Spring 6-30-2022

Empirical assessment of big data technology adoption factors for organizations with data storage systems

Ahmad B. Alnafoosi
DePaul University, Ahmad@Alnafoosi.net

Empirical Assessment of Big Data Technology Adoption Factors for Organizations with Data Storage Systems

by

Ahmad B. Alnafoosi

A DISSERTATION SUBMITTED TO THE SCHOOL OF COMPUTING,

COLLEGE OF COMPUTING AND DIGITAL MEDIA OF

DEPAUL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

Dissertation Committee

Theresa Steinbach, Ph.D., Chair
Raffaella Settimi-Woods, Ph.D.
Olayele Adelakun, Ph.D.
Ioan Raicu, Ph.D.
Zarreen Farooqi, Ph.D.
David Rudden, MA

DePaul University

Chicago, Illinois

2022

Acknowledgments

To God almighty, who created all possible.
To my grandparents, who showed me what is possible.
To my parents, who showed me how it is possible.
To my wife, who helped me to make it possible.
To my children, who is why it was possible.
To America and Iraq, the lands of great possibles.

EMPIRICAL ASSESSMENT OF BIG DATA TECHNOLOGY

ADOPTION FACTORS FOR ORGANIZATIONS

WITH DATA STORAGE SYSTEMS

ABSTRACT

Many organizations have on-premises data storage systems. Data storage systems are evolving in multiple ways. One way is the adoption of Big Data. Big Data is a data storage system with the ability to analyze large volumes, velocity, and a variety of data. Per the Economist, data is now the most valuable resource (Parkins, 2017). Big Data holds the promise of unlocking a substantial value of data stored. Yet many organizations are not implementing Big Data. There is a need to identify key factors affecting adoption for such organizations. The literature review revealed multiple gaps in studied adoption factors (un-studied or under-studied) such as data storage latency, ability to compute, data storage interface compatibilities, open-source software, enterprise sourced software, cost, perceived industry pressure, legislation barriers, and market turbulence. These factors are studied in this research using The Diffusion of Innovation (DOI) theory and Technology-Organization-Environment (TOE) framework with qualitative (semi-structured interviews, Interpretive Phenomenological Analysis (IPA), and structured interviews) and quantitative (survey) methods. Quantitative analysis is based on Partial Least Squares – Structural Equation Model (PLS-SEM) analysis. This analysis revealed that six of the nine studied factors are significant. Industry pressure, enterprise-sourced software, storage interface compatibility, market turbulence, open-source software, and cost are significant factors positively correlated to Big Data adoption.

Keywords

Big Data, Data Storage, Diffusion of Innovation (DOI), Technology-Organization-Environment Framework (TOE), antecedent adoption factors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Data Storage to Big Data

Data storage systems have changed from wall paintings, cuneiform writing, 3.5-inch diskette, solid-state drive, and recently a cloud service. These innovations were answers to new challenges and requirements of the environment. Some data storage innovations were adopted, whereas others were not. One of the major innovations that can extend data storage capabilities is Big Data technology. This research attempts to find significant factors that affect Big Data adoption by organizations.

Big Data allows the processing of large volume, heterogeneous variety, and high velocity of data that, distinguishes it from other analytics technologies (McAfee & Brynjolfsson, 2012). These 3Vs (Volume, Velocity, and Variety) are one of the most cited definitions of Big Data. Big Data capability opens new horizons of broader and deeper insights than ever before. Big Data allows queries that process vast amounts and varied categories of data at fast rates that other systems cannot handle in a timely manner.

Big Data capabilities contrast with traditional IT technologies and tools that are not able to acquire, manage and process those attributes of data (volume, variety, and velocity) in tolerable time (M. Chen, Mao, & Liu, 2014). Big Data can also

further extend the value of the data to discover patterns and uncover answers. For example, Big Data can enable Machine Learning and Artificial Intelligence (H. Chen, Chiang, & Storey, 2012). Big Data (BD), Big Data Service (BDS), Big Data Technology (BDT), or Big Data Analytics (BDA) are used in literature to refer to this technology interchangeably. This research will refer to this technology as Big Data.

Many organizations have data storage systems that store and retrieve their data with varying technical specifications. Data storage systems and the data they hold are essential parts required for Big Data technology. Yet not all organizations with data storage systems and data stored in them are able to take the next step toward Big Data (Ajimoko, 2017; H.-M. Chen, Kazman, & Matthes, 2015; Dubey, Gunasekaran, Childe, Wamba, & Papadopoulos, 2016). This research intends to find undiscovered significant factors that affect the adoption decision of Big Data technologies for organizations with data storage systems.

In this research context, the data storage system is defined as any storage system (no specification on medium, capacity, encoding, interface, or locality including the cloud). What is excluded from this research is purchasing the Big Data as a feature, which is not a significant challenge. For example, organizations with cloud storage systems can purchase Big Data as a feature from the cloud provider. Implementing Big Data on data storage systems (including the cloud) can be a significant challenge which is the focus of this research.

Although Big Data is at the near peak of hype and inflated expectations, Big Data adoption is still in its early stages (Hall, 2013). Most organizations have not adopted Big Data in production. Around 13% of companies put Big Data into production use, and Big Data projects had failure rates of 55% (H.-M. Chen et al., 2015). This paradox of high expectations and excitement for Big Data and its low adoption needs further study. There is a need to identify the key determinants affecting adoptions and identify multiple contexts (de Camargo Fiorini, Seles, Jabbour, Mariano, & de Sousa Jabbour, 2018).

This research will explore innovation theories and frameworks. It will be grounded in the Diffusion of Innovation (DOI) theory (Everett M Rogers, 2003) and the Technology-Organization-Environment (TOE) Framework (DePietro, 1990). DOI provides an in-depth analysis of innovation, emphasizing technology adoption factors. In comparison, TOE offers a broad context framework of innovation adoption factors. This combination of DOI theory and TOE framework addresses Big Data adoption factors from a technical, organizational, and environmental perspective in organizations with data storage systems.

This research will develop a collection of Big Data adoption factors in technology, organization, and environment based on gaps from prior research. The research will use a mixed-method approach by combining a qualitative semi-structured interview method with a quantitative survey method. The first phase is qualitative, using interviews of data storage researchers and practitioners who have

implemented Big Data on data storage systems to explore and refine adoption factors. The validated factors from the interviews will be tested using surveys to validate the hypotheses. The second phase is a pilot survey to validate the survey and get a narrow sample of responses from a small set of participants. The third phase is a general survey to test the hypotheses.

This research intends to find some of the significant factors affecting Big Data adoption in organizations with data storage systems. In addition, it plans to determine if these factors are enabling or impeding Big Data adoption for these organizations. Identifying significant adoption factors and their association with Big Data adoption will assist data storage practitioners and researchers in ascertaining challenges, opportunities, solutions, and strategies that address these factors. This can enable more organizations with data storage systems to adopt Big Data, thus increasing its diffusion.

1.2 Motivation for the study

Over 9.3 Zettabytes of data were stored in 2016 (Westervelt, 2017). The estimated growth rate in 2011 was 23%. However, some industries discard 80-90% of their data (Martin Hilbert, 2016). It is estimated that only 3% of the data stored is analyzed (David Reinsel, Gantz, & Rydning, 2017). There is substantial potential for analyzing a larger percentage of the data. Traditional analysis tools are not able

to handle these quantities and varieties of data in a reasonable time. Increasing adoption of Big Data technology can increase the amount of data analyzed.

Big Data technology offers the ability to analyze these large volumes, variety, and velocity. Even if the data is residing on multiple data servers or locations. The economic benefits of using Big Data were estimated to be in the hundreds of billions of dollars in 2011 (M. Chen et al., 2014). Big Data holds the promise of unlocking substantial value (Hirsch, 2014). Lately, per the Economist article, data is now the most valuable resource (Parkins, 2017). This research intends to find the significant factors that enable or hinder Big Data adoption for organizations with data storage systems. Thus, contributing to unlocking the stored data value with Big Data.

Many data analysis tools have limitations on the amount of data or type of data they can process. For example, Excel (as with many other spreadsheet tools) expects data to be in columns and row format (structured data). They expect the data to be in text format to be analyzed (not pictures, videos, or multidimensional matrices). There is a limit to the number of columns and rows that can fit. The processing usually happens on a single machine or a limited number of devices that needs to process the stored data in the memory of that machine.

Conversely, Big Data can process not only large amounts; but also varied types of data that can reside on multiple machines. Big Data can process structured (ex. Spreadsheets and database tables), semi-structured (ex. XML and JSON), and

unstructured (ex. text, pictures, and videos). The amount of data to be processed is not limited by a single machine but by the storage capacity of multiple machines where the data reside. The processing capacity is not limited by a single machine, but it can be spread across multiple machines. Thus, Big Data enables the processing of large volumes and varying types of data at a velocity that is not matched by other technologies.

Big Data has and is expected to expand its impact on all aspects of our lives. Big Data has added value to data by discovering patterns, predicting outcomes, and correlating factors (Martin Hilbert, 2016; Siegel, 2016). Big Data has allowed the data to "speak." Big Data enables data to speak and present data in substantially different ways than before. It can, in some instances, offers the entire population instead of a sample. It can present data with all the real-world anomalies.

It presents data with the ability to search and correlate this massive, heterogeneous, and rapid data within a reasonable time. The possibilities of other applications are still open for this technology. The range of Big Data impact is broad and expanding to all aspects of our lives in healthcare, transportation, communication, governance, and astronomy, to name a few aspects mentioned in "Big Data: A revolution that will transform how we live, work, and think" (Mayer-Schönberger & Cukier, 2017).

Data storage systems are evolving (Yianilos & Sobti, 2001). Big Data is an innovation that presents new possibilities for data utilization. One way to abstract

Big Data is data, storage, management, and analysis (or described as compute in some literature) (M. Chen et al., 2014; Nguyen & Petersen, 2017). Another abstraction is network, data storage systems (centralized or decentralized), and the ability to compute capacity (Martin Hilbert, 2016). Data storage systems have multiple components (including CPU, memory, network, and interface, to name some). To access the data, one must traverse the technical capabilities of the storage system. Thus, storage systems with data have several major components of a Big Data system. That does not include the processing ability and the technology for data management and analysis. Every storage system with data is a potential implementation for Big Data. Yet not every storage system is a Big Data system (Ajimoko, 2017; M. Chen et al., 2014; Dubey et al., 2016). Moreover, a minority of organizations have adopted Big Data in production, despite the Big Data applications and expected potential.

1.3 Contribution to the discipline

The main goal of the research is to assist researchers and practitioners on data storage systems to empirically determine the significant factors that are enabling or inhibiting Big Data technology adoption. These discovered factors should assist researchers and practitioners in providing more targeted solutions that address these factors. Thus, enabling more storage technologies to adopt Big Data.

Consequently, supporting data storage and larger data sets to realize the benefits of Big Data like analytics, machine learning, artificial intelligence, and other technologies.

Machine learning can extract even more value from the data by analyzing data sets (and very large data sets with Big Data). Artificial intelligence can add a value by utilizing the data sets available. These technologies enable data analysis, identification of solutions to problems, and implementation of solutions with human intervention or without. These analyses include classification, clustering, association, and network analysis. One distinguishing trait that distinguishes machine learning and artificial intelligence from regular analytics is that human design is not required to provide a specific solution or an algorithm to a problem. These technologies offer solutions based on the data they analyze. Human intervention is needed in delivering learning and intelligence algorithms for these technologies to examine, learn and decide. Big Data provides larger data sets and thus more potential to increase the breadth and depth of insight into data (H. Chen et al., 2012).

Prior research addressed the challenge of Big Data adoption and diffusion with multiple theories and frameworks. Various factors have been studied that affect the adoption of Big Data positively or negatively. There are numerous gaps in the evaluated factors. Some factors are not studied; while others are understudied, or too broad. Un-studied factors, such as organizational governance

of technology sources (e.g. open versus enterprise software) is a determining factor in adoption (Felin & Zenger, 2014). Understudied factors, such as perceived industry pressure, appeared only once (Nam, Kang, & Kim, 2015). Other factors are too broad to provide specific solutions like compatibility (K. Agrawal, 2015).

This research will test the significance of the proposed adoption factors. Cost, perceived industry pressure, regulations, and storage interface compatibility adoption factors have been under-studied or too broad in previous research. Thus, this research intends to further previous research and provide a more granular understanding. Data storage latency, ability to compute, interface compatibility, open-source, and enterprise-sourced software are adoption factors that have not been studied before. This research will contribute to understanding the significance of these unstudied adoption factors to Big Data. In addition, this research will also examine whether these factors positively or negatively affect the Big Data adoption decision.

