driven business decisions, should start by creating or using a data lake offering from the cloud. A data lake as described by Nadal et al. (2017) is a centralized repository that unlike DW which follows a schema-on-write approach, stores data in its native format (structured, semi-structured and unstructured) at any scale. A data lake stores data as they are produced without any requirement for preprocessing and can be utilized directly for analytical purposes using the load-first model-later principle (Nadal et al., 2017). Research participants all provided numerous arguments through their lived experiences to highlight the importance of having a data lake when transitioning to big data technologies from DW technologies.

Research participants through their various interventions stressed the importance of data lakes in big data transition projects. P1O4 stated that “for me, the first step is to establish a data lake" while answering to the question on what strategy should organizations use to work with multiple data types when transitions to big data. He further explained that developers working with multiple data types should be allowed to push all the data they generate (structured, semi-structured and unstructured) into a data lake. P1O2 and P1O3 recommended not building a data lake locally but using a data lake in the cloud. They explained that storing huge amounts of data locally could be expensive depending on the size of the organization and compliance requirements and installing and maintaining a big data infrastructure could be very expensive. P1O6 explained that "due to technological innovations, high-speed broadband has become widespread and easily available meaning data can easily be moved between on-premise datacenters to the cloud." The significance of that he explained is that data produced locally by organizations can effortlessly be moved to the cloud for analysis and must not be analyzed locally. P1O5 and P1O3 mentioned that another reason to use a data lake in the cloud rather than building one is that more and more organizations produce applications that are cloud-native or cloud-based. Using a data lake in the cloud has the advantage of providing storage as required for data produced by applications and systems and ready for analysis without the need to invest in extensive infrastructures to meet additional storage requirements or spikes in demand. Hence, a hybrid setup will be well-suited for big data applications they asserted.

Table 4

Theme 4 Frequency of Occurrence

Source of data collection	*t*	*n*

Interviewees	10	11
Documents	15	14

Note. Theme 4 create a data lake – big data in the cloud; *t* = total number; *n* = frequency of

Organizational document reviews for the research also highlighted the importance of a data lake in a big data transition process. The “Data Lakes: A Solution or a New Challenge for Big Data Integration” stated that a data lake is an absolute requirement for integrating data from multiple sources when transitions to big data as it would “provide a uniform way to query and retrieve data, ignoring its original format.” The document concludes that a data lake would help organizations transitioning to big data technologies from DW technologies to adopt an extract load transform methodology (ELT) rather than the extract transform load (ETL) as in traditional DW. Similarly, the “Building a Data Lake with AWS” document states that as a fact that a data lake is necessary for successful big data implementations and went further to explain why the cloud data lake solution provided by Amazon cloud (AWS) should be strongly considered as organizations plan their big data transition projects. It claimed that data lake from AWS is a centralized, secure, and durable solution that will allow organizations to ingest and store all sorts of data and transform the data as required. This assertion is in line with P1O3 who mentioned that their organization relied on the data lake solution on the AWS platform to store, explore and analyze their data during their transition process without the need to invest in expensive hardware.

P1O4 mentioned he would recommend a data lake in the cloud not only because of the prize advantage but because cloud providers have already done the necessary research in building an optimizing data lake for optimal success. This is a thought also shared in the "*Data Lakes, Purposes, Practices, Patterns, and Platforms*" document which asserted that a well-designed data lake is an effective data-driven design pattern for capturing a wide range of data types at scale and cloud providers have invested both in time, knowledge and infrastructures to make big data transitions easy for organizations who leverage these cloud services. P105 also mentioned that some big data applications that require compute-intensive hardware like in research are also good candidates for cloud data lakes as organizations could rely on cloud hardware to run complex research sequences and store the data in the cloud data lakes for quick analysis. This thought was also consistent with the "*Cloud Customer Architecture for Big Data and Analytics*" document which highlighted that a cloud data lake would be optimal in Research and Development environments (R&D) where a data lake can be used to automate the management and analytics of data as it flows in, out and through the data lake.

Similar evidence was also found in existing scholarly literature. Wahyudi, Kuk and Janssen (2018) found that a data lake is vital in a big data transformation project as it enables organizations to store any type of data, extract and transform the dataset based on the task at hand. They also found that a data lake is not only necessary for storing data from various sources, but also facilitates access to the source as well as data reuse. Similarly, Kitchens, Dobolyi, Li and Abbasi (2018) and Nadal et al. (2017) reported that a data lake is the solution to the current problem of collecting, aggregating and storing data from different sources. Further, Nadal et al. (2017) found that data lakes remove some of the barriers encountered in other technologies like DW to support seamlessly sharing data from different sources across an organization to support analytical processes. This thought is shared by P1O4 who mentioned that it was critical for them to not only make sure data from the different sources were stored in a central location but made available on time to the different teams working on different big data analytics pipelines. Stefanowski, Krawiec, and Wrembel (2017) also found that data lakes are the only feasible technologies for a successful big data transformation. The value that a data lake brings in aggregating and working with data of different types is superior to what DW technologies offered they asserted.

Study participants found that using a data lake greatly improved their performance during their respective big data transition projects. P1O2 and P1O5 mentioned that a data lake enabled them to work with different data types efficiently and become more productive. This countenance is consistent with Venkatesh and Davis who found in TAM2 that perceived ease of use significantly influences the acceptance of technology. Shafiee, Barker, and Rasekh (2018) found that a data lake facilitates big data analytics and eases the work of technical teams implementing big data projects. The perceived usefulness of data lakes was key to big data technology acceptance and usage as reported by study participants. P1O3 mentioned that using a data lake in the cloud reduce their overall big data implementation time, increased team collaboration, especially with remote team members. The TAM2 model suggests that subjective norm has a positive influence on image because a user’s image is elevated in a group if that user's workgroup considers it important to perform a task (like using a technology). This assertion is highlighted by Lehrer, Wieneke, vom Brocke, Jung and Seidel (2018) who found that by using data lakes in their big data initiatives to improve customer relations, employees from four business sectors (insurance, banking, telecommunications, and e-commerce) created a positive impact between customers and their respective organizations. Using a data lake gave employees a positive feeling and sense of importance or elevated their image amongst their peers.

P1O4 explained that mandatory usage of cloud data lakes for their big data transition project was critical for success, improved efficiency, and facilitated knowledge sharing amongst peers. Using technology in a mandatory setting aligns with Venkatesh and Davis (2000) who described in TAM2 that people will likely use a technology whenever somebody with power thinks they should do so, even if they are not totally in favor of using the technology and their potential consequences. The perceived usefulness of data lakes as described in TAM2 will be directly influenced by these external expectations and indirectly by the individual's self-image of the others. A view underscored by Munshi and Mohamed (2018) who found that data scientists will likely use data lakes in big data implementations because it allows quick consolidation of various data types into one repository. Similarly, Shepherd, Kesa, Cooper, Onema and Kovacs (2018) found that the perceived usefulness construct in TAM2 could be measured by data managers' belief that data lakes would improve efficiency, facilitate teamwork and would be important for success. These studies support participants' views that using a data lake was important to their job and important for success as they transitioned to big data technologies.

Transitioning to big data can lead to significant breakthroughs that can lead to more data-driven decisions and give a competitive edge for organizations that decide to leverage the potential of big data technologies. Transitioning to big data can be complicated if the right processes are not properly executed. Organizations should leverage innovative solutions like data lake either by building one or using a cloud provided data lake. The latter should always be considered except for regulatory purposes, for reasons explained in multiple sections above. A successful transition to big data technologies from DW technologies will enable organizations to ask and answer more questions, empower the workforce, and enable a more transparent and accurate decision-making process in the organization all powered by data, and implementing or using a cloud data lake might just be the first step in the right direction.

Applications to Professional Practice

The apparent lack of strategies by data managers to transition to big data technologies from traditional DW technologies is the specific IT problem that constituted the basis for this research. Participants through semi-structured interviews provided strategies that data managers use to transition to big data technologies from DW technologies.

Relying on available documentation on big data and industry standards is a significant requirement to consider by organizations as they plan to transition big data technologies. Most participants stated that they strongly relied on best practices as outlined in the existing documents and in some cases where applicable, relied on industry standards as guidance and implemented additional components as required. Adhering to existing documentation and recommended industry standards will enable organizations to create organization-specific workflows and document their progress as they execute the strategies to transition to big data technologies. This process will instill confidence and trust among all stakeholders and in the process. The documentation and workflow process should include the input of every stakeholder to ensure everyone is on the same page as the project progresses. The documentation process should also capture what worked, unify all information related to the big data transition project and provide guidelines including but not limited to step by step implementation processes, walk-throughs, and deep dives into the different big data solutions.

As one participant mentioned, transitioning to big data technologies will not only be a choice but a requirement for organizations that want to retain their competitive advantage. As organizations plan the transition process to big data technology from DW technologies, adopting an agile framework or methodology will be key to success. Adoption of agile practices will facilitate exchange and communication within the organization as agile principles will enable fast decision cycles and introduce rapid learning methodologies enabled by technology that will create value for all stakeholders. This will allow organizations to focus on the areas that will make them have quick and notable successes in the early stages of the transition period that they can build upon. The degree of success will ultimately be determined by the role senior leaders will play. Providing strategic and actionable guidance that is not only clear but coherent around the priorities and expected outcomes at both the technology and team levels as well as ensuring the proper channels exist for frequent feedback and coaching that will boost confidence, facilitate autonomous work and increase the likelihood of a positive outcome is the role expected from senior leadership. Participants unanimously agreed that effective leadership from senior management will be boost morale and increase the probability of success of the transition process.

Participants highlighted the fact that investing in employee professional training is something to strongly consider for data managers or organizations aspiring to benefit from the innovations that come with transitioning to big data technologies. Transitioning to big data technologies from DW technologies could lead to significant employee improvement that will be reflected in the business if team members and staff are trained with the precise skills to work with chosen big data technologies to improve their knowledge. While some organizations could rely solely on external contractors to execute their transition process to big data technologies, research participants cautioned about outsourcing or relying on external contractors for the execution of the transition process as this could be risky and would hurt the organization's flexibility and expected impact over a sustained period. As organizations look to either maintain or increase their competitive advantage by transitioning to big data technologies, employee training is a key competitive priority to consider. A well-trained team capable of working un-assisted with big data technologies is an important part of the overall strategies for transitioning to big data technologies from traditional DW technologies.

Organizations aspiring to successfully transition to big data technologies from traditional data warehousing technology should perform a readiness assessment. This assessment will enable organizational leaders to map their resources to the requirements and either get the assurance that their organization is ready for the transition to big data or make the necessary adjustments. Due to the complexity both in terms of technology and people when transitioning to big data technologies, organizational readiness assessment will provide organizational leaders the assurance and knowledge that the proposed venture or endeavor (transitioning to big data technologies) will be successful and the success can be maintained and build upon. As one of the participants mentioned, "without readiness assessment, the transition was heading to a failure before we even started."

Implications for Social Change

This study explored the strategies used by data managers to transition to big data technologies from traditional DW technologies. The findings from this study will add to the existing knowledge in scientific research and literature in the field of big data. The findings can be used by an organization not only to increase or maintain its competitive advantage but additionally serve as a basis for positive social change. The results from this could serve as a premise to raise awareness in support of developing effective strategies and policies to transition from traditional DW technologies to big data technologies.

The process of transitioning to big data technologies from traditional DW technologies brings together people of diverse backgrounds with a different perspective and culture and diverse skills set to accomplish a common goal. Knowledge sharing is a common trend when people work together to achieve a common objective which could have a direct impact on an organization and society in general. One of such implications for positive social change is that data managers and organizational leaders might use the research findings to optimize their data management strategies and decisions which will have a significant benefit to consumers using their services. As data management becomes refined and structured in an organization, so will be the services and products offered. Transitioning to big data might give the possibility for service or product consumers to track and monitor their actions and engagements with these services or products, and thus gain a better understanding of their own lives.

Information security and transparency is another area where transitioning to big data technologies could have a positive social change. Transitioning to big data would lead to the development of applications that run both on stationary and mobile devices to provide much needed access to services like transportation, medical care, nutrition, and security, which are vital pieces of any society. By transitioning to big data technologies, organizations can implement features that consumers can use to verify the type of information accessed and what they were used for on an individual level. This will increase transparency and consumer confidence. Fear of government sanctions and loss of consumers will push more organizations to transition to big data technologies and enable more transparency which will have a downstream effect on consumers and society in general.

Recommendations for Action

I explored the strategies that Data managers use to transition to big data technologies from DW technologies. The findings obtained from the study showed that successful transition to big data technologies from DW technologies requires substantial investments in resources in terms of finance, people, time and executive commitment.

Top management support is an absolute requirement for a successful transition to big data technologies from DW technologies. Senior leadership in organizations need to understand that gone are the days where having more expensive and powerful machine gave you a competitive edge and view their data (structured and unstructured) not just as random storage, but as an asset. As such, executive leaderships need to clarify strategic objectives that highlight goals and desired outcomes and secure necessary resources to meet those goals as far as using their data asset to increase or maintain their competitive advantage. Additionally, the organization must undergo a cultural change and be willing and open to using new technologies, implement new processes and explore new business models. Given that transitioning to and using big data technologies is a continuous process, senior leaders should encourage and verify adherence to best practices through peer review processes at every stage of the transition process.

Organization leaders should encourage and set standards for information and knowledge sharing within participating teams during and after the transition process. The whole process should be geared towards building a culture of sharing and knowledge development within the organization. Transitioning to big data will require employees to work with and adapt to new technologies. For some employees it could probably be the first time they use such technologies, so developing a strategy to make information available at all levels is vital for success. One way of developing this strategy is for the organization to invest in effective training to develop a knowledge-sharing culture within the project and the organization. The strategy should ensure that team members share a common vision with an acceptable team-working culture void of any competition amongst the employees. To avoid any competition amongst employees, giving that people would normally have the habit of hiding knowledge from their peers, senior leaders should develop as part of the knowledge-sharing culture effective interventions such as coaching and mentoring to address any potential crises or lack of knowledge sharing.

Senior leaders need to set standards for evaluating success through every stage of the implementation by effectively mapping and defining stages, key milestones, deliverables and criteria for acceptance for all the teams involved in the transition process. Senior leaders could use industry standards and benchmarked practices available in official documents to define and set the deliverables. Documenting success as well as failures at every step will help in evaluating the success of the process and highlight areas of improvement. Documenting every aspect of the transition process will have each team member to have complete knowledge of their responsibilities, a clear understanding of the expectation and how should contribute and manage their work.

The findings from this multiple case study research will be shared with the organizations and participants via email. Overall, this research might be of benefit to big data professionals, data scientists and the entire big data community including governmental and non-governmental organizations. Wherever possible, I intend to share the findings of this research through presentations, conferences, lectures, and speeches to select individuals or groups of interests. It is also my hope that the final published study will be available and open for public searches when organizations search for strategies to transition to big data technologies from traditional DW technologies.

Recommendations for Further Study

For this multiple case study, I interviewed data managers who had transitioned to big data technologies from traditional DW technology. This study was only limited to exploring strategies data managers use to transition to big data technology from traditional DW technologies without considering any security and privacy implications for the data owners. I recommend additional research where the entire big data pipeline will be reexamined with security and privacy in mind. Further research could investigate the appropriate privacy and access control policies to be enforced to ensure big data usage does not infringe on the privacy of data owners and is used only for legitimate purposes. I would also recommend further exploration of the themes identified during this study. Extensive analysis or examination of each theme could highlight and record the impact and influence of these themes within organizations and provide them with supplementary information to execute their transition process and mitigate any unexpected outcomes. This research was also limited to the fact that there was no in-depth look into a specific big data technology and its impact on the transition process. This research was too broad and was not limited to exploring the role played by one specific big data technology and giving there are multiple big data technologies, a deep dive into one specific big data technology might be warranted. Lastly, another recommendation would be for researchers to explore the strategies to transition to big data technologies but focus or limit their research to a specific big data technology like Hadoop, Spark, HIVE, and Cloud big data offerings for example.

Reflections

As I reflect upon this study, which was not only filled with obstacles but as well with enlightenment, I have an overall feeling of satisfaction. Though everything from the start of the doctoral project was not all that clear in terms of expectations, patience was a crucial and essential factor required for such a demanding and long project. Every time I encountered an obstacle, though difficult, I persevered and grew my knowledge in the topic and the process overall. I learned to be patient as the project advanced and communication was a significant factor to grow my patience. It was important to maintain constant communication with my research mentor and peers through regular discussion forums, emails, and telephone calls as required.

My doctoral project, “strategies to transition from traditional data warehousing technologies to big data technologies”, was very significant to my career goals as an IT professional. Before researching this topic, I could not understand all the concepts around big data and how they related to traditional data warehousing technologies. Although working as a big data consultant at Microsoft exposed me to big data technologies and DW technologies, researching the topic enable me to dig deeper into areas I never explored in my work as a big data consultant. I learned through research methodology processes and interactions with participants how to conduct research and report findings at an academic level as well as understand the different ways big data technologies can be implemented and used and the impact they could have on society. Due to the semi-structured nature of the interview questions, there is a possibility that through my interactions with the participants, I unintentionally biased the research. I acknowledged potential biases throughout the study wherever possible and did my utmost best to ensure the reliability and credibility of the study.

Summary and Study Conclusions

Transitioning to big data technologies is important for organizations to improve their data management capabilities and have more qualitative analysis for data-driven decisions. Transitioning to and using big data technologies effectively in organization is complex and challenging. Applying a lessons-learned approach can significantly increase the likelihood of success pre and post-transition as it will contain enough detail to provide value for future use and should be consistent with lessons learned from previous implementations outside of the organization. Leadership must lead to the organization embracing a multi-faceted approach to innovation and make it a corporate objective. A change of culture is another area where leadership in the organization needs to push and sustain to accommodate any risk and uncertainty that will follow as a result of corporate focus on innovation. To fully enable the culture change, leadership must empower their employees to innovate, accept failure and act appropriately without fear.

Ease of use, system reliability, employee training and senior management support have a direct effect on the usefulness of big data technologies. Senior leaders should create or initiate recognition and award programs to motivate their employees for a job well done or to improve or change work habits to help the organization grow and become more effective. These activities will improve the relationship between employees and senior leadership and hence the overall corporate culture.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2). Retrieved from http://aisel.aisnet.org/jais/

Abelson, H., Anderson, R., Bellovin, S. M., Benaloh, J., Blaze, M., Diffie, W., . . . Rivest, R. L. (2015). Keys under doormats: mandating insecurity by requiring government access to all data and communications. *Journal of Cybersecurity*, *1*(1), 69-79. doi:10.1093/cybsec/tyv009

Abma, T. A., & Stake, R. E. (2014). Science of the particular: An advocacy of naturalistic case study in health research. *Qualitative Health Research*, *24*(8), 1150-1161. doi:10.1177/1049732314543196

Adnan, V. E. (2015). Consumer insight as competitive advantage using big data and analytics. *International Journal of Commerce and Finance, 1*(1), 45-51. Retrieved from http://ijcf.ticaret.edu.tr/index.php/ijcf

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. doi:10.1016/0749-5978(91)90020-T

Ajzen, I. (2012). Martin Fishbein’s legacy: The reasoned action approach. *ANNALS of the American Academy of Political and Social Science*, *640*(1), 11-27. doi:10.1177/0002716211423363

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Alalwan, A. A., Dwivedi, Y. K., Rana, N. P., & Williams, M. D. (2016). Consumer adoption of mobile banking in Jordan: examining the role of usefulness, ease of use, perceived risk and self-efficacy. *Journal of Enterprise Information Management*, *29*(1), 118-139. doi:10.1108/JEIM-04-2015-0035

Amankwaa, L. (2016). Creating protocols for trustworthiness in qualitative research. *Journal of Cultural Diversity*, *23*(3), 121-127. Retrieved from http://tuckerpub.com/jcd.htm

Arnott, D., Lizama, F., & Song, Y. (2017). Patterns of business intelligence systems use in organizations. *Decision Support Systems*, *97*, 58-68. doi:10.1016/j.dss.2017.03.005

Arsel, Z. (2017). Asking questions with reflexive focus: A tutorial on designing and conducting interviews. *Journal of Consumer Research*, *44*(4), 939-948. doi:10.1093/jcr/ucx096

Asadollahi, M., Bostanabad, M. A., Jebralli, M., Mahallei, M., Rasooli, A. S., & Abdolalipour, M. (2015). Nurses' knowledge regarding hand hygiene and its individual and organizational predictors. *Journal of caring sciences*, *4*(1), 45. doi:10.5681/jcs.2015.005

Asrani, D., & Jain, R. (2016). Designing a framework to standardize DW development process for effective data warehousing practices. *International Journal of Database Management Systems*,*8*(4), 15-32. doi:10.5121/ijdms.2016.8402

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, 3-15. doi:10.1016/j.jpdc.2014.08.003

Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., . . . Del Signore, S. (2016). Making sense of big data in health research: towards an EU action plan. *Genome medicine*, *8*(1), 71. doi:10.1186/s13073-016-0323-y