Executives voted Big Data as "the most hated buzzword of 2013" (Datoo, 2014). It has also been the new oil with its negative connotations of breaches and contaminations (Hirsch, 2014). Yet in 2015, Gartner's research indicated that 75% of companies are investing or planning to invest in Big Data (Van der Meulen & Woods, 2015). With this direction toward investment, there has been a spur in research in Big Data. Big Data research started with few publications in peer

reviewed publications in 2009-2011. Then it reached over 100 by 2014 see Figure 1 (Frizzo-Barker, Chow-White, Mozafari, & Ha, 2016).

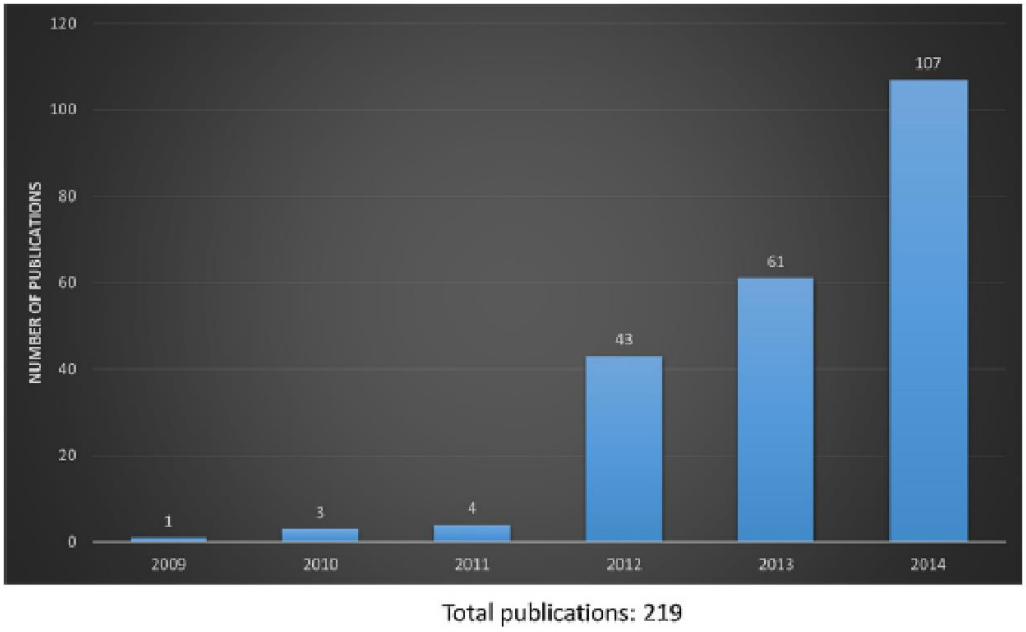

Figure 1 The Number of Publications on Big Data Per Year (Frizzo-Barker et al., 2016).

28% of Big Data research is empirical versus conceptual research 72% versus (Frizzo-Barker et al., 2016). Within the empirical research of Big Data research few has been in factors affecting Big Data adoption (Nguyen & Petersen, 2017). Only 11% of the Big Data research is mixed methods. (Frizzo-Barker et al., 2016). This research is designed to contribute to the Big Data research in the segment of empirical, mixed method research to find significant factors that are affecting the Big Data adoption.

1.4 Research Agenda

Big Data technology presents great potential for unlocking what will transform many aspects of our lives (Hirsch, 2014). Organizations with data storage systems have data. Enabling organizations with data storage systems to implement Big Data solutions is the goal of this research. Big Data adoption antecedent factors have been studied earlier from a variety of perspectives. This research focuses on studying the Big Data adoption factors that are specific organizations with data storage that have not been studied or under-studied. Thus, expanding the potential for Big Data adoption to every organization with a data storage system.

1.4.1 Problem Statement

The value of Big Data has been demonstrated in many organizations for both public and private sectors (Romijn, 2014). Big Data has become an important factor in growth and competition for many organizations (McGuire, Manyika, & Chui, 2012). Tapping into the large volume, heterogenous variety and the velocity of what being stored, because of human and instrument activities, requires Big Data technology.

Big Data is characterized as having the ability to ingest data at high velocity, store large volumes of accumulated data and can compute these large sets of data

(Martin Hilbert, 2016). Barker et al. have described Big Data as data collection, storage and processing (Frizzo-Barker et al., 2016). Most organizations (69%) store their data locally in data storage systems versus cloud (Maddox, 2015). In other words, many organizations have the data and the storage system components of Big Data systems. This presents a great opportunity for organizations with data storage systems to adopt Big Data and reap its benefits.

Yet not all organizations are able to adopt Big Data for various challenges. Big Data adoption is around 20% in some sectors (Columbus, 2017) while the potential has been illustrated to be substantial and hype of Big Data is almost at its peak (Nguyen & Petersen, 2017). This disparity can be understood within DOI and TOE frameworks. Many Big Data adoption factors have been studied in an unbalanced manner that focused mainly on organizational and environmental factors as illustrated in the literature review chapter. Novel Big Data adoption factors can be explored using the DOI and TOE framework. These novel factors will expand the current understanding of what factors that can advance or hinder Big Data adoption. One can ascertain which of these factors are significant. Factors can be tested whether they affect Big Data adoption positively or negatively.

This research can be the input for further enhancement or mitigation plans to assist in Big Data adoption. This research will empirically explore adoption factors of Big Data and identify their significance for organizations. This study will cover technical, organizational, and environmental factors. The research is intended to

find these significant factors to enable organizations, practitioners, and academics to provide more tailored solutions that address these factors. This will hopefully enable further diffusion of Big Data technologies for organizations with data storage systems.

1.4.2 Scope of the Research

This research focuses on organizations with data storage systems that will implement and adopt Big Data. This excludes organizations that will buy Big Data as a feature from their data storage system provider. Out of scope example is an organization with data storage on cloud that purchases big Data as a feature. In scope example is an organization with data storage system (local or cloud) that implement Big Data solution on top of its own data storage system. Thus, this research will not explore organization with their data in the cloud providers' Big Data adoption factors. Since many cloud offerings (Amazon, Google, IBM, Oracle, and others as of 2020) offer Big Data as a feature that can be added to the cloud deployment.

This research studies selected technological, organizational, and environmental factors based on current research literature review. This research is a mixed method of qualitative and quantitative methods. The selected set of factors is selected based on a review of the literature. Then, these factors are refined by conducting semi-structured interviews of data storage and Big Data practitioners and researchers

who implemented Big Data. The semi-structured interviews findings are then validated using a pilot survey then a large-scale survey. This will capture a snapshot selection of practitioners and academics in organizations with data storage technologies. The survey will test the significance of Big Data adoption factors identified above.

1.4.3 Research Questions

To tackle the lingering Big Data adoption issue in organizations with data storage system, we propose the following research questions:

1) What are the significant technical factors that affect the Big Data adoption?
2) What are the significant organizational factors that affect the Big Data adoption?
3) What are the significant environmental factors that affect the Big Data adoption?

1.4.4 Research Model

Big Data is near its peak of the hype cycle (Hall, 2013), yet the adoption of Big Data is significantly lagging (H.-M. Chen et al., 2015; Columbus, 2017). To get a better understanding of this mismatch between the expectations and reality, this research investigated adoption theories and frameworks. There are multiple

adoption theories and frameworks with varying goals and scopes. This research has surveyed and evaluated multiple adoption theories and frameworks. The best fit based on organizational adoption unit, technological factors and wider scope factors were Diffusion of Innovation (DOI) theory (Everett M Rogers, 2003) and Technology-Organization-Environment (TOE) framework (DePietro, 1990) as a combination. DOI theory provides in depth understanding of adoption theory with explanatory capability from multiple aspects of innovation adoption. TOE provides a wider breadth framework that covers organizational and environmental factors in addition to the technical factors.

Based on the TOE framework we have surveyed Big Data adoption specific research in technical, organizational, and environmental perspectives. There are 11 technical, 40 organizational and 28 environmental adoption factors studied from previous research. This survey helped identify the disparity of limited number and scope of technical factors compared to organizational and environmental factors. More in depth exploration of the literature will be explored in the second's chapter.

Thus, based on the surveyed literature, the researcher was able to identify new DOI-TOE factors that were not studied before. This presents an opportunity to add to the knowledge in the Big Data adoption space. New technical factors identified are data storage latency, ability to compute large amounts of data, and network compatibility factors that have not been studied in the Big Data context. These factors give more granularity and depth to the generic compatibility adoption factor

which is one of the most studied factors. Organizational factors had much more breadth in coverage, yet governance of open source versus enterprise sourced has not been covered in the Big Data context. The cost of Big Data implementation has been studied but not specifically for organizations with Data Storage. Perceived industry pressure is one of the environmental adoption factors that has not been studied extensively. Furthermore, there are other environmental factors that have not been studied much previously to strengthen the repeatability of the studied factors. Legislation barriers are a factor that fall into that category.

There are multiple technical factors that affect that ability to adopt Big Data. Technical factor of compatibility has been studied in multiple research papers (K. Agrawal, 2015; H.-M. Chen et al., 2015; Esteves & Curto, 2013; Mahesh, Vijayapala, & Dasanayaka, 2018; Salleh & Janczewski, 2018; Verma & Bhattacharyya, 2017). Yet, there is so much depth and dimensions to compatibilities that can be studied further. Some data storage latency can range from milliseconds to hours. Data storage latency to store and retrieve data is a technical factor that can affect the ability to adopt Big Data, yet this has not been studied. It also presents an opportunity to provide a technical solution to enable Big Data adoption.

The ability to compute large amounts of data is another technical factor that can give further granularity to compatibility. Data locality is one of the features of Big Data. Data locality seeks to collocate ability to compute with the data

(Xiaoqiang, Xiaoyi, Jiangchuan, Hongbo, & Kai, 2017). Yet not all data storage systems may not have the capacity to process it. That can be a prospect to expand solutions that offer compute power to existing data storage systems that need additional compute to enable Big Data.

Data storage interface is another technical factor that can give further detail on compatibility. Storage systems can have multiple interfaces (file system, block, objects, and other variations). This variety can present a compatibility challenge to adopt Big Data technology (Qingchen et al., 2014).

Organizational governance of open sourced versus enterprise systems adoption factors has been studied in general IT technology adoption (Ajila & Wu, 2007; Fan, Stallaert, & Whinston, 2000). Yet, this has not been studied from a Big Data perspective. Many enterprise companies like IBM, Microsoft, and Oracle, to name a few, offer enterprise systems Big Data. Whereas Big Data technology can be found as open source such as Apache Hadoop and Apache Spark. These varying sources of the Big Data technology present an adoption choice that needs to be understood within the realm of Big Data.

Cost of Big Data implementation as an organizational factor has been surprisingly studied only once (Verma & Bhattacharyya, 2017). Accordingly, this study will include the "perceived cost" of implementing Big Data solutions.

There are many adoption factors that are not technology or organizational specific but rather environmental. Many environmental Big Data adoption factors

have been studied. Remarkably, perceived industry pressure was found only once among these studied factors. Perceived industry pressure is the degree that the firm is affected by competitors and partners in the market in their adoption decision making (Nam et al., 2015). Other environmental factors that were studied only once are legislation barriers and market turbulence.

Summary of the research model is presented in Figure 1.

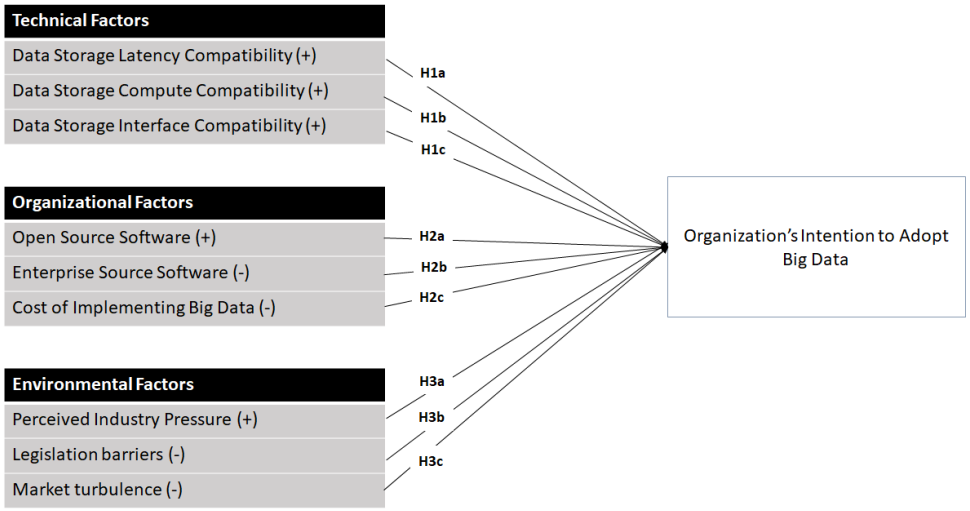

Figure 2 The Research Model

1.4.4 Hypotheses

1.4.4.1 Hypothesis 1

Technological compatibility adoption factors are studied broadly in previous research (K. Agrawal, 2015; H.-M. Chen et al., 2015; Esteves & Curto, 2013; Mahesh et al., 2018; Salleh & Janczewski, 2018; Verma & Bhattacharyya,

2017). Compatibility is defined as "The degree to which innovations are perceived as being consistent with existing methods for executing their mission." The more compatible an innovation is to the existing systems leads to less uncertainty for the potential adopters of the innovation (Everett M Rogers, 2003). However, compatibility has multiple dimensions (M. P. Johnson, 1989) that can be studied. To expand our understanding of compatibility, the more granular dimensions of compatibility will be explored. The following are the novel and more granular Big Data adoption factors and their definitions:

1- Data Storage Latency Compatibility: The degree to which Big Data is perceived as being consistent with existing data storage system latency for executing their mission.
2- Data Storage Compute Capability: The degree to which Big Data is perceived as being consistent with existing storage system ability to compute (Especially large amounts of data) for executing their mission.
3- Data Storage Interface Compatibility: The degree to which Big Data is perceived as being consistent with existing storage system interfaces for executing their mission.