Aydin, G., Hallac, I. R., & Karakus, B. (2015). Architecture and implementation of a scalable sensor data storage and analysis system using cloud computing and big data technologies. *Journal of Sensors*, *2015*. doi:10.1155/2015/834217

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Trasormational issues of big data and analytics in networked business. *MIS Quarterly*, *40*(4). Retrieved from https://www.misq.org/

Bagchi, S. (2015). Performance and quality assessment of similarity measures in collaborative filtering using mahout. *Procedia Computer Science*, *50*, 229-234. doi:10.1016/j.procs.2015.04.055

Bai, Y., Bai, X., Lin, L., Huang, J., Fang, H. W., & Cai, K. (2017). Big data technology in establishment and amendment of water management standard. *Applied Ecology and Environmental Research*, *15*(3), 263-272. Retrieved from http://aloki.hu/

Baker, S. R., & Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from Google search data. *Review of Economics and Statistics*, *99*(5), 756-768. Retrieved from https://www.mitpressjournals.org/loi/rest

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122. doi:10.1037/0003-066X.37.2.122

Banica, L., & Hagiu, A. (2015). Big data in business environment. *Scientific Bulletin-Economic Sciences*, *14*(1), 79-86. Retrieved from https://ideas.repec.org/s/pts/journl.html

Bansal, S., & Rana, D. A. (2014). Transitioning from relational databases to big data. *International Journal of Advanced Research in Computer Science and Software Engineering*, *4*(1). Retrieved from https://www.ijarcsse.com/

Baptista, G., & Oliveira, T. (2015). Understanding mobile banking: The unified theory of acceptance and use of technology combined with cultural moderators. *Computers in Human Behavior*, *50*, 418-430. doi: 10.1016/j.chb.2015.04.024

Barczak, G. (2015). Publishing qualitative versus quantitative research. *Journal of Product Innovation Management*, *32*(5), 658. doi:10.1111/jpim.12277

Barnham, C. (2015). Quantitative and qualitative research: Perceptual foundations. *International Journal of Market Research*, *57*(6), 837-854. doi:10.2501/IJMR-2015-070

Barrett, M., Davidson, E., Prabhu, J., & Vargo, S. L. (2015). Service innovation in the digital age: Key contributions and future directions. *MIS quarterly*, *39*(1), 135-154. Retrieved from https://www.misq.org/

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, *30*(1), 89-116. 10.1007/s00146-014-0549-4

Bedeley, R. T., Ghoshal, T., Iyer, L. S., & Bhadury, J. (2016). Business analytics and organizational value chains: A relational mapping. *Journal of Computer Information Systems*, 1-11. doi:10.1080/08874417.2016.1220238

Begam, N., & Sandhya, E. C. (2017). Big data challenges and techniques. *International Journal of Engineering Science*, *10353*. Retrieved from https://www.journals.elsevier.com/international-journal-of-engineering-science

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28*, 45-59. doi: 10.1016/j.inffus.2015.08.005

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. Open, 2, 8-14. doi: 10.1016/j.npls.2016.01.001

Berg, K. L., Seymour, T., & Goel, R. (2013). History of Databases. *International Journal of Management & Information Systems (IJMIS), 17*(1), 29. doi:10.19030/ijmis.v17i1.7587

Berger, R. (2015). Now I see it, now I don’t: Researcher’s position and reflexivity in qualitative research. Qualitative research, 15(2), 219-234. doi: 10.1177/1468794112468475

Berman, J. J. (2013). *Principles of big data: preparing, sharing, and analyzing complex information*. Amsterdam: Elsevier.

Bhattacherjee, A., & Lin, C. P. (2015). A unified model of IT continuance: three complementary perspectives and crossover effects. *European Journal of Information Systems, 24*(4), 364-373. Retrieved from https//www.palgrave-journals.com/ejis/

Big Data. (n.d.). In Gartner glossary. Retrieved from http://www.gartner.com/it-glossary/big-data/

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member checking: a tool to enhance trustworthiness or merely a nod to validation?. Qualitative Health

Research, 26(13), 1802-1811. doi: 10.1177/1049732316654870

Blake, M., & Gallimore, V. (2018). Understanding academics: a UX ethnographic research project at the University of York. *New Review of Academic Librarianship*, 1-25. Retrieved from https://www.tandfonline.com/

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change, 130*, 99-113. doi: 10.1016/j.techfore.2017.07.027

Botkin, J. R., Belmont, J. W., Berg, J. S., Berkman, B. E., Bombard, Y., Holm, I. A., ... & Wilfond, B. S. (2015). Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. *The American Journal of Human Genetics, 97*(1), 6-21. doi: 10.1016/j.ajhg.2015.05.022

Brière, S., Proulx, D., Flores, O. N., & Laporte, M. (2015). Competencies of project managers in international NGOs: Perceptions of practitioners. *International Journal of Project Management*, *33*(1), 116-125. doi:10.1016/j.ijproman.2014.04.010

Bromley, E., Mikesell, L., Jones, F., & Khodyakov, D. (2015). From subject to participant: Ethics and the evolving role of community in health research. American journal of public health, 105(5), 900-908.

Brown, S. A., Massey, A. P., Montoya-Weiss, M. M., & Burkman, J. R. (2002). Do I really have to? User acceptance of mandated technology. *European journal of information systems*, *11*(4), 283-295. Retrieved from https://link.springer.com/journal/41303

Brubacher, S. P., Powell, M., Skouteris, H., & Guadagno, B. (2015). The effects of e-simulation interview training on teachers' use of open-ended questions. *Child abuse & neglect*, *43*, 95-103. doi: 10.1016/j.childyouth.2016.02.018

Busse, C., Kach, A., & Wagner, S. (2016). Boundary conditions: What they are, how to explore them, why we need them, and when to consider them. *Organizational Research Methods*, 1-36. doi: 10.1177/1094428116641191

Cabrera-Sánchez, J. P., & Villarejo-Ramos, Á. F. (2020). Acceptance and use of big data techniques in services companies. *Journal of Retailing and Consumer Services*, *52*, 101888. doi: 10.1016/j.jretconser.2019.101888

Caesarius, L. M., & Hohenthal, J. (2018). Searching for big data: How incumbents explore a possible adoption of big data technologies. *Scandinavian Journal of Management*, *34*(2), 129-140. Retrieved from https://www.journals.elsevier.com/scandinavian-journal-of-management

Campbell, S. (2014). What is qualitative research? *Clinical Laboratory Science*, 27(1), 3. Retrieved from http://www.proquest.com

Casado, R., & Younas, M. (2015). Emerging trends and technologies in big data processing. *Concurrency and Computation: Practice and Experience*, *27*(8), 2078-2091. doi: 10.1002/cpe.3398

Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. Qualitative Report, 21(5), 811–831. Retrieved from http://nsuworks.nova.edu/tqr/

Charman, A. J., Petersen, L. M., Piper, L. E., Liedeman, R., & Legg, T. (2017). Small area census approach to measure the township informal economy in South Africa. *Journal of Mixed Methods Research*, *11*(1), 36-58. doi: 10.1177/1558689815572024

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, 1165-1188. Retrieved from https://pdfs.semanticscholar.org/fbb1/8d40150508ef2f344359ec345da7fb77cc0a.pdf

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171-209. doi:10.1007/s11036-013-0489-0

Chen, Y. H., & Chengalur-Smith, I. (2015). Factors influencing students' use of a library Web portal: Applying course-integrated information literacy instruction as an intervention. The Internet and Higher Education, 26, 42-55. doi: 10.1016/j.iheduc.2015.04.005

Chen, Z. (2015). Towards Integrated Study of Data Management and Data Mining. *Procedia Computer Science*, *55*, 1331-1339. doi: 10.1016/j.procs.2015.07.117

Cheng, E. W. (2019). Choosing between the theory of planned behavior (TPB) and the technology acceptance model (TAM). *Educational Technology Research and Development*, *67*(1), 21-37. doi: 10.1007/s11423-018-9598-6

Cheung, R., & Vogel, D. (2013). Predicting user acceptance of collaborative technologies: An extension of the technology acceptance model for e-learning. *Computers & Education*, *63*, 160-175. doi: 10.1016/j.compedu.2012.12.003

Cho, M. K., Magnus, D., Constantine, M., Lee, S. S. J., Kelley, M., Alessi, S., ... &

Diekema, D. (2015). Attitudes toward risk and informed consent for research on medical practices: a cross-sectional survey. *Annals of internal medicine, 162*(10), 690-696. doi: 10.7326/M15-0166

Christy, M., Gupta, A., Grumbach, E., Mandell, L., Furuta, R., & Gutierrez-Osuna, R. (2017). Mass digitization of early modern texts with optical character recognition. Journal on Computing and Cultural Heritage (JOCCH), 11(1), 6. doi: 10.1145/3075645

Chu, T. H., & Chen, Y. Y. (2016). With good we become good: Understanding e-learning adoption by theory of planned behavior and group influences. *Computers & Education*, *92*, 37-52. doi: 10.1016/j.compedu.2015.09.013

Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., & Reis, M. S. (2016). How can SMEs benefit from big data? Challenges and a path forward. *Quality and Reliability Engineering International*, *32*(6), 2151-2164. doi: 10.1002/qre.200

Cridland, E. K., Jones, S. C., Caputi, P., & Magee, C. A. (2015). Qualitative research with families living with autism spectrum disorder: Recommendations for conducting semistructured interviews. Journal of Intellectual and Developmental Disability, 40(1), 78-91. Retrieved from https://www.tandfonline.com/loi/cjid20

Cronin, C. (2014). Using case study research as a rigorous form of inquiry. *Nurse Researcher, 21*(5), 19 – 27. doi:10.7748/nr.21.5.19.e1240

Cruz, E. V., & Higginbottom, G. (2013). The use of focused ethnography in nursing research. *Nurse Researcher, 20*(4), 36 – 43 8p.

doi:10.7748/nr2013.03.20.4.36.e305

Cugini, M. (2015). Successfully navigating the human subject's approval process. *Journal of Dental Hygiene, 89*(1), 54-56. Retrieved from http://jdh.adha.org/content/89/suppl_1/54.full.pdf

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13, 319-340. doi: 10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982 – 1003. doi:10.1287/mnsc.35.8.982

Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: implications for smart tourism destinations. *Information Processing & Management*, *54*(5), 847-860. Retrieved from www.elsevier.com/locate/infoproma

De Maio, C., Fenza, G., Loia, V., & Orciuoli, F. (2017). Unfolding social content evolution along time and semantics. *Future Generation Computer Systems*, *66*, 146-159. doi: 10.1016/j.future.2016.05.039

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. AIP conference proceedings 1644(1) 97. doi: 10.1063/1.4907823

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. doi: 10.1108/LR-06-2015-0061

Deuter, K., & Jaworski, K. (2016). Assuming vulnerability: Ethical considerations in a multiple-case study with older suicide attempters. *Research Ethics*, 1-12. doi:10.1177/1747016116649994

Dikko, M. (2016). Establishing construct validity and reliability: Pilot testing of a qualitative interview for research in Takaful (Islamic Insurance). *Qualitative Report*, *21*(3), 521–528. Retrieved from http://nsuworks.nova.edu/tqr/.

Donaldson, D. R. (2016). The digitized archival document trustworthiness scale. International Journal of Digital Curation, 11(1), 252–270. doi:10.2218/ijdc.v11i1.387

Dremel, C., Wulf, J., Herterich, M. M., Waizmann, J. C., & Brenner, W. (2017). How AUDI AG Established Big Data Analytics in Its Digital Transformation. *MIS Quarterly Executive, 16*(2). Retrieved from https://aisel.aisnet.org/misqe/

Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in Big Data and Business Analytics. Industrial and Commercial Training, 47(4), 174-181. doi: 10.1108/ICT-08-2014-0059

Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., Wamba, S. F., & Roubaud, D. (2019). Can big data and predictive analytics improve social and environmental sustainability? *Technological Forecasting and Social Change*, *144*, 534-545. doi: 10.1016/j.techfore.2017.06.020

Dumbill, E. (2012). *Planning for big data*. Sebastopol, CA: O'Reilly.

Dunne, B., Pettigrew, J., & Robinson, K. (2016). Using historical documentary methods to explore the history of occupational therapy. British Journal of Occupational Therapy, 79(6), 376–384. doi:10.1177/0308022615608639

Dutta, D., & Bose, I. (2015). Managing a big data project: the case of ramco cements limited. *International Journal of Production Economics*, *165*, 293-306. doi: 10.1016/j.ijpe.2014.12.032

Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers*, *21*(3), 719-734. Retrieved from https://link.springer.com/journal/10796

Eastin, M. S., Brinson, N. H., Doorey, A., & Wilcox, G. (2016). Living in a big data world: Predicting mobile commerce activity through privacy concerns. *Computers in Human Behavior*, *58*, 214-220. doi: 10.1016/j.chb.2015.12.050

Elgendy, N., & Elragal, A. (2014). Big Data Analytics: A Literature Review Paper. *Advances in Data Mining. Applications and Theoretical Aspects Lecture Notes in Computer Science,* 214-227. doi:10.1007/978-3-319-08976-8_16

Evans, J. R. (2015). Modern analytics and the future of quality and performance excellence. *The Quality Management Journal, 22*(4), 6-17,4. Retrieved from http://asq.org/pub/qmj/index.html

Fathema, N., Shannon, D., & Ross, M. (2015). Expanding the Technology Acceptance Model (TAM) to Examine Faculty Use of Learning Management Systems (LMSs) In Higher Education Institutions. *Journal of Online Learning & Teaching*, *11*(2).

Retrieved from http://jolt.merlot.org/

Fayolle, A., & Liñán, F. (2014). The future of research on entrepreneurial intentions. *Journal of Business Research*, *67*(5), 663-666. https://doi. 10.1016/j.jbusres.2013.11.024

Fernández, A., del Río, S., López, V., Bawakid, A., del Jesus, M. J., Benítez, J. M., & Herrera, F. (2014). Big Data with Cloud Computing: An insight on the computing environment, MapReduce, and programming frameworks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(5), 380-409. doi: 10.1002/widm.1134

Fletcher, J., Sarkani, S., & Mazzuchi, T. A. (2014). A technology adoption model for broadband Internet adoption in India. *Journal of Global Information Technology Management*, 150 – 168. doi:10.1080/1097198x.2014.951294

Furuta, M., Sandall, J., & Bick, D. (2014). Women's perceptions and experiences of severe maternal morbidity–A synthesis of qualitative studies using a meta-ethnographic approach. *Midwifery*, *30*(2), 158-169. doi:10.1016/j.midw.2013.09.001

Fusch, P. I., Fusch, G. E., & Ness, L. R. (2017). How to conduct a mini-ethnographic case study: A guide for novice researchers. *The Qualitative Report*, *22*(3), 923-941. Retrieved from http://nsuworks.nova.edu/tqr/vol22/iss3/16

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The qualitative report*, *20*(9), 1408. http://www.proquest.com

Galdas, P. (2017). Revisiting Bias in Qualitative Research: Reflections on Its

Relationship with Funding and Impact. *International Journal of Qualitative Methods*. doi: 10.1177/1609406917748992

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137-144. doi: https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gangwar, H., Date, H., & Ramaswamy, R. (2015). Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. *Journal of Enterprise Information Management*, *28*(1), 107-130. doi:10.1108/JEIM-08-2013-0065

Gaus, N. (2017). Selecting research approaches and research designs: a reflective essay. Qualitative Research Journal, 17(2), 99-112. Retrieved from http://ezp.waldenulibrary.org/

Gentles, S. J., Charles, C., Ploeg, J., & McKibbon, K. A. (2015). Sampling in qualitative research: Insights from an overview of the methods literature. *The Qualitative Report*, *20*(11), 1772-1789. Retrieved from http://nsuworks.nova.edu/tqr/vol20/iss11/5/

Gonzales, R., Wareham, J., & Serida, J. (2015). Measuring the impact of DW and business intelligence on enterprise performance in Peru: A developing country. *Journal of Global Information Technology Management*, *18*(3), 162-187. Retrieved from http://www.tandfonline.com/loi/ugit20

Gopalani, S., & Arora, R. (2015). Comparing apache spark and map reduce with performance analysis using k-means. *International journal of computer applications*, *113*(1). Retrieved from https://www.ijcaonline.org/

Grady, C., & Fauci, A. S. (2016). The role of the virtuous investigator in protecting human research subjects. *Perspectives in biology and medicine, 59*(1), 122-131. Retrieved from https://www.press.jhu.edu/journals/perspectives-biology-and-medicine

Greener, S. (2018). Research limitations: the need for honesty and common sense. *Interactive Learning Environments*, *26*(5), 567-568. doi: 10.1080/10494820.2018.1486785

Guba, E., & Lincoln, Y. (1985). Naturalistic inquiry. Newbury Park, CA: SAGE

Guha, R. V., Brickley, D., & Macbeth, S. (2016). Schema. org: evolution of structured data on the web. Communications of the ACM, 59(2), 44-51. doi: 10.1145/2844544

Gunawan, J. (2015) Ensuring trustworthiness in qualitative research. *Belitung Nursing Journal 1*(1). 10-11. Retrieved from http//belitungraya.org/BRP/index.php/bnj

Gupta, K. P., Bhaskar, P., & Singh, S. (2017). Prioritization of factors influencing employee adoption of e-government using the analytic hierarchy process. *Journal of Systems and Information Technology*, *19*(1/2), 116-137. doi: 10.1108/JSIT-04-2017-0028

Hadi, M. A., & Closs, S. J. (2016). Ensuring rigour and trustworthiness of qualitative research in clinical pharmacy. *International journal of clinical pharmacy, 38*(3), 641-646. doi: 10.1007/s11096-015-0237-6

Hammer, M. J. (2016). Informed consent in the changing landscape of research. *Oncology Nursing Forum*, *43*(5), 558. doi: 10.1188/16.ONF.558-560

Harriss, D. J., MacSween, A., & Atkinson, G. (2017). Standards for ethics in sport and exercise science research: 2018 update. *International journal of sports medicine, 38*(14), 1126-1131. Retrieved from https://www.thieme.de/

Hart, M., & Porter, G. (2004). The impact of cognitive and other factors on the perceived usefulness of OLAP. Journal of Computer Information Systems, 45(1), 47-56. Retrieved from https://www.tandfonline.com/loi/ucis20

Harvey, L. (2015). Beyond member-checking: A dialogic approach to the research interview. International Journal of Research & Method in Education, 38(1), 23–38. doi:10.1080/1743727X.2014.914487

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management, 36*(5), 748-758. doi: 10.1016/j.ijinfomgt.2016.05.002

Hashem, I. A. T., 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. doi: https://doi.org/10.1016/j.is.2014.07.006

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics, 154*, 72-80. doi:10.1016/j.ijpe.2014.04.018

Heintz, B., Chandra, A., Sitaraman, R. K., & Weissman, J. (2016). End-to-end optimization for geo-distributed mapreduce. *IEEE Transactions on Cloud*

Computing, *4*(3), 293-306. doi: 10.1109/TCC.2014.2355225

Hewitt, J. (2007). Ethical components of researcher-researched relationships in qualitative interviewing. *Qualitative Health Research, 17*(8), 1149–1159. doi:10.1177/1049732307308305

Heyvaert, M., Hannes, K., Maes, B., & Onghena, P. (2013). Critical appraisal of mixed methods studies. *Journal of Mixed Methods Research*, *7*, 302-327. doi:10.1177/1558689813479449

Hilbert, M. (2016). The bad news is that the digital access divide is here to stay: domestically installed bandwidths among 172 countries for 1986–2014. *Telecommunications Policy*, *40*(6), 567-581. 10.1016/j.telpol.2016.01.006

Hoover, S., & Morrow, S. L. (2015). Qualitative researcher reflexivity: A follow-up study with female sexual assault survivors. *The Qualitative Report*, *20*(9), 1476-1489. Retrieved from https://nsuworks.nova.edu/tqr/

Hormann, P., & Campbell, L. (2014). Data storage energy efficiency in the Zettabyte Era. *Australian Journal of Telecommunications and the Digital Economy*, *2*(3). Retrieved from https://ajtde.telsoc.org/index.php/ajtde

Hoyland, S., Hollund, J. G., & Olsen, O. E. (2015). Gaining access to a research site and participants in medical and nursing research: A synthesis of accounts. *Medical Education*, *49*(2), 224 – 232. doi:10.1111/medu.12622

Hu, H., Wen, Y., Chua, T., & Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *IEEE Access, 2*, 652-687. doi:10.1109/access.2014.2332453

Hussein, A. (2015). The use of Triangulation in Social Sciences Research: Can qualitative and quantitative methods be combined? Journal of comparative social work, 4(1). Retrieved from http://journal.uia.no/index.php/JCSW

Islam, M. M. (2014). Dealing with Qualitative Methods in Bangladesh: The Potentials and Pitfalls of Research Design and Fieldwork. Oriental Anthropologists, 14(1), 13-26. Retrieved from http://www.printspublications.com/.