Based on the above discussion, it is hypothesized in this study as follows:

- **H1a: With respect to adopting Big Data in an organizational level, data storage latency compatibility will positively correlate with Big Data adoption decisions.**
- **H1b: With respect to adopting Big Data in an organizational level, data storage compute compatibility will positively correlate with Big Data adoption decisions.**
- **H1c: With respect to adopting Big Data on an organizational level, data storage interface compatibility will positively correlate with Big Data adoption decisions.**

1.4.4.2 Hypothesis 2

This research on the organizational innovation adoption factors covers two main areas. The first area is source of the Big Data software source which determines the support type, confidence, and affordability of the solution. The organization's governance on open sources versus enterprise sourced technology is a new factor that has not been studied in Big Data adoption. Technology source can be a collaborative open-source project that is distributed with open-source type use license. Open-source licenses can offer free usage of the technology, access to the source code and limited or no dedicated support. In contrast enterprise sources offer for fee usage of the technology, custom private code, and dedicated support. Some organizations may prefer open-source software since it offers lower cost. On

other hand, other organizations may elect to have dedicated support that assist in realizing the trialability and full potential of the technology. It will be interesting to determine the profile of the companies to select either or both options as a significant factor of Big Data adoption.

Second area of the Big Data adoption factor is perceived cost of implementing Big Data. In the organizational context within TOE, surprisingly, perceived cost is one of the least studied factors in Big Data adoption. Cost is defined as the expenses of implementing necessary technologies in organizations and efforts devoted to organizational restructuring and process re-engineering (Verma & Bhattacharyya, 2017). Cost of adopting innovation in general and Big Data is an important factor in determining the ability of an organization to adopt Big Data.

Based on the above discussion, it is hypothesized in this study as follows:

- **H2a: With respect to adopting Big Data in an organizational level, open-source software availability of Big Data solutions will positively correlate with Big Data adoption decisions.**
- **H2b: With respect to adopting Big Data in an organizational level, enterprise source software availability of Big Data solutions will negatively correlate with Big Data adoption decisions.**

- **H2c: With respect to adopting Big Data in an organizational level, perceived cost of Big Data will negatively correlate with Big Data adoption decisions.**

1.4.4.3 Hypothesis 3

There are many environmental factors affecting Big Data adoption decisions. The environmental adoption factors are not in direct control of organizations. Yet these adoption factors affect organizations' decisions to adopt new technologies. Perceived industry pressure is an environmental issue that has been studied only once before. It can be defined as "the degree that the firm is affected by competitors and partners in the market" (Nam et al., 2015). The demand for more customized data will grow as personalized solutions become more prevalent in the marketplace from other vendors. Big Data is an enabler of personalization and data customization (Anshari, Almunawar, Lim, & Al-Mudimigh, 2018).

Legislation barriers adoption factor is an environmental factor in the TOE framework. It is defined as government policy that provides inadequate legal protection or business laws (H.-M. Chen et al., 2015). It is one of the least studied Big Data adoption factors. Government legislation can be a barrier or a promoter of change (Gibbs & Kraemer, 2004).

Market turbulence is another environmental factor that affects Big Data adoption. This can be defined as “Changes in customers’ product preferences, demands, and needs in a big data environment” (Sun, Cegielski, Jia, & Hall, 2018).

Based on the above discussion, it is hypothesized in this study as follows:

- **H3a: With respect to adopting Big Data in an organizational level, perceived industry pressure for Big Data will positively correlate with Big Data adoption decisions.**
- **H3b: With respect to adopting Big Data in an organizational level, legislation barriers of Big Data will negatively correlate with Big Data adoption decisions.**
- **H3c: With respect to adopting Big Data in an organizational level, market turbulence of Big Data will negatively correlate with Big Data adoption decisions.**

1.5 Summary

Data Storage Systems are changing and evolving and will continue to do so. One major innovation that offers great potential to extract value from the stored data is Big Data. It can be elucidating innovation that enables the query of large

and heterogeneous data in high velocity. Organizations who have data storage systems and large sets of data have the potential of advancing their knowledge with Big Data technology. This potential of adopting Big Data technology is affected by multiple factors.

Using the TOE framework this research intends to discover unexplored factors that affect Big Data adoption from a data storage perspective. The adoption decision is based on technological, organizational, and environmental factors. In addition to the identification of significant factors, this research intends to find the enabling and inhibiting factors of the Big Data adoptions.

The following is a list of Big Data adoption factors to be studied in this research. Technological Big Data adoption factors are data storage latency compatibility, ability to compute large amounts of data and data storage interface compatibility. Organizational Big Data adoption factors are concerned with governance. They are open-source software, enterprise source software and the cost of implementing Big Data. Environmental Big Data adoption factors are perceived industry pressure, legislation barriers and security, privacy, and ethics. The main aim of this research is to find if these adoption factors are significant. In addition to find if these are enabling or hindering Big Data adoption.

These factors, once studied, can be used by data storage practitioners and researchers in finding solutions to enable wider Big Data technology adoption. Either by using enabler factors to advance Big Data adoption or finding solutions

to overcome the inhibitor factors. Thus, this research aims to contribute to the understanding and development of the adoption of Big Data technology to organizations with data storage systems.

CHAPTER 2: LITERATURE REVIEW

Big Data technology is relatively new technology. In 2005, the Big Data term was coined first by Roger Magoulas. In the same year, Hadoop, a Big Data solution, one of the first Big Data technologies, was created by Yahoo (Melby, 2013). Yet, the term Big Data was used to describe large data sets since the late 1960's (Halevi & Moed, 2012). Interestingly, a NASA paper in 1997 described Big Data as large data sets and heterogeneous among other traits (Cox & Ellsworth, 1997) which comes very close to the current definition. This historical perspective of nomenclature hints at the growing need to manage this "Big Data" that is different from other data sets in terms of size, variety, and the challenge of processing it. Big Data is an innovation that yet to reach the potential that is aspired to. To understand Big Data technology adoption, this research will rely on diffusion theories.

There are multiple innovation theories in the information system that attempt to find the factors that affect an innovation's adoption. Multiple innovation theories, frameworks and models will be explored to test their compatibility with this research. The best theoretical match for the purposes of this research are Diffusion of Innovation (DOI) theory and Technology-Organization-Environment (TOE) framework. DOI theory provides in depth understanding of innovation, adoption and diffusion (Everett M Rogers, 2003). The TOE framework has more

specific focus on adoption factor analysis but expands on DOI adoption factors (DePietro, 1990). There are multiple studies in Big Data adoption factors using DOI, TOE, both or other theories and frameworks.

Diffusion and adoption theories and frameworks are surveyed. These theories and frameworks are explored in the Big Data adoption context. Specifically, previously studied Big Data adoption factors are explored. Gaps in Big Data adoption factors are identified. This research will expand on factors that have not been studied or under-studied and will be focused on data storage systems evolution perspective.

2.1 Evolution of Data Storage Systems and Big Data

The challenge of storing data is seen from clay tablets writings over 5,000 years ago, to modern day storage systems. The challenge to communicate over time by storing data and retrieving it. Many innovations over the ages have attempted to answer this challenge. There is no final solution to this open challenge. Storing, holding and retrieving data are basic functions storage systems can offer (IEEE, 2001). Other innovations have added other features to data storage systems. For example, resilience to failures, searching, adding metadata among,

analysis and many other functionalities are added over the years. Beyond that, there is an innovation in data storage that offers deeper insights at a wider scale that has not been experienced before. The functionality to query and to analyze large swathes of stored data. That innovation is Big Data technology.

The quantity of the stored data has increased not only by human generated content but also by computer generated contents like the Internet of Things (IOT). Currently, the amount of data stored is rapidly increasing due to consumer, business, scientific and government generated content. In 2007, there were over 250 exabytes (exabyte is over a billion gigabytes) of optimally compressed data stored globally (M. Hilbert & López, 2011). In 2016, data stored was estimated at 9.3 zettabytes (zettabyte is over 1,000 exabyte) created, captured and replicated (Sh. Hajirahimova & S. Aliyeva, 2017; Westervelt, 2017). That is an increase of over 3600% in less than a decade. At the organizations' level, the demand for large data storage system has increased as well with current capacity of petabytes and exabytes (Alnafoosi & Steinbach, 2013; M. Hilbert & López, 2011; Ma, Vazhkudai, & Zhang, 2009).

Not only is capacity increasing, the variety of data is also grown. Data density of content has surged over the years from writing, pictures, ultra-high-definition videos, and other large data matrices. Some of the data can fit nicely into tables and databases (structured data), others are less able to (semi structured like XML, YML, JSON, etc.). While other types of data are completely different

(unstructured like standalone files, images, videos, etc.) (Alnafoosi & Steinbach, 2013; H. Chen et al., 2012).

Data at this large scale stored in data storage systems can not fit into tables or databases (In terms of volume, type variety and velocity). The challenge is to query and analyze large data (Terabytes, Petabytes and larger) that may be distributed on multiple machines. Big Data technology enables operators to query and analyze this data within "reasonable time" (M. Chen et al., 2014). Big Data enables the management and coordination of large amounts of data, Queries that can parse through multiple types of data files in multiple data storage devices and locations then to be processed. Other definitions of Big Data will be presented later in this chapter.

Currently, Zettabytes of data is stored, and the rate of storage is accelerating. Data storage systems can offer storage and retrieval of the data. But beyond these basic operations, the potential for broader and in-depth insight to the data is limited by the data ingestion and processing of current analysis systems. Isn't the whole "of the data" more than the sum of its parts as Aristotle posited (Hofstadter, 1979; Jara, Genoud, & Bocchi, 2014)? Big Data technology offers the ability to look at the "whole" data and can even be extended to Machine learning and Artificial Intelligence. Big Data offers values in the hundreds of Billions according to a McKinsey report published in 2011 (M. Chen et al., 2014).

This data in storage systems is waiting to be analyzed. According to the IDC report by 2025 only 3% of the data will be analyzed (David Reinsel et al., 2017). Yet not all storage systems are able to run Big Data technologies. Storage systems have the data and the ability to store it. Big Data offers in addition the ability to manage data and analyze it (compute) (M. Chen et al., 2014). There are factors that are affecting the adoption of Big Data technology. This research attempts to find these factors and their significance.

2.2 What is a Data Storage System?

IEEE standards define storage as "act of storing information." In addition to the storing information, it also needs to hold the information and to allow retrieval of stored information later. That is the basic functionality of a storage system. Storage itself in computer systems can be divided into primary and secondary storage. Primary storage (or main storage) is internal storage on computers. In contrast, secondary storage (mass storage, auxiliary storage, or bulk storage) is defined as "An area of storage, or a storage device, having a very large storage capacity" (IEEE, 2001). This research refers to secondary storage when it uses the term "data storage system".

The evolution of data storage remains in flux at a higher rate of change than ever before. In the last hundred years, more media with bigger storage density were

invented like paper, vinyl records, magnetic tape, magnetic disk, optical disk, solid state drives. Encoding the data varied greatly. Some of the encoding features include analog, digital, error correction, compression to name a few. The interface to the storage varied to deliver audio, video, text, or other media types. Some storage interfaces involved few functions like play, rewind, forward, stop and pause like cassette tapes. While other interfaces like file systems have more sophisticated functions that allow hierarchical directory and file structure with metadata for each element. Yet the life expectancy of the media is decreasing while storage density of characters per square inch is increasing, Figure – 3 (Conway, 1996). Data density trend in HDD is like the general trend, Figure – 4 (Morris & Truskowski, 2003).