Jaber, M. M., Ghani, M. K. A., Suryana, N., Mohammed, M. A., & Abbas, T. (2015). Flexible DW parameters: Toward building an integrated architecture. *International Journal of Computer Theory and Engineering*, *7*(5), 349. Retrieved from http://www.ijcte.org/

James, N. (2017). Using narrative inquiry to explore the experience of one ethnically diverse ESL nursing student. *Teaching and Learning in Nursing*, 1-6. doi:10.1016/j.teln.2017.08.002

Jee, K., & Kim, G. H. (2013). Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthcare informatics research*, *19*(2), 79-85. Doi: 10.4258/hir.2013.19.2.79

Ji, S., Mittal, P., & Beyah, R. (2016). Graph data anonymization, de-anonymization attacks, and de-anonymizability quantification: A survey. *IEEE Communications Surveys & Tutorials*, *19*(2), 1305-1326. Doi: 10.1109/COMST.2016.2633620

Johnson, P., & Weeks, J. (2016). On the Secure-Domination Number of the Full Balanced Binary Tree with 2n Leafs. International Journal Of Mathematics & Computer Science, 11(2), 173-186. Retrieved from http://ijmcs.future-in-tech.net/

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260. doi: 10.1126/science.aac4520

Jukić, N., Sharma, A., Nestorov, S., & Jukić, B. (2015). Augmenting DWs with Big Data. *Information Systems Management*, *32*(3), 200-209. doi:10.1080/10580530.2015.1044338

Jukic, N., Vrbsky, S., & Nestorov, S. (2016). Database systems: Introduction to databases and DWs. Prospect Press.

Jun, C. J., Lee, J.-H., & Jeon, I. S. (2014). Research about factor affecting the continuous use of cloud storage service: user factor, system factor, psychological switching cost factor. *Journal of Society for E-Business Studies, 19*(1). doi:10.7838/jsebs.2014.19.1.015

Kennedy-Martin, T., Curtis, S., Faries, D., Robinson, S., & Johnston, J. (2015). A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials, 16*(1), 495. doi: 10.1186/s13063-015-1023-4

Kharat, A. T., & Singhal, S. (2017). A peek into the future of radiology using big data applications. The Indian journal of radiology & imaging, 27(2), 241. Doi: 10.4103/ijri.IJRI49316

Kieras, D., & Polson, P. G. (1999). An approach to the formal analysis of user complexity. *International Journal of Human-Computer Studies, 51*(2), 405-434. doi:10.1006/ijhc.1983.0317

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78-85. doi: 10.1145/2500873

Kim, J. (2006). Toward an understanding of Web-based subscription database acceptance. Journal of the Association for Information Science and Technology, 57(13), 1715-1728. doi: 10.1002/asi.20355

Kim, M., Zimmermann, T., DeLine, R., & Begel, A. (2017). Data scientists in software teams: State of the art and challenges. *IEEE Transactions on Software Engineering*, *44*(11), 1024-1038. Retreived from https://www.computer.org/csdl/journal/ts

Kimpel, J. F., & Morris, R. (2013). Critical success factors for data warehousing: A classic answer to a modern question. *Issues in Information Systems*, *14*(1), 376-384. Retrieved from http://www.iacis.org/iis/iis.php

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal Of Management Information Systems*, *35*(2), 540-574. Retrieved from https://www.tandfonline.com/loi/mmis20

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 205395171452848. doi:10.1177/2053951714528481

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, *19*(2), 267-283. doi: 10.1007/s10796-016-9686-2

Koonrungsesomboon, N., Laothavorn, J., & Karbwang, J. (2015). Understanding of essential elements required in Informed Consent Form among researchers and institutional review board members. *Tropical medicine and health*, *43*(2), 117-122. doi: 10.2149/tmh.2014-36

Kornbluh, M. (2015). Combatting challenges to establishing trustworthiness in qualitative research. Qualitative Research in Psychology, 12(4), 397-414. doi: 10.1080/14780887.2015.1021941

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological methods*, *21*(4), 493. doi: 10.1037/met0000105

Koslowski, N., Klein, K., Arnold, K., Koesters, M., Schützwohl, M., Salize, H. J., & Puschner, B. (2016). Effectiveness of interventions for adults with mild to moderate intellectual disabilities and mental health problems: systematic review and meta-analysis. *The British Journal of Psychiatry, 209*(6), 469-474. doi:10.1192/bjp.bp.114.162313

Kostov, C. E., Rees, C. E., Gormley, G. J., & Monrouxe, L. V. (2018). ‘I did try and point out about his dignity’: a qualitative narrative study of patients and carers’ experiences and expectations of junior doctors. *BMJ Open*, *8*(1), e017738. Retrieved from https://bmjopen.bmj.com

Kour, A. (2015). Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Indispensable Elements to Support Decision-Making Process in Industrial World. *International Journal of Scientific and Research Publications*, *5*(1), 1-7. doi: 10.1.1.735.3953

Krogh, K., Bearman, M., & Nestel, D. (2015). Expert practice of video-assisted debriefing: an Australian qualitative study. Clinical Simulation in Nursing, 11(3), 180-187. doi: 10.1016/j.ecns.2015.01.003

Kucukusta, D., Law, R., Besbes, A., & Legohérel, P. (2015). Re-examining perceived usefulness and ease of use in online booking: The case of Hong Kong online users. *International Journal of Contemporary Hospitality Management*, *27*(2), 185-198. Retrieved from https://www.emeraldinsight.com/journal/ijchm

Lamb, D. (2013a). Promoting the case for using a research journal to document and reflect on the research experience. *The Electronic Journal of Business Research Methods, 11*(2), 84–91. Retrieved from http://www.ejbrm.com/.

Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the Hadoop ecosystem. *Journal of Big Data*, *2*(1), 24. doi: 10.1186/s40537-015-0032-1

Lee, D. Y., & Lehto, M. R. (2013). User acceptance of YouTube for procedural learning: An extension of the Technology Acceptance Model. *Computers & Education, 61*, 193 – 208. doi:10.1016/j.compedu.2012.10.001

Lee, H., Shao, B., & Kang, U. (2015). Fast graph mining with hbase. *Information Sciences, 315*, 56-66. doi: 10.1016/j.ins.2015.04.016

Lehrer, C., Wieneke, A., vom Brocke, J., Jung, R., & Seidel, S. (2018). How big data analytics enables service innovation: materiality, affordance, and the individualization of service. *Journal of Management Information Systems*, *35*(2), 424-460. doi: 10.1080/07421222.2018.1451953

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. Journal of Family Medicine and Primary Care, 4(3), 324–327. doi:10.4103/2249-4863.161306

Lewis, S. (2015). Qualitative inquiry and research design: Choosing among five approaches. *Health promotion practice*, *16*(4), 473-475. doi: 10.1177/1524839915580941

Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers*, *17*(2), 243-259. doi: doi:10.1007/s10796-014-9492-7

Li, X., Shang, W., Wang, S., & Ma, J. (2015). A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data. *Electronic Commerce Research and Applications*, *14*(2), 112-125.

Loudcher, S., Jakawat, W., Morales, E. P. S., & Favre, C. (2015). Combining OLAP and information networks for bibliographic data analysis: a survey. *Scientometrics*, *103*(2), 471-487. Retrieved from https://link.springer.com/journal/11192

Loui, R. P. (2016). From Berman and Hafner’s teleological context to Baude and Sachs’ interpretive defaults: an ontological challenge for the next decades of AI and Law. Artificial Intelligence and Law, 24(4), 371-385. doi:10.1007/s10506-016-9186-1

Lucas, T. W., Kelton, W. D., Sanchez, P. J., Sanchez, S. M., & Anderson, B. L. (2015). Changing the paradigm: Simulation, now a method of first resort. *Naval Research Logistics (NRL)*, *62*(4), 293-303.

Lyon, L., & Brenner, A. (2015). Bridging the data talent gap: Positioning the iSchool as an agent for change. International journal of digital curation, 10(1). Doi: 10.2218/ijdc.v10i1.349

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, *51*, 47-60. doi: 10.1016/j.future.2014.10.029

MacLennan, E., & Belle, J. P. (2014). Service-oriented architecture (SOA) takes an architectural approach to designing and implementing IT solutions. *Information Systems and e-Business Management*, *12*(1), 101-137. doi:10.1007/s10257-012-0212-x

Magdaleno, A. M., de Oliveira Barros, M., Werner, C. M. L., de Araujo, R. M., & Batista, C. F. A. (2015). Collaboration optimization in software process composition. *Journal of Systems and Software*, *103*, 452–466. doi:10.1016/j.jss.2014.11.036

Maher, C., Hadfield, M., Hutchings, M., & de Eyto, A. (2018). Ensuring rigor in qualitative data analysis: A design research Approach to coding Combining NVivo With traditional material methods. *International Journal of Qualitative Methods*, *17*(1). Retrieved from http://journals.sagepub.com/

Mahrt, M., & Scharkow, M. (2013). The Value of Big Data in Digital Media Research. *Journal of Broadcasting & Electronic Media*, *57*(1), 20-33. doi:10.1080/08838151.2012.761700

Malterud, K., Siersma, V. D., & Guassora, A. D. (2015). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 1–8. doi:10.1177/1049732315617444

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81-95. doi: 10.1007/s10209-014-0348-1

Marion, T. J., Eddleston, K. A., Friar, J. H., & Deeds, D. (2015). The evolution of interorganizational relationships in emerging ventures: An ethnographic study within the new product development process. *Journal of business Venturing*, *30*(1), 167-184. doi: 10.1016/j.jbusvent.2014.07.003

Martinez, Y., Cachero, C., & Melia, S. (2013). MDD vs. traditional software development: A practitioner's subjective perspective. *Information and Software Technology, 55*(2), 189 – 200. doi: 10.1016/j.infsof.2012.07.004

Martínez-Mesa, J., González-Chica, D. A., Duquia, R. P., Bonamigo, R. R., & Bastos, J. L. (2016). Sampling: how to select participants in my research study?. *Anais brasileiros de dermatologia*, *91*(3), 326-330. doi: 10.1590/abd1806-4841.20165254

Mayoh, J., & Onwuegbuzie, A. J. (2015). Toward a conceptualization of mixed methods phenomenological research. *Journal of mixed methods research*, *9*(1), 91-107. Retrieved from http://journals.sagepub.com/

McCusker, K., & Gunaydin, S. (2015). Research using qualitative, quantitative or mixed methods and choice based on the research. *Perfusion*, *30*(7), 537-542. doi:10.1177/0267659114559116

McIntosh, M. J., & Morse, J. M. (2015). Situating and constructing diversity in semi-structured interviews. Global Qualitative Nursing Research, 2, 1–12. doi:10.1177/2333393615597674

Miki, A. W., Kagiri, A., & Nganga, K. (2017). Factors influencing sustainability of international livestock research institute technology supported livestock projects in Kisumu County. *International Academic Journal of Information Sciences and Project Management*, *2*(1), 66-85. doi:10.24940/ijird/2017/v6/i4/apr17076

Miller, P. (2008b). Validity. In L. Given (Ed.), Encyclopedia of qualitative methods (p. 909). Thousand Oaks, CA: SAGE.

Miyan, M. U., Nuruzzaman, M., & Sagar, M. S. J. (2016). Systematic and Effective Change in HR Behaviour and Competencies Lead to Organizational Development. *Global Journal of Management And Business Research*. Retrieved from https://journalofbusiness.org

Mohanty, S., Jagadeesh, M., & Srivatsa, H. (2013). "Big Data" in the Enterprise. In *Big Data Imperatives* (pp. 1-24). Apress.

Morse, J. M. (2015a). "Data were saturated". *Qualitative Health Research*, *25*(5), 587-588. doi:10.1177/1049732315576699

Morse, J. M. (2015b). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, *25*(9), 1212-1222. doi:10.1177/1049732315588501

Moylan, C. A., Derr, A. S., & Lindhorst, T. (2015). Increasingly mobile: How new technologies can enhance qualitative research. *Qualitative Social Work*, *14*(1), 36–47. doi:10.1177/1473325013516988

Munshi, A. A., & Mohamed, Y. A. R. I. (2018). Data lake lambda architecture for smart grids big data analytics. *IEEE Access*, *6*, 40463-40471. doi:

1109/ACCESS.2018.2858256

Murrell, M. (2017). Out of print: the orphans of mass digitization. Current Anthropology, 58(S15), S149-S159. Retrieved from https://www.journals.uchicago.edu/toc/ca/current

Nadal, S., Herrero, V., Romero, O., Abelló, A., Franch, X., Vansummeren, S., & Valerio, D. (2017). A software reference architecture for semantic-aware Big Data systems. *Information and software technology*, *90*, 75-92. Retrieved from https://www.journals.elsevier.com/information-and-software-technology

Naheb, O. A., Sukoharsono, E. G., & Baridwan, Z. (2017). The influence of critical factors on the behavior intention to computerized accounting system (CAS) in cement manufactures in Libya. *The International Journal of Accounting and Business Society*, *25*(1), 38-60. Retrieved from https://www.neliti.com/journals/international-journal-of-accounting-and-business-society

Nasri, W., & Charfeddine, L. (2012). Factors affecting the adoption of Internet banking in Tunisia: An integration theory of acceptance model and theory of planned behavior. *The Journal of High Technology Management Research*, *23*(1), 1-14. doi: 10.1016/j.hitech.2012.03.001

Nathali Silva, B., Khan, M., & Han, K. (2017). Big data analytics embedded smart city architecture for performance enhancement through real-time data processing and decision-making. *Wireless Communications and Mobile Computing*, *2017*. doi: 10.1155/2017/9429676

Neale, J. (2016). Iterative categorization (IC): a systematic technique for analysing qualitative data. *Addiction, 111*(6), 1096-1106. doi:10.1111/add.13314

Nicod, E., & Kanavos, P. (2016). Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: a mixed methods study. Health Policy, 120(1), 35-45. doi: 10.1016/j.healthpol.2015.11.007

Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. Evidence-Based Nursing, ebnurs-2015. Doi: 10.1136/eb-2015-102054

Ohno-Machado, L., Sansone, S. A., Alter, G., Fore, I., Grethe, J., Xu, H., ... & Soysal, E. (2017). Finding useful data across multiple biomedical data repositories using DataMed. *Nature genetics*, *49*(6), 816. doi: 10.1038/ng.3864

O'Keeffe, J., Buytaert, W., Mijic, A., Brozović, N., & Sinha, R. (2016). The use of semi-structured interviews for the characterization of farmer irrigation practices. Hydrology and Earth System Sciences, 20(5), 1911-1924. doi: 10.5194/hess-20-1911-2016

O'leary, D. E. (2013). Artificial Intelligence and Big Data. *IEEE Intelligent Systems, 28*(2), 96-99. doi:10.1109/mis.2013.39

Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing Big Data with augmented and virtual reality: challenges and research agenda. Journal of Big Data, 2(1), 22. doi: /10.1186/s40537-015-0031-2

Oltmann, S. M. (2016). Qualitative interviews: A methodological discussion of the interviewer and respondent contexts. *In Forum: Qualitative Social Research, 17(*2). Retrieved from https://uknowledge.uky.edu/slis_facpub/32

Ooi, K. B., Lee, V. H., Tan, G. W. H., Hew, T. S., & Hew, J. J. (2018). Cloud computing in manufacturing: The next industrial revolution in Malaysia?. *Expert Systems with Applications*, *93*, 376-394. doi: 10.1016/j.eswa.2017.10.009

Ovčjak, B., Heričko, M., & Polančič, G. (2015). Factors impacting the acceptance of mobile data services–A systematic literature review. *Computers in Human Behavior*, *53*, 24-47. doi: 10.1016/j.chb.2015.06.013

Özköse, H., Arı, E. S., & Gencer, C. (2015). Yesterday, today and tomorrow of big data. *Procedia-Social and Behavioral Sciences, 195,* 1042-1050. doi.10.1016/j.sbspro.2015.06.147

Padgavankar, M. H., & Gupta, S. R. (2014). Big data storage and challenges. *International Journal of Computer Science and Information Technologies*, *5*(2), 2218-2223. Retrieved from http://www.ijcsit.com

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Administration and Policy in Mental Health and Mental Health Services Research, 42(5), 533-544. doi: 10.1007/s10488-013-0528-y

Panackal, J. J., & Pillai, A. S. (2015). Adaptive utility-based anonymization model: Performance evaluation on big data sets. *Procedia Computer Science*, *50*, 347-352. doi: 10.1016/j.procs.2015.04.037

Park, E., Baek, S., Ohm, J., & Chang, H. J. (2014). Determinants of player acceptance of mobile social network games: An application of extended technology acceptance model. *Telematics and Informatics*, *31*(1), 3-15. doi: 10.1016/j.tele.2013.07.001

Park, E., Kim, H., & Ohm, J. Y. (2015). Understanding driver adoption of car navigation systems using the extended technology acceptance model. *Behaviour & Information Technology, 34*(7), 741-751. doi:10.1080/0144929X.2014.963672

Park, N., Rhoads, M., Hou, J., & Lee, K. M. (2014). Understanding the acceptance of teleconferencing systems among employees: An extension of the technology acceptance model. *Computers in Human Behavior, 39,* 118 – 127. doi: 10.1016/j.chb.2014.05.048

Parker, M., & Kingori, P. (2016). Good and bad research collaborations: Researchers' views on science and ethics in global health research. *PloS ONE 11*(10) .doi: 10.1371/journal.pone.0163579

Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. PloS one, 10(10), e0137041. doi: 10.1371/journal.pone.0137041

Peek, S. T., Wouters, E. J., van Hoof, J., Luijkx, K. G., Boeije, H. R., & Vrijhoef, H. J. (2014). Factors influencing acceptance of technology for aging in place: a systematic review. *International journal of medical informatics, 83*(4), 235-248. doi: 10.1016/j.ijmedinf.2014.01.004

Percy, W. H., Kostere, K., & Kostere, S. (2015). Generic qualitative research in psychology. The Qualitative Report, 20(2), 76. Retrieved from http://nsuworks.nova.edu/tqr/vol20/iss2/

Peticca-Harris, A., deGama, N., & Elias, S. R. S. T. A. (2016). A dynamic process model for finding informants and gaining access in qualitative research. *Organizational Research Methods, 19*(3), 376–401. doi:10.1177/1094428116629218

Pettit, M. (2016). Historical time in the age of big data: Cultural psychology, historical change, and the Google Books Ngram Viewer. History of psychology, 19(2), 141. doi: 10.1037/hop0000023

Plamondon, K. M., Bottorff, J. L., & Cole, D. C. (2015). Analyzing data generated through deliberative dialogue: bringing Knowledge Translation into qualitative analysis. *Qualitative health research*, *25*(11), 1529-1539. Retrieved from http://journals.sagepub.com/

Polit, D. F., & Beck, C. T. (2012). Data collection in qualitative research. Nursing Research: Generating and Assessing Evidence for Nursing Practice. 9th ed. Philadelphia: Wolters Kluwer Health| Lippincott Williams & Wilkins, 542-543.

Power, D. J. (2016). Data science: supporting decision-making. Journal of Decision systems, 25(4), 345-356. doi: 10.1080/12460125.2016.1171610

Prescott, M. E. (2016). Big Data: Innovation and Competitive Advantage in an Information Media Analytics Company. *Journal of Innovation Management*, *4*(1), 92. Retrieved from https://journals.fe.up.pt/index.php/IJMAI/index

Prinsloo, P., Archer, E., Barnes, G., Chetty, Y., & Van Zyl, D. (2015). Big (ger) data as better data in open distance learning. The International Review of Research in Open and Distributed Learning, 16(1). doi: 10.19173/irrodl.v16i1.1948.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51-59. doi:10.1089/big.2013.1508

Pulgar-Rubio, F., Rivera-Rivas, A. J., Pérez-Godoy, M. D., González, P., Carmona, C. J., & del Jesus, M. J. (2017). MEFASD-BD: Multi-objective evolutionary fuzzy algorithm for subgroup discovery in big data environments-A MapReduce solution. *Knowledge-Based Systems, 117,* 70-78. Doi: 10.1016/j.knosys.2016.08.021

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems, 2*(1). doi:10.1186/2047-2501-2-3

Rahi, S. (2017). Research design and methods: A systematic review of research paradigms, sampling issues and instruments development. *International Journal of Economics & Management Sciences*, *6*(2), 1-5. doi: 10.4172/2162-6359.100040

Rahim, M. M. (2008). Identifying factors affecting acceptance of e-procurement systems: An initial qualitative study at an Australian City Council. *Communications of the IBIMA*, *3*(1), 7-17. Retrieved from http://ibimapublishing.com/journals/

Ramírez-Correa, P., Mariano-Melo, A., & Alfaro-Pérez, J. (2019). Predicting and Explaining the Acceptance of Social Video Platforms for Learning: The Case of Brazilian YouTube Users. *Sustainability*, *11*(24), 7115. doi: /10.3390/su11247115

Rana, N. P., Dwivedi, Y. K., & Williams, M. D. (2013). E-government adoption research: An analysis of the employee's perspective. *International Journal of Business Information Systems, 14*(4), 414–428. doi:10.1504/ijbis.2013.057497

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. Computer

Networks, 101, 63-80. Doi: 10.1016/j.comnet.2015.12.023

Renu, B., Ashish, A., & Nitin, G. (2013). Three-tier architecture of DW. International Journal of Latest Technology in Engineering, Management & Applied Science. *2*(5). Retrieved from http://www.ijltemas.in/

Rimando, M., Brace, A., Namageyo-Funa, A., Parr, T. L., Sealy, D. A., Davis, T. L., ... & Christiana, R. W. (2015). Data collection challenges and recommendations for early career researchers. The Qualitative Report, 20(12), 2025. Retrieved from http://nsuworks.nova.edu/tqr/vol20/iss12/8

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). New York, NY: Free Press.