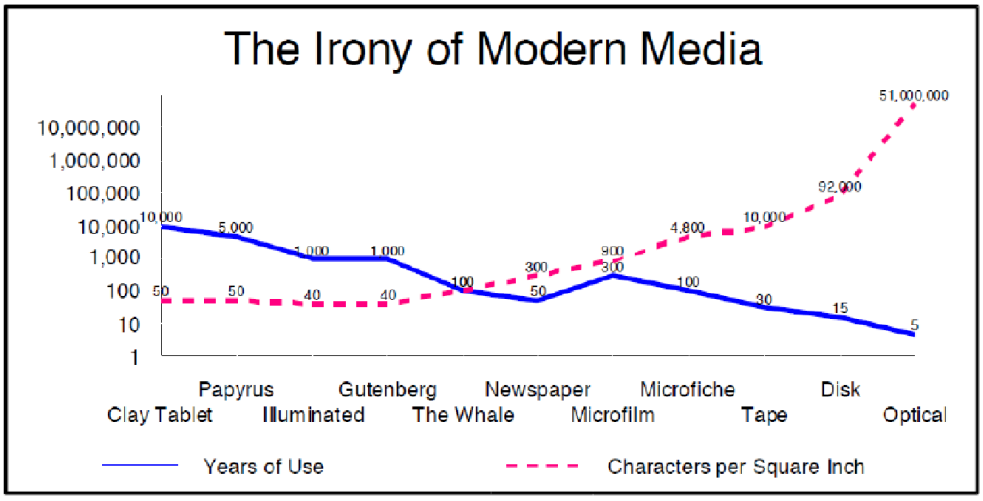

Figure 3 Life expectancy of media in Years through the ages (Conway, 1996).

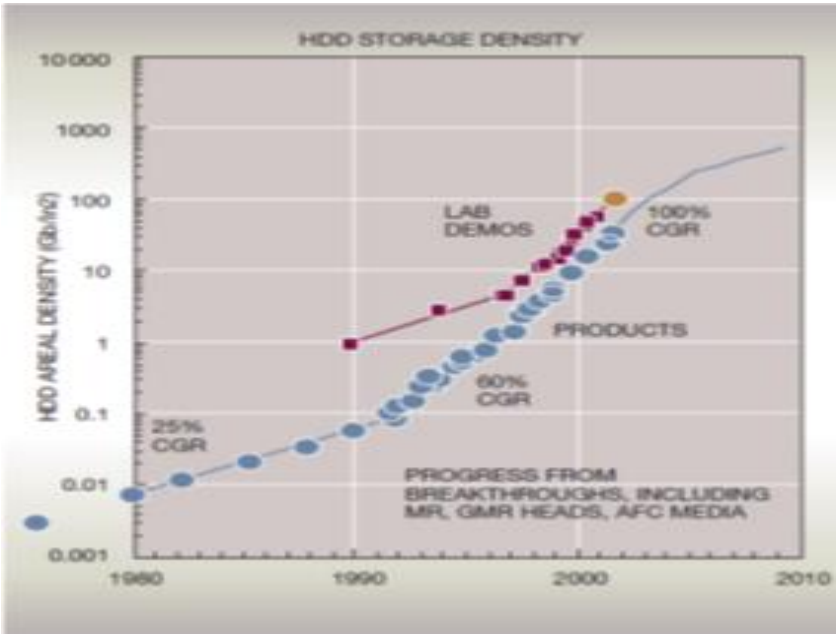

Figure 4 HDD Storage Density Trend from 1980s to 2010 (Morris & Truskowski, 2003).

Data storage systems have many different options for media, storage schemas, interfaces and algorithms and plurality of these and other features combinations to fit varying needs. Storage media can be paper (as in punch cards or more recently in QR code), magnetic tape or disk (Hard Disk Drive), optical (DVD or Blu-ray Disk) or Solid-State Drives (Alnafoosi & Steinbach, 2013).

The data that is stored on data storage systems that have multiple components and each component can be impaired independently due to failure or malicious attack see Figure - 5 (Xin, 2005). This is in addition to inevitable media failure (Xin, 2005). Simple storage of keeping one copy on a media is not enough

to ensure longevity and availability of the data. Multiple layers of technologies are needed to protect against media failures. In the IDC survey in 2018, nearly half of businesses suffered unrecoverable data events in the last 3 years (IDC, 2018).

The media failure in the data storage system is significant. The chance of media going bad is directly correlated to the number of storage components which can be very high for large systems (P. M. Chen, 1993). For example, a storage system with 1 Petabytes and expansion rate of 1.5 will require to have around 1,500 disk drives of 1Terabyte capacity. If there is a failure rate of 1% per year, then there will be around 15 disks failures per year. In other words, with these parameters there is nearly 100% chance of media failure for these large systems.

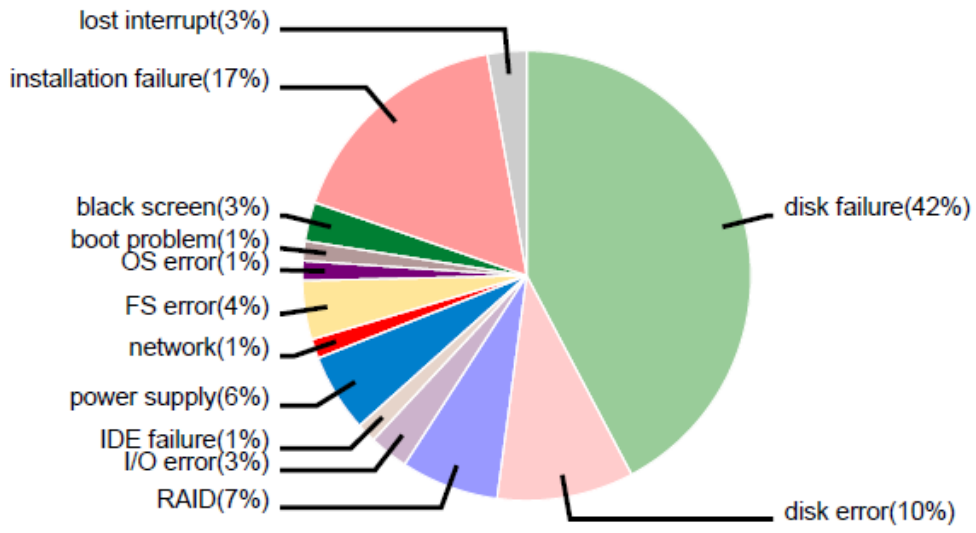

Figure 5 Hardware Failure Distribution at Internet Archive for 30 days (Xin, 2005).

To overcome these failures and keep the data available multiple storage schemes have been developed. To overcome data loss caused by components' failure, the data needs to be expanded. These schemes offer differing degrees of protections with differing degrees of data expansion. These techniques include replication, striping (decimation), splitting and information dispersal algorithms. These storage schemas can work individually or in combination within a storage system. There are multiple variations of each of these storage schemes. Refer to (Alnafoosi & Steinbach, 2013) for more in-depth information and resources.

The expense of storing data is dependent on media, expansion rate and other storage technologies. Where more latent technologies are usually cost less than the less latent technologies. For example, in 2015 the cost of archiving data on magnetic tape cost around $ 0.02 per GB. Whereas Solid State Drive SSD (Uses Flash storage) cost is $0.25 which is 12.5 X compared to magnetic tape storage (Coughlin, 2016). The general trend of cost is decreasing. However, some technologies like tape and HDD still hold cost advantages over other technologies like optical storage and SSD (D. Reinsel, 2013). Morris and Truskowski assert that raw storage price has been decreasing at 50%-60% per year since 1997 (Morris & Truskowski, 2003). See Figure -6

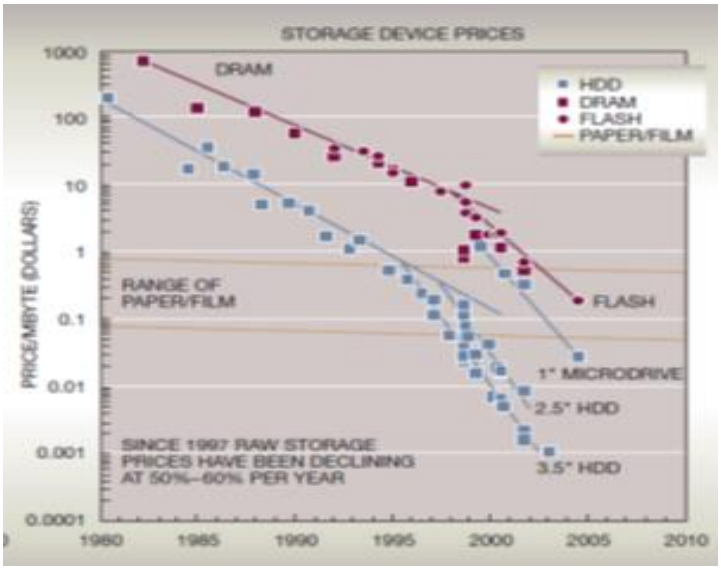

Figure 6 Cost Per MB for Multiple Storage Media from 1980 to 2010 (Morris & Truskowski, 2003).

This differentiation in cost led to tiering data into different classes of storage media and technologies (Chaudhuri & Dayal, 1997). This diversity of media and storage technology addresses the challenge of storing large volumes of data but also presents a challenge on how to deal with multiple storage technologies with differing performance characteristics. See figure 7 on the data storage distribution per media type since the 2000s (D. Reinsel, 2013).

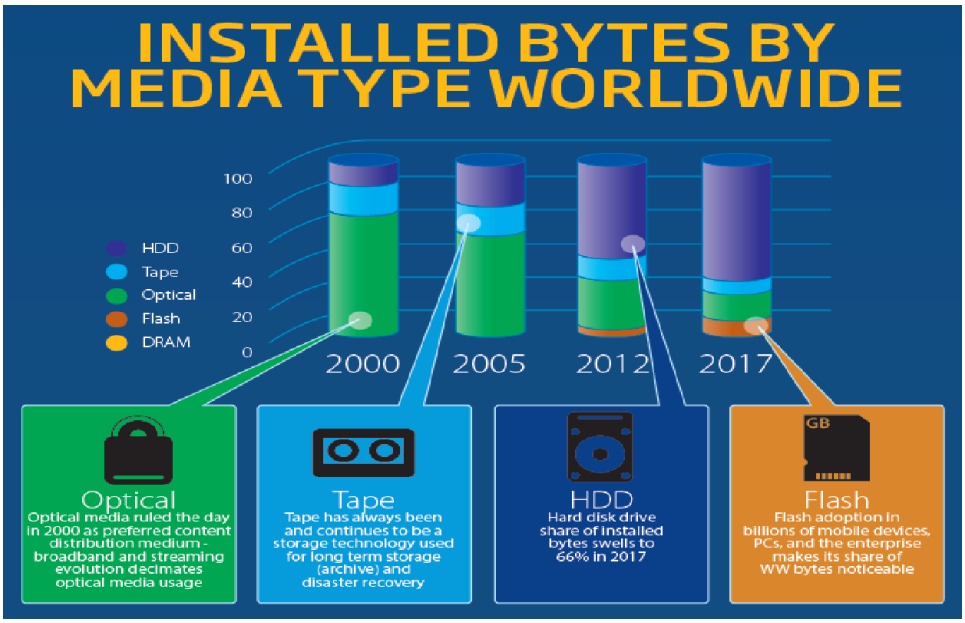

Figure 7 Installed Bytes by Media type Worldwide (D. Reinsel, 2013).

Data is stored on different storage systems. Thus, the data is presented in another set of characteristics. Data storage interface can be a file system or object storage. Within each interface there are multiple varieties with differing protocols of accessing the data. For example, a data interface can be a file system where data is presented in a hierarchy of directories (or folders) and files. There are multiple technologies that support this interface like NAS, SAN, Transfer Protocols. Object storage presents the data in a single layer of folders (other terms used like containers and buckets) and files are presented as objects (or Blobs). There are multiple

interface types for object storage like OpenStack, S3 and others (Alnafoosi & Steinbach, 2013).

There are many ways to classify storage systems. (Offline, near online and online) IOPS, Bandwidth, Media, technology, Interface, security privacy and operational features. Among other classifications. Further research in each classification to determine the significant factors of Big Data adoption at higher granularity might be needed.