Rolbiecki, A., Subramanian, R., Crenshaw, B., Albright, D. L., Perreault, M., & Mehr, D. (2017). A qualitative exploration of resilience among patients living with chronic pain. *Traumatology, 23*(1), 89. doi: 10.1037/trm0000095

Rondan-Cataluña, F. J., Arenas-Gaitán, J., & Ramírez-Correa, P. E. (2015). A comparison of the different versions of popular technology acceptance models: A non-linear perspective. *Kybernetes*, *44*(5), 788-805. doi: 10.1108/K-09-2014-0184

Rosenthal, M. (2016). Qualitative research methods: Why, when, and how to conduct interviews and focus groups in pharmacy research. Currents in Pharmacy Teaching and Learning, 8(4), 509-516. doi: 10.1016/j.cptl.2016.03.021

Roulston, K., & Shelton, S. A. (2015). Reconceptualizing bias in teaching qualitative research methods. *Qualitative Inquiry, 21*(4), 332-342. Retrieved from http://journals.sagepub.com/

Sahoo, S. S., Wei, A., Valdez, J., Wang, L., Zonjy, B., Tatsuoka, C., ... & Lhatoo, S. D.

(2016). NeuroPigPen: A Scalable Toolkit for Processing Electrophysiological Signal Data in Neuroscience Applications Using Apache Pig. *Frontiers in neuroinformatics, 10,* 18. doi: 10.3389/fninf.2016.00018

Sale, H. B., Patil, D., Thube, S., & Student, B. E. (2018). Crime Prevention with DW using OLAP through Business Intelligence. *International Journal of Engineering Science, 16017*. Retrieved from https://www.journals.elsevier.com/international-journal-of-engineering-science

Salinas, S. O., & Lemus, A. C. (2017). DW and big data integration. *Int. Journal of Comp. Sci. and Inf. Tech*, *9*(2), 1-17. Retrieved from http://airccse.org/journal/ijcsit.html

Sanjari, M., Bahramnezhad, F., Fomani, F. K., Shoghi, M., & Ali Cheraghi, M. (2014). Ethical challenges of researchers in qualitative studies: The necessity to develop a specific guideline. *Journal of Medical Ethics & History of Medicine*, *7*(14), 1 – 6. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4263394/

Santiago-Delefosse, M., Gavin, A., Bruchez, C., Roux, P., & Stephen, S. L. (2016). Quality of qualitative research in the health sciences: Analysis of the common criteria present in 58 assessment guidelines by expert users. Social Science & Medicine, 148, 142-151.

Santos-Vijande, M. L., López-Sánchez, J. Á., & Pascual-Fernandez, P. (2018). Co-creation with clients of hotel services: the moderating role of top management support. *Current Issues in Tourism*, *21*(3), 301-327. doi: 10.1080/13683500.2015.1078781

Sarma, S. K. (2015). Qualitative research: Examining the misconceptions. *South Asian Journal of Management, 22*(3), 176. Retrieved from http://sajm-amdisa.org/

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: Current state and future potential. Journal of Business Logistics, 36(1), 120-132.

Sebaa, A., Chick, F., Nouicer, A., & Tari, A. (2017). Research in big data warehousing using Hadoop. *Journal of Information Systems Engineering & Management*, *2*(2), 10. doi: 10.20897/jisem.201710

Seitz, S. (2016). Pixilated partnerships, overcoming obstacles in qualitative interviews via Skype: A research note. Qualitative Research, 16(2), 229-235 Retrieved from http://journals.sagepub.com/.

Seng, J. K. P., & Ang, K. L. M. (2019). Multimodal Emotion and Sentiment Modeling From Unstructured Big Data: Challenges, Architecture, & Techniques. *IEEE Access*, *7*, 90982-90998. Doi: 1109/ACCESS.2019.2926751

Shafiee, M. E., Barker, Z., & Rasekh, A. (2018). Enhancing water system models by integrating big data. *Sustainable Cities and Society*, *37*, 485-491. Retrieved from https://www.journals.elsevier.com/sustainable-cities-and-society/

Shariati, M. H., & Seyyedrezaei, S. H. (2015). The Effect of the Use of Demonstration and Pictures on Children’s Vocabulary Learning and Retention. *International Journal of Basic Sciences & Applied Research*, *4*, 14-18. Retrieved from http://www.isicenter.org

Shepherd, A., Kesa, C., Cooper, J., Onema, J., & Kovacs, P. (2018). Opportunities and challenges associated with implementing Data Lakes for enterprise decision-making. *Issues in Information Systems*, *19*(1). Retrieved from www.iacis.org/iis

Siddique, S. N., & Ganguly, S. K. (2019). Critical Success Factors for Manufacturing Industries in India: A Case Study Analysis. *International Journal of Applied Engineering Research*, *14*(8), 1898-1905. Retrieved from http://www.ripublication.com

Singh, J., & Singla, V. (2015). Big Data: Tools and Technologies in Big Data. *International Journal of Computer Applications*, *112*(15). Retrieved from http://www.ijcaonline.org/

Smith, B., & McGannon, K. R. (2018). Developing rigor in qualitative research: Problems and opportunities within sport and exercise psychology. *International review of sport and exercise psychology*, *11*(1), 101-121. doi: 10.1080/1750984X.2017.1317357

Smith, C. A., & Fogarty, S. (2015). A survey of study participants' understanding of informed consent to participate in a randomised controlled trial of acupuncture. *BMC complementary and alternative medicine*, *16*(1), 10. doi: 10.1186/s12906-015-0975-y

Snelson, C. L. (2016). Qualitative and Mixed Methods Social Media Research: A Review of the Literature. *International Journal Of Qualitative Methods*, *15*(1), 1-15. doi:10.1177/1609406915624574

Sohn, B. B., Greenberg, K. H., Thomas, S. P., & Pollio, H. R. (2017). Hearing the voices of students and teachers: A phenomenological approach to educational research. *Qualitative Research in Education* (2014-6418), *6*(2), 121-148. doi:10.17583/qre.2017.2374

Sommerhoff, D., Szameitat, A., Vogel, F., Chernikova, O., Loderer, K., & Fischer, F. (2018). What Do We Teach When We Teach the Learning Sciences? A Document Analysis of 75 Graduate Programs. Journal of the Learning Sciences. Doi: 10.1080/10508406.2018.1440353

Sorsa, M. A., Kiikkala, I., & Astedt-Kurki, P. (2015). Bracketing as a skill in conducting unstructured qualitative interviews. *Nurse researcher 22*(4). doi: 10.7748/nr.22.4.8.el1317

Stefanowski, J., Krawiec, K., & Wrembel, R. (2017). Exploring complex and big data. *International Journal of Applied Mathematics and Computer Science*, *27*(4), 669-679. doi: //10.1515/amcs-2017-0046

Storey, V. C., & Song, I. Y. (2017). Big data technologies and Management: What conceptual modeling can do. Data & Knowledge Engineering, 108, 50-67. Doi: 10.1016/j.datak.2017.01.001

Sumbal, M. S., Tsui, E., & See-To, E. W. (2017). Interrelationship between big data and knowledge management: an exploratory study in the oil and gas sector. *Journal of Knowledge Management*, *21*(1), 180-196. doi:10.1108/jkm-07-2016-0262

Sun, S., Cegielski, C. G., Jia, L., & Hall, D. J. (2018). Understanding the factors affecting the organizational adoption of big data. *Journal of Computer Information Systems*, *58*(3), 193-203. doi:10.1080/08874417.2016.1222891

Sun, S., Hall, D. J., & Cegielski, C. G. (2019). Organizational intention to adopt big data in the B2B context: An integrated view. *Industrial Marketing Management*. doi: 10.1016/j.indmarman.2019.09.003

Sun, Z., Strang, K., & Firmin, S. (2017). Business analytics-based enterprise information systems. *The Journal of Computer Information Systems, 57*(2), 169-178. doi:10.1080/08874417.2016.1183977

Sutton, J., & Austin, Z. (2015). Qualitative research: Data collection, analysis, and management. *Canadian Journal of Hospital Pharmacy, 68*(3). doi:10.4212/cjhp.v68i3.1456

Svendsen, G. B., Johnsen, J. A. K., Almas-Sorensen, L., & Vitterso, J. (2013). Personality and technology acceptance: The influence of personality factors on the core constructs of the technology acceptance model. *Behaviour & Information Technology, 32*(4), 323–334. doi:10.1080/0144929X.2011.553740

Tan, G. W. H., Ooi, K. B., Leong, L. Y., & Lin, B. (2014). Predicting the drivers of behavioral intention to use mobile learning: A hybrid SEM-Neural Networks approach. *Computers in Human Behavior*, *36*, 198-213. doi:10.1016/j.chb.2014.03.052

Tarhini, A., Hone, K., & Liu, X. (2014). Measuring the moderating effect of gender and age on e-learning acceptance in England: A structural equation modeling approach for an extended technology acceptance model. *Journal of Educational Computing Research, 51*(2), 163-184. doi:10.2190/EC.51.2.b

Tobin, M., Nugroho, D., & Lietz, P. (2016). Large-scale assessments of students' learning and education policy: Synthesising evidence across world regions. *Research Papers in Education*, *31*(5), 578-594. doi:10.1080/02671522.2016.1225353

tom-Dieck, M. C., & Jung, T. (2018). A theoretical model of mobile augmented reality acceptance in urban heritage tourism. *Current Issues in Tourism*, *21*(2), 154-174. doi: 10.1080/13683500.2015.1070801

Tom-Dieck, M. C., & Jung, T. (2018). A theoretical model of mobile augmented reality acceptance in urban heritage tourism. *Current Issues in Tourism*, *21*(2), 154-174. doi: 10.1080/13683500.2015.1070801