This plurality of media, encoding schemas and interfaces produces a variety of storage systems with varying characteristics. Big Data, as mentioned earlier, can be abstracted to the following components: data, data storage system, ability to process the large volume of data and the technology to enable that (Nguyen & Petersen, 2017). Adding Big Data technology to the data storage system presents its sets of technical challenges that are unique to this domain. This research intends to explore technical factors within data storage systems that affect Big Data adoption. These factors will offer insight and feedback into what factors can be addressed in future Big Data research and solutions to enable wider adoption.

2.3 What is Big Data Technology?

"Big Data is the Information asset characterized by such a high Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value" (De Mauro, Greco, & Grimaldi, 2016).

The main features of Big Data are data volume, data velocity and data variety which was defined in 2001 (Laney, 2001; Russom, 2011). This definition is often referred to as 3Vs. It was used for 10 years by many corporations and research entities (M. Chen et al., 2014). Although these are the main features of Big Data, others have identified other features that distinguish Big Data from data storage and analytics technologies (H. Chen et al., 2012).

In 2011, IDC added Value as another V (M. Chen et al., 2014; Gantz & Reinsel, 2011). Value is economically extracted from the other 3Vs. This highlights the underlying reasons for Big Data. Bello-Orgaz et al. added data veracity to that list as well as extracting value from the stored data (Bello-Orgaz, Jung, & Camacho, 2016). Conversely, variability (inconsistency of the Big Data) was added in 2012. Other researchers did not adhere to starting Big Data features with letter V. For example, the complexity of connecting and linking the data are identified as features of Big Data (Hood-Clark, 2016). Others have added visualization making up a 7Vs (Nguyen & Petersen, 2017). Although these are some of the distinguishing features of Big Data. They are also some of its challenges. Data

storage systems in general need to handle the volume, velocity, and variety of data to extract value of data. This research will focus on the first 3Vs of Big Data which are Volume, Velocity and Variety (De Mauro et al., 2016).

Volume and velocity of data being generated have increased significantly (M. Hilbert & López, 2011). Data variety has increased on multiple dimensions to accommodate different needs (Z. H. Liu & Krishnamurthy, 2012). Veracity is the challenge of ensuring the accuracy of the data. The differentiation of data volume, velocity, variety, veracity, and values led to the tiering of data and differentiation of Storage Solution systems. For example, the data storage requirements for on-line transaction processing systems (OLTP) data storage systems are significantly different from on-line analytic processing (OLAP) data storage system requirements in velocity and volume (Chaudhuri & Dayal, 1997).

2.3.1 Data Volume

As more sensors, cameras, X-ray machines, MRIs and other devices record data, the more data that needs to be stored. This is in addition to the human storage needs for text files, pictures, videos, and others. The data created in 2016 is estimated between 9.3 ZBs (Sh. Hajirahimova & S. Aliyeva, 2017; Westervelt, 2017) and 16 ZBs (David Reinsel et al., 2017). This data needs to be stored on actual media. To put it in perspective 16 ZB is 16 *10 ^ 9 Terabyte. If they were to be stored in 1TB disks, this will translate to 16,000,000,000 disks. This number

does not account for expanding the data to protect it from failures and errors. The actual raw data stored with expansion factors (RAID, Erasure Coding, Replication) will be multiplied 3X (or more) so the data can have 99.999 availability or more (X. Li, Lillibridge, & Uysal, 2011). By 2025 the total size of the data stored is expected to be 163 ZB which is 10x the 2016 baseline see Figure 8 (David Reinsel et al., 2017).

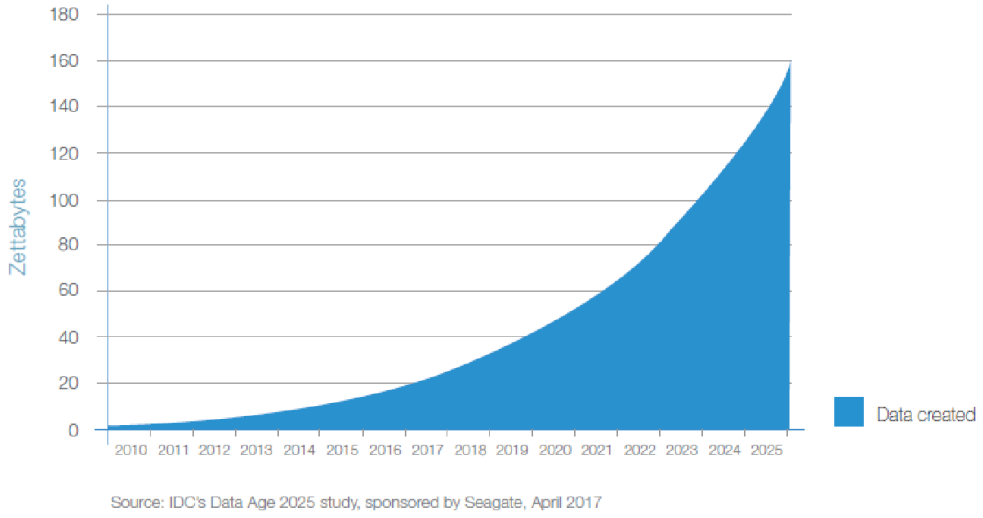

Figure 8 Annual Size of the Global Data sphere (David Reinsel et al., 2017).

2.3.2 Data Velocity

In 2007, there were over 250 exabytes (exabyte is over a billion gigabytes) of optimally compressed data stored globally and data to be stored is growing at a rate of 23 percent (M. Hilbert & López, 2011). Capacity and lifetime of the storage system before it reaches its capacity limit are important variables of consideration.

Similarly, at the organizational level, the demand for big-data storage has been in the capacity of petabytes and exabytes (M. Hilbert & López, 2011; Ma et al., 2009). Facebook, for example, is expected to have ingested 500 TB of data per day in 2012 (Hogan & Shepherd, 2015).

From a media perspective, media capacity is not growing at the rate of data growth. Hard Disk Drive (HDD) and Solid-State Drive (SSD) abilities to grow their capacity density are expected to increase at a rate of 20 percent (Fontana Jr, Decad, & Hetzler, 2012). This rate of capacity density growth breaks Kryder's Law of doubling HDD capacity annually (Walter, 2005) and will pose a challenge to the data growth rate of 23 percent.

Thus, scaling data storage for continual growth is a challenge. It involves the scaling of capacity, performance, throughput, and proximity. Adding additional media devices is not an optimal solution to the problem. For example, using earlier social network 2012 ingest example, transferring 500TB in a day will require a network that is able to handle 500TB/Day of raw data (not expanded) which translates to 56 Gbps of dedicated throughput. Expanding the data to account for error correction and reliability can increase that number significantly. That is not an easy engineering task that will only become more challenging as time passes.

2.3.3 Data Variety

Data types can be classified from multiple perspectives. For example, data differ by source, format, sizes, content, structure, latency to name a few. Sources can vary from human generated content, social media, machine generated data, Internet of Things (IOT) data, transactions data, sensing data (Hood-Clark, 2016).

Data are classified as structured (a relational database), semi-structured (XML) and unstructured (text files, pictures, videos and PDFs) (Z. H. Liu & Krishnamurthy, 2012). Structured data can fit into a table or a database. Semi structured data have semantic tags that have a certain structure but needs some processing and does not fit neatly into structured data format. Unstructured data can have any format that may not fit into the current data classification paradigm. Another data classification is determined by how data is stored. The categories are document oriented, column oriented, graph database information and key value information. These varying data formats require different storage characteristics (Hood-Clark, 2016).

The data can also be classified in terms of its size (KBs, MBs, GBs or larger) (Mesnier & Akers, 2011). Data are classified in terms of its retrieval time/latency (time to first byte, or last byte) (Ganger et al., 2001). Data are classified in terms of retention times (Bsharah & Less, 2000). There are other data classification dimensions as well that are based on the data or the metadata (Kune, Konugurthi, Agarwal, Chillarige, & Buyya, 2016).

2.4 Data Storage System vs. Big Data

Big Data has four main components: information, technology, methods and impact (De Mauro et al., 2016). De Maro et al defined Information as data structured in a way that can be useful for specific purposes. Basically, data that can be accessed in a way that enables Big Data other components to be used. Technology component has hardware and software sub-components. Hardware sub-components are data storage system, computational ability. Software sub-component is a Big Data application (ex. Hadoop, HDFS and MapReduce). The software sub-component provides the ability to query and analyze the data. Methods are defined as procedure and algorithms that can process the large volume of data (ex. Cluster analysis, genetic algorithms, natural language processing, machine learning, neural networks, predictive modeling to name a few). Finally, impact is defined as the pervasive and adaptable nature of Big Data. Big Data gives flexibility to solve increasingly complex problems that covers large data sets (De Mauro et al., 2016; Hood-Clark, 2016).

Information and data storage technology hardware are already present in data storage systems. There are other components of compute and Big Data technologies that enable the processing and management of the data that might be missing. These missing components might be needed to build a Big Data system. That indicates data storage systems have the potential to be Big Data systems. Still

there are other differences between data storage systems and Big Data systems. The following are some of the differences.

Consistency and completeness of the data increase in uncertainty as the data volume and variety increases (Verma & Bhattacharyya, 2017). Unlike databases for example, Big Data does not have a correlation check option for the data ingest to ensure consistency with the stored data. Thus, data preparation might be needed in Big Data to ensure the consistency of the data (Siriweera, Paik, Zhang, Kumara, & Services, 2016).

Others have added variability and complexity to the issue of veracity. Variability is defined as inconsistencies in data. Complexity is the need to connect and correlate the data to produce the desired data (Hood-Clark, 2016). For example, social media and web traffic are affected by spam. Data types are susceptible to noise that affect the identification of actual data (Abbasi, Sarker, & Chiang, 2016). The data noise can originate from internal or external sources. It can be random or deliberate.

Big Data is a subset of Data storage systems. In addition to storing and retrieving the data in accordance with service level agreement (SLA), Big Data storage solutions are also required to provide the ability to manage, analyze, validate, visualize and disseminate the data stored (Kune et al., 2016). This additional functionality may not be available in all storage solutions where analytic

software like Hadoop and SPARK can retrieve and process the data to answer questions that may not be designed into the data to begin with.

Once the Big Data is stored, analyzing these large data sets will be a major challenge. There are two approaches to analytics. The first approach is for data to be moved to where the computing power is located. The second approach is to move the computation to where the data is located. The ability to analyze stored data is an important feature for those interested in extracting more value out of the data stored (Apache, 2019). Big Data Analytics (BDA), or Big Data (BD) in short, refers to tools and methodologies that are used to transform massive quantities of raw data (structured and unstructured) into "data about the data" for analytical purposes (Ochieng, 2015).

2.5 Why Big Data?

More individuals, organizations and governments digitize their data and processes than ever before. This digital transformation has forced more data to be stored and a selected amount to be analyzed. The scale of this digitization is enormous. For instance, in 2017 Gartner research estimated that there are 8.4 billion devices online today in the realm of the Internet of Things (IOT). That is an increase of more than 30% in one year (Newman, 2017). Yet, only 3% of data is estimated to have been analyzed by 2025 (David Reinsel et al., 2017). Since this

data is being produced anyways, that begs the question: How can this data be analyzed, and further value can be extracted?

Big Data is an answer. It can analyze large volume, varying data content types at speeds that current technologies are not able to compete with yet. It is an enabler of many IT innovations and part of the mega trend of the digital information revolution (Nguyen & Petersen, 2017). It facilitates further analysis and dynamic insight into stored data from these activities. Big Data has and has bigger potential to transform the operations of organizations to add further value than just the data (Hood-Clark, 2016).

Big Data impact is significant and wide ranging. In one of the frequently cited papers in Big Data, Martin Hilbert described Big Data's impact "on the social sciences can be compared with the impact of the invention of the telescope for astronomy and the invention of the microscope for biology (providing an unprecedented level of fine-grained detail)." (Martin Hilbert, 2016). Big Data is only at the initial phase of the cycle of information analytics technologies in 2012 (Hall, 2013) see Figure – 8.

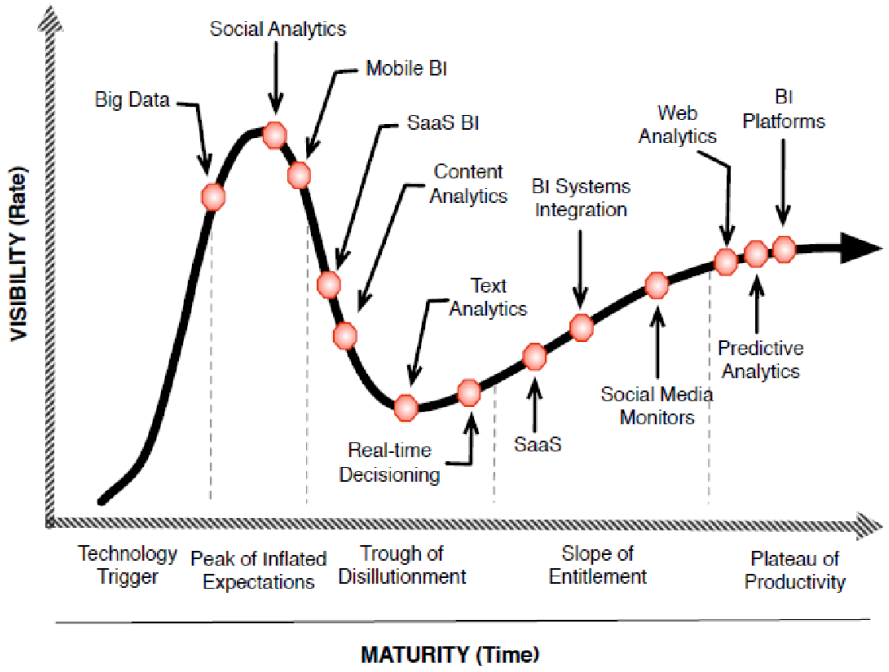

Figure 9 Information Analytics Hype Cycle (Hall, 2013).

Big Data is being used extensively in private organizations specifically in retail, E-commerce and market intelligence such as Walmart, Sears and Amazon (H. Chen et al., 2012). Big Data is also in the financial sector such as Morgan Stanley and online companies such as Google, Facebook, and Twitter. They are using Big Data to gain more comprehensive and in-depth understanding of customers, their choices, personalize marketing, increase satisfaction, increase sales, identify market trends, increase productivity, cut cost, improve decision making among many other goals (H. Chen et al., 2012; Nguyen & Petersen, 2017). A study from

MIT concluded that companies engaged in data-driven decision-making were, on average, 5% more productive and 6% more profitable than their competitors (Brynjolfsson, Hu, & Simester, 2011).

In 2012, the Whitehouse issued a Strategic Plan for Big Data Research and Development (Marzullo, 2016). It states the following the following as part of the objectives of this plan:

"*We envision a Big Data innovation ecosystem in which the ability to analyze, extract information from, and make decisions and discoveries based upon large, diverse, and real-time datasets enables new capabilities for Federal agencies and the Nation at large; accelerates the process of scientific discovery and innovation; leads to new fields of research and new areas of inquiry that would otherwise be impossible; educates the next generation of 21st century scientists and engineers; and promotes new economic growth.*"

Public organizations are also using Big Data to unlock the ability to ask new and existing questions, transform government, empower citizens, increase transparency and to improve their decision-making process. Governments have used Big Data to summarize and understand public comments and concerns. Democratic campaigns have also used Big Data in campaign research, mobilization, donations, opinion mining, social network analysis and social media analytics techniques (H. Chen et al., 2012).

Science and technology sectors have also used Big Data to advance scientific impact and improve safety and security. Big Data is used in The Large Hadron Collider to answer scientific questions. Biological, healthcare and medical fields are using Big Data as well in so many endeavors such as genome, healthcare outcomes and neuroscience (G.-H. Kim, Trimi, & Chung, 2014; Romijn, 2014). Big Data is being used in improving healthcare, outcomes and patient empowerment (H. Chen et al., 2012).

Furthermore, (Tambe, 2014) examined the extent to which early adopters of Big Data technology would have distinct advantages over their competitors. The study demonstrated that firms' investments in such technology, for the period 2006 to 2011, were associated with 3% faster productivity growth. This performance gap is predicted to continue growing as more relevant data are generated. Similarly, the European Commission (2016) predicts that the use of Big Data by the top 100 EU manufacturers could lead to savings worth €425 billion. For the year 2020, employing BIA on Big Data could bring the EU economic growth by an additional 1.9%, equivalent to a GDP increase of €206 billion (Jourová, 2016).

BD extends BI and BA. Types of analysis are descriptive, predictive and prescriptive (Nguyen & Petersen, 2017). Big Data is also referred to as Big Data Analytics is the natural next step of Business Analytics and Business Intelligence from information societies to knowledge societies (Bellinger, Castro, & Mills, 2004; Martin Hilbert, 2016).

Big Data has been compared to oil in multiple aspects. They are similar in providing large amounts of value, needing refinement, providing outputs for other processes, and in causing considerable damage in case of breaches (Hirsch, 2014; Skaale & Rygh, 2018). Big Data can provide further analytics and value in visualization, artificial intelligence, machine learning and data mining. Visualization and tabulation of data is the most basic analytical ability of Big Data. Artificial intelligence (AI) technologies are used to simulate intelligence on problems that were thought impossible to solve without human involvement. Artificial intelligence can be enhanced significantly by providing large data sets. Common types of AI include machine learning and data mining. Machine learning can be defined as algorithms that can perform supervised or unsupervised learning in classification and regression. The learning can be performed using a subset of the data against the full data set. Another machine learning type is regression where the data is analyzed to predict a value and compared to actual values. Data mining is the process of "mining" finding valuable information "ores" from large sets of "dirt" data. These findings can present significant value and can lead to the advantage that other competitors may not have these insights (Skaale & Rygh, 2018).

Machine learning, and data mining were not significantly adopted initially in the 1990s and 2000s. Nevertheless, they have made a major comeback in the last decade. This success has contributed to the availability of large sets of data

especially with Big Data. Big Data is an enabler of machine learning and data mining. These technologies are not guided by specific theory, but they are based on actual data and their correlations (Martin Hilbert, 2016). Big Data provides the ability to examine the data in very large, dynamic, evolving ways that the data were not designed or thought to be used in those ways initially.

Big Data can replace random sampling with access to the whole population data when that is available (Martin Hilbert, 2016). Instead of just analyzing a small percentage of the population, Big Data can enable the analysis of the whole population in some instances (for example, the complete sales data for a particular company). This approach reduces estimation and increases confidence in the data extracted. For example, all active cell phone types of data can be obtained (versus an estimate) by the carrier. This permits the carrier to capture usage and determine rollout of upgrades. Nevertheless, Big Data does not capture all the data on all aspects of reality. It is bound by what is measured, captured, and stored.

Big Data has been used in tracking words, location, nature (sensing), transactions, behaviors, production, and many other facets of data with the dimension of time. All these aspects can be analyzed individually or in any combinations to answer questions and find correlations (Martin Hilbert, 2016). Big Data enables decision making processes to be data driven and applied to a greater extent (H. Chen et al., 2012). The potential of Big Data and what is still being developed and discovered.

2.6 Innovation Adoption, Diffusion, and Frameworks

There are multiple definitions of innovation. An innovation can be defined as any idea, product, program, or technology that is new to the adopting unit (Nguyen & Petersen, 2017; Premkumar & Roberts, 1999). Everett M. Rogers defines innovation as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (Everett M Rogers, 2003). It can also be defined as a "new knowledge applied in some tangible form to achieve human goals" (Gregor & R. Hevner, 2014). Thus, innovation is a novelty (that can be idea, technology, artifact, object, process, etc.) that is adopted by humans (individuals, groups, organizations, countries, etc.) to achieve at least a goal.

Innovation adoption is a decision to implement or use a new idea made by an individual, group of individuals, or authoritative individual. Units of adoption can be individual, groups, organizations, or other larger social units (countries for example). Diffusion includes adoption on a wider scale that has social, communication and time elements. Rogers defined diffusion as

"***The process by which an innovation is communicated through certain channels over time among the members of a social system***" (Everett M Rogers, 2003).

An earlier definition of diffusion cited in Rogers *Diffusion of Innovation* book is from Gabriel Tarde's book *The Laws of Imitation* published in 1903

"***A slow advance in the beginning, followed by rapid and uniformly accelerated progress, followed again by progress that continues to slacken—until it finally stops: These are the three ages of invention. If taken as a guide by the statistician and by the sociologist, [they] would save many illusions.***"

Innovation adoption decision can happen by an individual, individuals representing teams, individuals representing organizations, markets, or societies at large. Ontologically, a unit of adoption can be a person adopting a novel gadget which is also referred to as micro level. Team or group adopting a new technology that is also described as Meso or medium level. Organizations adopting innovation that can be described as Macro level (Z. Liu, Min, & Ji, 2008). Others argued that macro levels refer to the market at large (Verma & Bhattacharyya, 2017). This research will focus on the organizational level unit of adoption as the macro level adoption unit of Big Data technology.

Diffusion of innovation is the change that occurs and is communicated at the social level. Diffusion deals with new innovations/ideas and their uncertainty, unpredictability and how that information is used to decide among alternatives over time (Everett M Rogers, 2003). Human actors as individuals or groups make decisions within organizations to determine the adoption for that organization. They are part of communication channels and social groups. This research will

study data storage practitioners and researchers to determine what are the factors that are affecting their organizations' decision of adopting Big Data technologies.

There are many theories, frameworks and models that describe how innovations are adopted and diffused. Some of them are used separately while others can complement each other. They have been applied in multiple fields, especially, in social sciences. They were also used in Information Systems research (Z. Liu et al., 2008; Everett M Rogers, 2003). Most of their use focuses on micro level adoption units (Over 60%). Meso level adoption research is around 3%. Macro level research is at 35% (Z. Liu et al., 2008). This distribution of research suggests scarcity for Meso, and macro level adoption research compared to micro level. Researchers need to select the appropriate theory, framework or model that works with the intended goal and unit of adoption.

There are many salient research questions. For example, what are the traits of different adaptors? What are the determinant factors that affect the decision of adoption? What stage of adoption is technology at? What is the rate of diffusion and how to control it? (Everett M Rogers, 2003). To answer these questions, a variety of topics are explored. Innovations' adoption, unit of adoption, determining antecedent factors of adoption, rate of adoption, channels of diffusion, classes of adapters and their traits have been studied (Everett M Rogers, 2003). In the antecedent factors that affect adoption, Rogers defined DePietro et al. have studied the social context of innovation and the environmental context of innovation

(DePietro, 1990). Others have studied the process of innovation (Nguyen & Petersen, 2017), intention to use, perceived system use, continuance of system use, actual system use, mix of the previous factors and others (Z. Liu et al., 2008).

A list of theories that are used in the Information Systems field was produced by Liu et al. in 2008. These theories and frameworks have varying research questions and foci. In the literature review phase of this research. Other theories were found in the Big Data and Information Systems technology adoption research. These theories were added to the list and produced the Table 1 below (modified and expanded from (Z. Liu et al., 2008)). This table is used as a starting point to narrow down which theory or framework to be used in this research.

	Theoretical foundation	**Main contributing author(s)**
1	Social Learning Theory/ Social Cognitive Theory (SLT/SCT)	(Miller & Dollard, 1979); (Bandura, 1986)
2	Institutional Theory	(Selznick)
3	Diffusion of Innovation Theory (DOI)	(Everett M. Rogers, 1962)
4	Theory of Reasoned Action (TRA)	(Fishbein & Ajzen, 1975)
5	Triandis' Model	(Triandis, 1980a)
6	Social Influence Theory	(Fulk, 1987)
7	Diffusion/Implementation Model	(T. H. Kwon & Zmud, 1987)
8	Technology Acceptance Model (TAM)	(Fred D. Davis, 1989)
9	User Satisfaction Theory	(Melone, 1990)
10	Technology-Organization-Environment Framework (TOE framework)	(DePietro, 1990)
11	IT Fashion Theory	(Abrahamson, 1991)
12	Theory of Planned Behavior (TPB)	(Ajzen, 1991)
13	Perceived Characteristics of Innovating (PCI)	(Moore & Benbasat, 1991)

14	Motivational Model (MM)	(Fred D Davis, Bagozzi, & Warshaw, 1992)
15	Capability Maturity Model Integration (CMMI)	(Paulk, Curtis, Chrissis, & Weber, 1993)
16	Tri-Core Model	(Swanson, 1994)
17	Task-Technology Fit Model (TTF)	(Goodhue & Thompson, 1995)
18	Usability Studies	(Nielsen, 1999)
19	Extended Technology Acceptance Model (TAM2)	(Venkatesh & Davis, 2000)
20	IS Continuance Model	(Bhattacherjee, 2001)
21	Unified Theory of Acceptance and Use of Technology (UTAUT)	(V. Venkatesh, M. G. Morris, G. B. Davis, & F. D. Davis, 2003)

Table 1 Theories used in IT Adoption Studies (modified and expanded from (Z. Liu et al., 2008)).