Tong, A., & Dew, M. A. (2016). Qualitative research in transplantation. *Transplantation*, *100*(4), 710-712. doi:10.1097/tp.0000000000001117

Tsang, E. W. (2014). Case studies and generalization in information systems research: A critical realist perspective. *The Journal of Strategic Information Systems*, *23*(2), 174-186. doi: 10.1016/j.jsis.2013.09.002

United States Department of Health and Human Services. (1979). *The Belmont report*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation, 84*, 21-36. doi: 10.1016/j.technovation.2018.07.004

Ur-Rehman, M. H., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. International Journal of Information Management, 36(6), 917-928. doi: 10.1016/j.ijinfomgt.2016.05.013

Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice*, *6*(5), 100-110. doi:10.5430/jnep.v6n5p100

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies, *Management Science*, *46* (2), 186 204, doi:10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478. Retrieved from https://misq.org/

Verma, S., Bhattacharyya, S. S., & Kumar, S. (2018). An extension of the technology acceptance model in the big data analytics system implementation environment. *Information Processing & Management*, *54*(5), 791-806. doi: 10.1016/j.ipm.2018.01.004

Vicary, S., Young, A., & Hicks, S. (2016). A reflective journal as learning process and contribution to quality and validity in interpretative phenomenological analysis. *Qualitative Social Work*. doi:10.1177/1473325016635244

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626-639. doi: 10.1016/j.ejor.2017.02.023

Vroman, K. G., Arthanat, S., & Lysack, C. (2015). “Who over 65 is online?” Older adults’ dispositions toward information communication technology. *Computers in Human Behavior*, *43*, 156-166. doi:10.1016/j.chb.2014.10.018

Wahyudi, A., Kuk, G., & Janssen, M. (2018). A process pattern model for tackling and improving big data quality. *Information Systems Frontiers, 20*(3), 457-469. doi: 10.1007/s10796-017-9822-7

Wallace, L. G., & Sheetz, S. D. (2014). The adoption of software measures: A technology acceptance model (TAM) perspective. *Information & Management*, *51*(2), 249 – 259. doi:10.1016/j.im.2013.12.003

Waller, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. Journal of Business Logistics, 34(2), 77-84. doi:10.1111/jbl.12010

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How ‘big data’can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165*, 234-246. doi: 10.1016/j.ijpe.2014.12.031

Wang, C. C., & Geale, S. K. (2015). The power of story: Narrative inquiry as a methodology in nursing research. *International Journal of Nursing Sciences, 2(*2), 195-198. doi: 10.1016/j.ijnss.2015.04.014

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics, 176*, 98-110. doi:10.1016/j.ijpe.2016.03.014

Wang, S., Li, J., & Zhao, D. (2017). Understanding the intention to use medical big data processing technique from the perspective of medical data analyst. *Information Discovery and Delivery*, *45*(4), 194-201. doi: 10.1108/IDD-03-2017-0017

Wiesenberg, M., Zerfass, A., & Moreno, A. (2017). Big data and automation in strategic communication. *International journal of strategic communication*, *11*(2), 95-114. doi: 10.1080/1553118X.2017.1285770

Wijndaele, K., Westgate, K., Stephens, S. K., Blair, S. N., Bull, F. C., Chastin, S. F., ... & Granat, M. H. (2015). Utilization and harmonization of adult accelerometry data: review and expert consensus. *Medicine and science in sports and exercise*, *47*(10), 2129. doi: 10.1249/MSS.0000000000000661

Williamson, K. M., & Muckle, J. (2018). Students' perception of technology use in nursing education. *CIN: Computers, Informatics, Nursing*, *36*(2), 70-76. Retrieved from https://journals.lww.com/cinjournal/pages/default.aspx

Wimmer, H., & Aasheim, C. (2019). Examining Factors that Influence Intent to Adopt Data Science. *Journal of Computer Information Systems*, *59*(1), 43-51. doi: 10.1080/08874417.2017.1295790

Win, N. W., & Thein, T. (2015). An efficient big data analytics platform for mobile devices. *International Journal of Computer Science and Information Security*, *13*(9), 1. doi: https://sites.google.com/site/ijcsis/

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming–a review. *Agricultural Systems*, *153*, 69-80. doi: 10.1016/j.agsy.2017.01.023

Wong, H. S. P., & Salahuddin, S. (2015). Memory leads the way to better computing. *Nature nanotechnology*, *10*(3), 191. doi: doi:10.1038/nnano.2015.29

Yan, J., Meng, Y., Lu, L., & Li, L. (2017). Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance.

IEEE Access, *5*, 23484-23491. doi: 1109/ACCESS.2017.2765544

Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences*, *112*(47), 14473-14478. doi: 10.1073/pnas.1515373112

Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *The Qualitative Report*, *20*(2), 134-152. Retrieved from http://nsuworks.nova.edu/tqr/vol20/iss2/12

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, *48*(2), 311 – 325. doi:10.1111/ejed.12014

Yoon, C. (2018). Extending the TAM for Green IT: A normative perspective. *Computers in Human Behavior*, *83*, 129-139. doi: 10.1016/j.chb.2018.01.032

Yousafzai, S. Y., Foxall, G. R., & Pallister, J. G. (2007). Technology acceptance: a meta-analysis of the TAM: Part 1. *Journal of Modelling in Management, 2*(3), 251-280. doi: 10.1108/17465660710834453

Yu, D., Dou, W., Zhu, Z., & Wang, J. (2015). Materialized view selection based on adaptive genetic algorithm and its implementation with Apache hive. *International Journal of Computational Intelligence Systems*, *8*(6), 1091-1102. Retrieved from https://www.atlantis-press.com/journals/ijcis

Yurovich, D. P. (2015). A quantitative examination of employees’ perceptions of adopting a learning management system (Doctoral dissertation, Capella University).

Zadvinskis, I. M., Smith, J. G., & Yen, P. Y. (2018). Nurses' experience with health information technology: Longitudinal qualitative study. *JMIR medical informatics, 6(*2) Retrieved from https://www.ncbi.nlm.nih.gov

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... & Ghodsi, A. (2016). Apache spark: a unified engine for big data processing. *Communications of the ACM*, *59*(11), 56-65. doi: 10.1145/2934664

Zakir, J., Seymour, T., & Berg, K. (2015). Big data analytics. *Issues in Information Systems*, *16*(2). Retrieved from http://www.iacis.org/iis/iis.php

Zhang, W., Yan, T. T., Barriball, K. L., While, A. E., & Liu, X. H. (2015). Post-traumatic growth in mothers of children with autism: A phenomenological study. *Autism*, *19*(1), 29-37. Retrieved from http://journals.sagepub.com/

Zheng, Y., Zhao, K., & Stylianou, A. (2013). The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation. *Decision Support Systems*, *56*, 513-524. doi: 10.1016/j.dss.2012.11.008

Zhou, T., & Li, H. (2014). Understanding mobile SNS continuance usage in China from the perspectives of social influence and privacy concern. *Computers in Human Behavior*, *37*, 283-289. doi:10.1016/j.chb.2014.05.008

Zolnikov, T. R., & Blodgett Salafia, E. (2016). Improved relationships in eastern Kenya from water interventions and access to water. Health Psychology, 35(3), 273.doi: 10.1037/hea0000301

Appendix A: National Institute of Health Office of Extramural Research

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Mbah Fortem** successfully completed the NIH Web-based training course “Protecting Human Research Participants”.

Date of completion: 01/11/2016

Certification Number: 1946885

.

Appendix B: Interview Protocol

Interviewee (Title): ____

Interviewer: ________________Mbah Johnas Fortem_________________

Background:

_____ A: Interviewee Background

_____ B: Demographics

Other Topics Discussed: ____

____

Documents Obtained: ____

____

Post Interview Comments or Leads:

____

Introductory Protocol

To facilitate our note-taking and collaboration, I would like to audio tape the entire interview conversations today and perhaps during any follow-up interviews. For your information, only researchers involved in the project will have access to the tapes which will be eventually be destroyed after they are transcribed. This document Essentially states that: (1) all information regardless will be held confidential, (2) your participation is voluntary and you may stop at will at any time if you so desire without explanation, and (3) I do not intend to inflict any harm during this interview. Thank you for accepting willingly to participate. I have allocated a 45 minutes time slot for this interview. During this time, There are series of questions I would like to discuss with you.

Introduction

You have been identified as someone who has the experience and knowledge to share about big data and DW. My research project focuses on big data technologies and the topic is “What strategies do data managers use to transition to big data technologies from DW technologies. My study is not intended to evaluate your knowledge, techniques or experiences. Rather, I am trying to learn more about your strategies using big data technologies.

A. Interviewee Background

How long have you been in your current position at this organization?

B. Demographics

What role and position do you currently hold at your organization?

Post Interview Comments and/or Observations:

Appendix C: Interview Questions

Demographic Questions

1. What role and position do you currently hold at your organization?
2. How long have you been in your current position?
3. How many years of experience do you have in data management and analytics?
4. What types of diplomas, degrees or industry certifications do you possess?

Interview Questions

8. How would you describe big data and how would you compare big data to relational database technologies like DW?
9. What are the tools and methods your team uses to for big data initiatives and How would you describe the usefulness of those tools and methods?
10. What challenges have you had if any, with relational database management systems like DW in your big data initiatives?
11. What strategies have you used to transition from data warehousing to big data technologies?
12. Has your business analytics methodology and tools changed as a result to the transition to big data technologies? .
13. Why did you consider making changes to your data management and analytics strategies?
14. How has the transition to big data technologies affected the overall performance of your team in performing their day to day tasks?

Appendix D: Interview Question Matrix

Interview Questions	**Usefulness**	**Ease of use**	**Intention of use**	**Job Relevance**	**Result Demonstrability**	**Output Quality**	**Voluntariness**	**Social Influences**	**Usage Behaviors**
How would you describe big data and how would you compare big data to relational database technologies like DW?	X			X					X
What are the tools and methods your team uses for big data initiatives and how would you describe the usefulness of those tools and methods?	X	X	X	X	X	X	X		X
What challenges have you had, if any with RDMS like DW?	X			X	X	X	X		X
What strategies have you used to transition to big data technologies from DW technology?	X		X	X	X	X		X	X
Has your business analytics methodology and tools changed as a result of the transition to big data?	X	X		X	X	X		X	X
Why did you consider making changes to your data management and analytics strategies	X	X	X	X	X	X	X	X	X
Has the transition to big data affected the overall performance of your team in performing their day-to-day task?	X	X	X	X	X	X	X	XX	

Note. RDMS = relational database management systems
DW