Not all these theories, frameworks and models fit the scope and objectives of this research. The scope of this research is to understand the Organizational level adoption decision. Some of these have a micro unit of analysis (individual) which is not the focus of this research. Individual unit of analysis in the following theories exclude them from selection for this research: Technology Acceptance Model (TAM), Theory of Planned Behavior (TPB), Perceived Characteristics of Innovating (PCI), Theory of Reasoned Action (TRA), Extended Technology Acceptance Model (TAM2), Task-Technology Fit Model (TTF), User Satisfaction Theory, IS continuance model, Triandis' Model and Social Influence Theory (Bhattacherjee, 2001; Fulk, 1987; Schmitt, Thiesse, & Fleisch, 2007; Triandis, 1980b). The diffusion/implementation model by Kwon & Zmud has expanded DOI factors to include environmental, organizational, task, individual and innovation

characteristics. The individual unit of analysis excludes the Diffusion/Implementation model. Although, it may have many factors that are worth investigating like task and innovation characteristics (Prescott & Conger, 1995).

Social cognitive theory (SCT) also has an individual characteristic component that focuses on the personal factors. Social learning theory (SLT) has behavioral factors such as skills, practice, and self-sufficiency. It can be applied to a group, but the social aspect is not the focus of this research. The Tri-Core model is interested more in innovation from business, technical and administrative aspects which does not share the interest of this research that focuses on environmental, organizational and technical factors for macro level (Grover, Fiedler, & Teng, 1997). Institutional theory (Selznick) and IT Fashion theory (Abrahamson, 1991) investigate deep social structures that affect the adoption decision of technology within an organization.

The Unified Theory of Acceptance and Use of Technology (UTAUT) has wide insight into individual or organizations intent and behavior to use technology (Venkatesh, Morris, Davis, & Davis, 2003). Conversely, UTAUT focus is on the intention of the usage and not the factors. Many of the studied factors in UTAUT are individual specific (micro level) like gender, age, experience, etc. whereas other factors are focused on the technology fit like performance and effort expectancy which don't align with the stated goals of this research. Capability Maturity Model

Integration is an organizational level model that focuses on internal aspects of the organizations. That limits the scope of the study of adoption factors to internal factors.

Motivational model explores intrinsic and extrinsic motivations that affect the psychology of the user to adopt a technology. While this is valuable from motivations and intention perspectives, it does not address the external technical, organizational, and environmental factors that this research is interested in. Usability is another aspect of adoption study. It explores factors that affect the use and adoption of different technologies from a user perspective. Usability studies investigate how to make technology more usable and more accessible to the user which is micro level (Nielsen, 1999).

2.7 Why DOI and TOE?

The above theories do explore many important aspects of technology adoption. Conversely, Diffusion of Innovation theory (DOI) and Technology-Organization-Environment (TOE) framework fit the goals, scope and unit of analysis desired for this research. They both enable the study of antecedent factors that affect the technology adoption decision. TOE expands the antecedent adoption factors to include technology, organization, and environment. DOI and TOE offer organization level units of adoption analysis. They have both been used in many

studies in the Big Data arena specifically and other technologies in general. These studies have corroborated some of the found adoption factors and enables the research for other under explored or unexplored factors.

DOI theory supports the understanding of the process of Big Data adoption in organizations. It offers the ability to investigate the key determinant of Big Data Adoption (Soon, Lee, & Boursier, 2016). DOI provides an in-depth theoretical understanding of diffusion of innovation. It has also antecedent factors affecting adoption decisions. Specifically, this is relevant to this research since it covers factors of compatibility and technology cluster concept (Everett M Rogers, 2003). DOI theory offers 5 adoption factors (Everett M Rogers, 2003). This small number of adoption factors is limiting since there are many other factors that are studied in other theories.

The TOE framework has been used in much Information Systems adoption research. Technology-Organization-Environment framework has a wider perspective than DOI. It expands on the DOI adoption factors by adding other contexts for factors (technical, organizational, and environmental). In addition to adding these contexts, TOE framework offers the flexibility to identify other factors within each context. It allows for exploring other innovation adoption factors that are in technology, organization, and environment contexts. That expanded perspective and flexibility allow exploration of other factors that are not mentioned in DOI (DePietro, 1990).

This research will focus on the antecedent factors that determine Big Data adoption positively or negatively. It will use Diffusion of Innovation (DOI) theory (Everett M Rogers, 2003) with Technology-Organization-Environment (TOE) framework (DePietro, 1990). Multiple research that used this combination of theory and framework to study the Big Data adoption factors.

2.8 Diffusion of Innovation (DOI) Theory

The first edition of *Diffusion of Innovation* theory (DOI) was first published in 1962 by Everett M. Rogers (Everett M. Rogers, 1962). This book was republished several times. The 5th edition was the last one published in 2003 (Everett M Rogers, 2003). Rogers developed DOI to explain how 1) innovations are spread in 2) social systems through 3) communication channels 4) over time among adoption actors. Each of these elements are explored on how it affects the decision to adopt the innovation.

Historically, not all useful and more efficient innovations are adopted. For example, Dovark Keyboard is more efficient than the traditional qwerty keyboard. Yet not many people have heard or used it. Scurvy is a disease that killed thousands of sailors over the centuries. Its preventative cure was discovered over 260 years before it was implemented by the British navy (from 1601 to 1865 CE). DOI endeavors to answer many questions within the subject of diffusion in multiple

fields. Some of these questions are: Why do some innovations get adopted by individual actors and diffused into larger society? Why do some innovations take much longer to be adopted? What are the factors that play into the success and the speed of adoption?

Entities seek innovation to attain objectives or alleviate obstacles. Yet, there is uncertainty that any innovation and its implementation can accomplish these goals in each context. This ambiguity is greater in new innovations. The purpose of adopter actors is to reduce the uncertainty of achieving the adopter goals. One aspect of reducing this uncertainty is to study the factors that affect the adoption decision of an innovation by other adoption actors. Some of these factors are internal to the unit of adoption like the perceived relative advantage of adopting an innovation. Other factors are external like the complexity of the technology to be adopted.

One of the most cited diffusion theories in social sciences is Diffusion of Innovation by Evert Rogers (Nguyen & Petersen, 2017). The interest in diffusion for Rogers before the 1960s with research in rural sociology on agricultural innovations in rural areas. Rogers was not the first researcher into diffusion. He was a major pioneer and contributor to synthesize this research into a comprehensive theory (Xian, 2013). Rogers then published the first edition of his book *Diffusion of Innovation*. The motivation was to bring awareness to diffusion

research in general. Since then, 5 editions of the book have been published, the last edition being in 2003.

Diffusion is a special type of communication since it is only concerned with innovation. Rogers defines Diffusion as "The process by which an innovation is communicated through certain channels over time among the members of the social system." Adoption is done at the unit of adoption level, while diffusion happens at larger social context. The DOI research was initially developed in rural sociology context. It researched why and how some farming innovations were adopted while did not get the same rate of adoption.

Rogers traced innovation research historically from the early 1900's with Gabriel Tarde's law of limitation. He also described diffusion in different research traditions from sociology, anthropology, education, geography, communications, and others. He identified 8 areas of diffusion research as: earliness of knowing innovation, rate of innovation of different innovations, innovativeness, opinion leadership, diffusion networks, rate of innovation of different social systems, communication channel usage and consequences of innovation. He described successful innovation (in contrast to Plateaued adoption) as S shaped curve (Figure - 10).

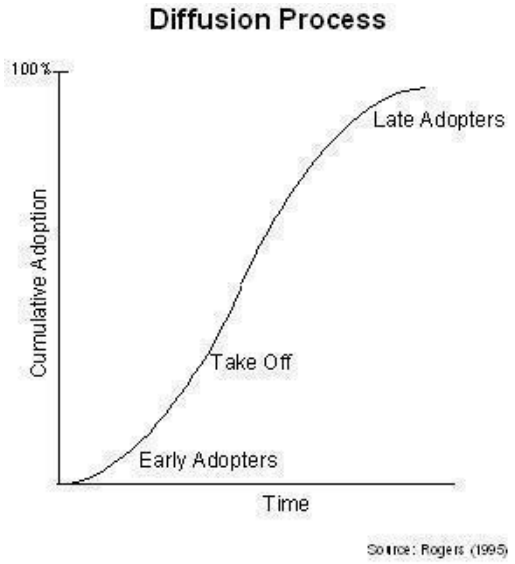

Figure 10 S Shaped Adoption Rate (Everett M Rogers, 2003).

Rogers explored steps of innovation generation as: need, research, development, commercialization, diffusion, and adoption and then consequences. He also developed a model of five stages in the innovation decision process. This model contains traits of the adoption decision making unit and antecedent factors that affect the adoption decision. Persuasion is the portion of the model that will be used in this research (Figure – 11).

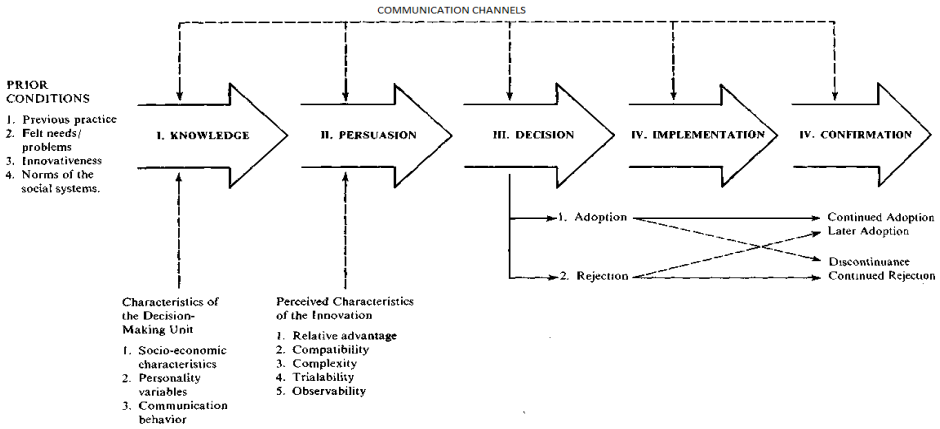

Figure 11 Roger's Model of Stages in the Innovation Decision Process (Everett M Rogers, 2003).

The following innovation adoption factors have been studied in many fields. In the Information Systems field, these factors have been studied for many technologies. They have also been studied for Big Data technology innovation. These factors in this model are:

1- Relative Advantage:

"The degree to which an innovation is perceived as being better than the idea it supersedes."

2- Compatibility:

"The degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters."

3- Complexity:

"The degree to which an innovation is perceived as relatively difficult to understand and use."

4- Trialability:

"The degree to which an innovation may be experimented with on a limited basis."

5- Observability:

"The degree to which the results of an innovation are visible to others."

These factors are expanded by combining DOI to other frameworks like TAM and TOE to cover other possible factors in the persuasion stage. These factors are summarized in Figure-12 (Nguyen & Petersen, 2017).

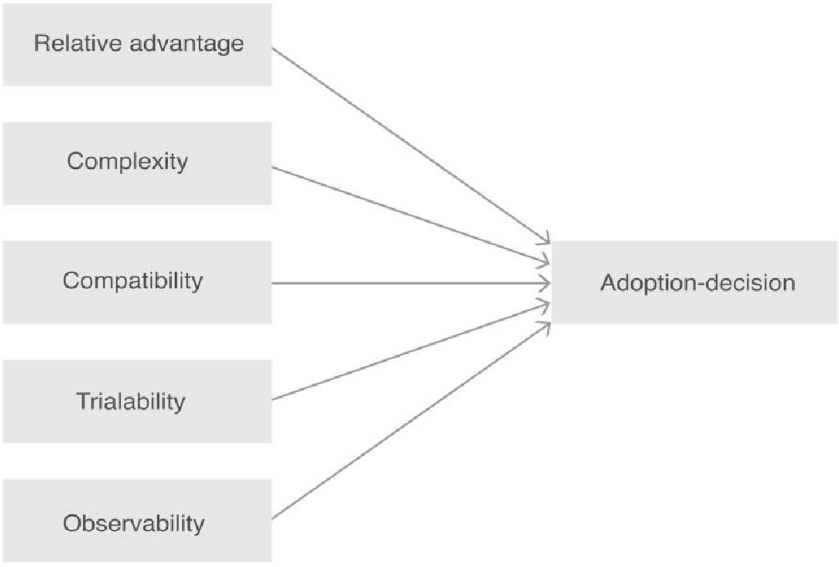

Figure 12 DOI Innovation Adoption Factors (Nguyen & Petersen, 2017).

2.9 Technology-Organization-Environment (TOE) Framework

Innovations and their adoption do not materialize in a vacuum. They are carried out by individuals in environments with specific requirements and challenges. These innovations are entangled in the environment and its requirements and are not only determined by the innovation itself or the adopter. Innovation can also be developed as a response to an environmental pressure (ex. Regulations). Thus, taking the environment and organizational factors is an interest of innovation research.

There are multiple attempts to address the organizational and environmental factors (in some cases described as contexts) in innovation adoption decisions. Wejnert in 2002 proposed a framework that classified innovation adoption factors into:

1- Characteristics of Innovations
2- Characteristics of innovators
3- Environmental context.

Li et al. in 2011 classified factors into: (Nguyen & Petersen, 2017)

1- Decision entity factors.
2- Decision object factors.
3- Context Factor.

Yet, the most recognized attempt in IT to include the organizational and environmental factors was part of a book titled *The Process of Technological Innovation* by Tornatzky, Fleischer and multiple other authors in 1990 (Tornatzky & Fleischer, 1990). This book had multiple authors for each chapter. Chapter seven is titled "*The Context for Change: Organization, Technology and Environment*" was authored by DePietro, Wiarda, & Fleischer. That is the chapter where technological, organizational, and environmental factors were defined and thus TOE was identified. Although most TOE framework references point to (Tornatzky & Fleischer, 1990), the TOE framework is produced by DePietro,

Wiarda, & Fleischer. Thus, this research will reference (DePietro, 1990) to this framework.

TOE framework proposed classifying the innovation adoption factors into:

1- Technology Factors:

These factors are technology specific to innovation.

2- Organizational Factors:

These factors are organizational specific to determine the adoption of the innovation.

3- Environmental Factors:

These factors are environmental specific to determine the adoption decision of the innovation.

Putting these contexts in this framework provides a holistic approach to innovation. Where these contexts are progressively wider in scope. Yet, deal with the same decision of adopting an innovation. As DePietro et al describe these factors as "both constraints and opportunities for technological innovation" (DePietro, 1990). These factors were summarized in Figure-13 (Nguyen & Petersen, 2017).

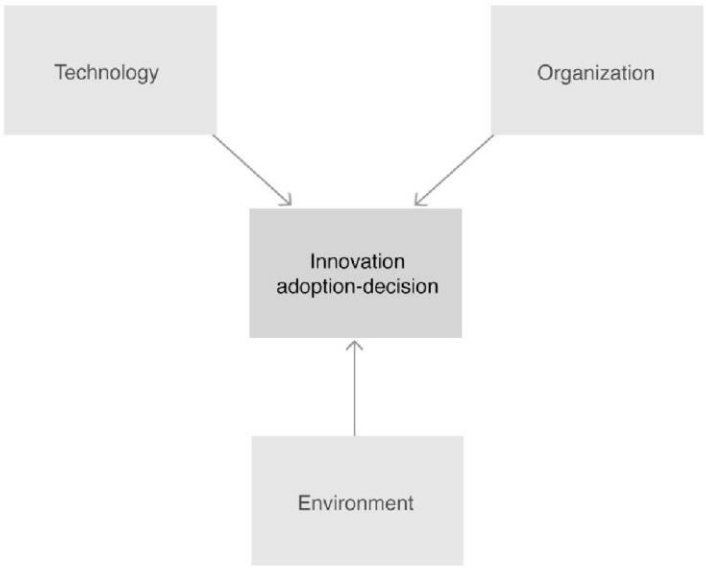

Figure 13 TOE Innovation Adoption Factors (Nguyen & Petersen, 2017).

The technological factors include both internal and external factors to the organization and the environment at large. The technical factors studied are in many cases the same or similar as the technological factors studied. Others have extended these factors to include other technical factors of the innovation itself. Other research covered the technological readiness of the organizations (Nguyen & Petersen, 2017). These factors will be explored further from the Big Data adoption perspective in the upcoming section of this chapter.

The organizational factors are factors within the organization that affect the innovation adoption decision. These organizational factors include the organizations' resources, their skills that promote or hinder the organization's ability to adopt innovation (DePietro, 1990).

2.10 Studied Factors Affecting Adoption of Big Data

There have been several studies on the factors affecting the adoption of Big Data. These studies utilized multiple theories and frameworks. Some studies used single theory or framework to explore these factors like DOI (Micheni, 2015), TOE (K. Agrawal, 2015; Nam et al., 2015; Verma & Bhattacharyya, 2017), Theory of Planned Behavior (Esteves & Curto, 2013), TAM (Lombardo, 2018) and maturity model (Olszak & Mach-Król, 2018). Most studies have used combinations of theories and frameworks to study the Big Data adoption factors. Combinations surveyed in this literature review included DOI+TOE (K. P. Agrawal, 2013; Bremser, 2018), TOE+IT Fashion (H.-M. Chen et al., 2015), DOI+TOE+ Institutional Theory (K. Agrawal, 2015), Resource Based View + isomorphism (O. Kwon, Lee, & Shin, 2014), DOI+TOE+TAM (Ajimoko, 2017; Nguyen & Petersen, 2017) and EMATEL-ANFIS + TOE (Yadegaridehkordi et al., 2018).

There is a diversity of aspects that factors affecting Big Data adoption are studied. There are multiple research methodologies applied like: interviews (Mach-Król, 2017; Skaale & Rygh, 2018; Verma & Bhattacharyya, 2017; Zanabria & Mlokozi, 2018), case studies (Dremel et al., 2017; Gong & Janssen, 2017) and

surveys (Nguyen & Petersen, 2017; Verma & Bhattacharyya, 2017; L. Wang, Yang, Pathan, Salam, & Shahzad, 2018). These studies are also varied in geography. They are studied in China (L. Wang et al., 2018), Germany (Bremser, 2018), India (Verma & Bhattacharyya, 2017), Korea (M.-K. Kim & Park, 2017), Norway (Nguyen & Petersen, 2017), Poland (Mach-Król, 2017), Sweden (Zanabria & Mlokozi, 2018) and USA (Ghosh, 2018).

I have expanded and modified the literature review (Nguyen & Petersen, 2017; Sun et al., 2018) and have produced a summary of the studied antecedent factors affecting Big Data adoptions. This summary includes the adoption factors, the definition of the adoption factor and the literature that have studied that adoption factor. The factors are listed alphabetically and filtered for recent studies post 2009. Big Data adoption factors can be categorized as technical, organizational, and environmental according to the TOE framework. Therefore, the adoption factors are categorized in 3 tables corresponding to the TOE framework. From the literature examined, there are 11 technical factors, 40 organizational factors and 28 environmental factors. See Tables 2, 3 and 4 for table summary of the literature.

Many of the technical adoption factors studied are the same as DOI technical factors (Relative Advantage, Compatibility, Complexity, Trialability and Observability). There were six factors that do not correspond with the original DOI factors. In contrast, the organizational and environmental are more innovative and diverse in coverage. They go beyond the classical factors that are mentioned in the

original literature. They cover a wider range of factors that were not studied before. This disproportionate emphasis on technical adoption factors presents an opportunity to add more factors in the technical category. It also presents an opportunity to study organizational and environmental factors that may have not been covered.

2.10.1 Summary of Studied Technological Factors Affecting Adoption of Big Data

This research survey of the literature on Big Data adoption factors shows technological adoption factors of Big Data to be the least studied. There are 11 technological factors that were studied in the Big Data realm. Five of these were the classical DOI factors and the rest were new factors. The literature survey of technological adoption factors of Big Data is summarized alphabetically in Table 2. Since DOI only covers technological factors, this distinction will only be here and will not be in organizational or environmental factors. The following paragraph will cover the current research starting with DOI factors first.

2.10.2 DOI Studied Technological Factors Affecting Adoption of Big Data

Relative Advantage is defined as "The degree to which an innovation is perceived as being better than the idea it supersedes". It was also described as perceived usefulness of the technology (Nguyen & Petersen, 2017; Sun et al.,

2018). It is the most cited factor in technological context. Relative Advantage in Big Data context was studied in multiple papers (K. Agrawal, 2015; K. P. Agrawal, 2013; Ajimoko, 2017; Boonsiritomachai, 2014; H.-M. Chen et al., 2015; Esteves & Curto, 2013; Hung, Huang, Lin, Chen, & Tarn, 2016; M.-K. Kim & Park, 2017; Mahesh et al., 2018; Nam et al., 2015; Nguyen & Petersen, 2017; Park, Kim, & Paik, 2015; L. Wang et al., 2018; Yadegaridehkordi et al., 2018; Zanabria & Mlokozi, 2018).

Compatibility is defined as "The degree to which innovations are perceived as being consistent with existing methods for executing their mission". It was studied by multiple research papers (K. Agrawal, 2015; H.-M. Chen et al., 2015; Esteves & Curto, 2013; Mahesh et al., 2018; Salleh & Janczewski, 2018; Verma & Bhattacharyya, 2017).

Complexity is defined as "The degree to which an innovation is perceived to be relatively difficult to understand and use". It was also described as perceived ease of use (Nguyen & Petersen, 2017). It is the second most cited factor in technological context. It was studied by multiple research papers (K. Agrawal, 2015; K. P. Agrawal, 2013; Boonsiritomachai, 2014; Esteves & Curto, 2013; Hung et al., 2016; Mahesh et al., 2018; Nguyen & Petersen, 2017; Salleh & Janczewski, 2018; Verma & Bhattacharyya, 2017; Yadegaridehkordi et al., 2018; Zanabria & Mlokozi, 2018).

Observability is defined as “The degree to which the results of an innovation are visible to others”. Observability was studied by (Boonsiritomachai, 2014). Another DOI technological factor is Trialability. Trialability is defined as “The degree to which an innovation may be experimented with on a limited basis”. It was studied by (Ramdani, Lorenzo, & Kawalek, 2009) and mentioned by Nguyen & Petersen in 2017 (Nguyen & Petersen, 2017).

2.10.3 Non-DOI Studied Technological Factors Affecting Adoption of Big Data

The following adoption factors are studied as technological adoption factors for Big Data. They are not part of the DOI theory but an extension of it. IT assets is an adoption factor that was studied by Verma & Bhattacharyya, 2017. It is defined as “The degree to which an innovation is associated with complex procedures” (Verma & Bhattacharyya, 2017). Perceived Indirect Benefit is a similar factor to relative advantage/perceived usefulness, but it has an important distinction of strategic or indirect benefit that will take effect indirectly and not immediately (Nam et al., 2015). Scalability is defined as “The need for innovative solutions for data models, algorithms and architectures have to be designed providing the necessary scalability and flexibility for novel Big Data analytics applications” (Motau & Kalema, 2016).

Security is another technological adoption factor that has organizational and environmental dimensions as well (which will be covered in the following sections). In the technology context security is defined as the ability to protect Big Data and privacy from malicious attacks and the misuse of data (Motau & Kalema, 2016; Nguyen & Petersen, 2017). Technology infrastructure factor is the ability of the internal infrastructure to adopt Big Data (Mach-Król, 2017; Malladi & Krishnan, 2013; Motau & Kalema, 2016; Olszak & Mach-Król, 2018; Yeh, Lee, & Pai, 2015). Technology readiness/maturity is the last technological adoption factor discovered in this survey. It is defined as "The maturity of the information technology within an organization and its information technology capabilities encourage the organization to apply information technology to achieve its strategic goals" (L. Wang et al., 2018; Yeh et al., 2015). The technology readiness/maturity does not only cover the technical readiness of the IT infrastructure but also the ability to use that technology (for example, appropriate data collection and data storage).

Technological Factors

No.	Factor	Definition	Sources
1	Compatibility	The degree to which innovations are perceived as being consistent with existing methods for executing their mission	(Esteves & Curto, 2013), (K. Agrawal, 2015), (H.-M. Chen et al., 2015), (Verma & Bhattacharyya, 2017), (Salleh & Janczewski, 2018), (Mahesh et al., 2018)
2	IT Assets	The degree to which an innovation is associated with complex procedures.	(Verma & Bhattacharyya, 2017)
3	Observability	The degree that potential adopters of an innovation can perceive the results of using that innovation from users who have already adopted it	(Boonsiritomachai, 2014)

4	Perceived ease of use/Complexity	The degree to which an innovation is perceived to be relatively difficult to understand and use	(K. P. Agrawal, 2013), (Esteves & Curto, 2013), (Boonsiritomachai, 2014), (K. Agrawal, 2015), (Hung et al., 2016), (Verma & Bhattacharyya, 2017), (Nguyen & Petersen, 2017), (Salleh & Janczewski, 2018), (Mahesh et al., 2018), (Yadegaridehkordi et al., 2018), (Zanabria & Mlokozi, 2018)
5	Perceived Indirect Benefit	The strategic benefits, i.e., development of corporate strategies through the building of external relationships with customers, partners, and competitors.	(Nam et al., 2015)

6	Perceived usefulness/Relative advantage	The degree to which an innovation is perceived as being better than the idea it supersedes	(K. P. Agrawal, 2013), (Esteves & Curto, 2013), (Boonsiritomachai, 2014), (K. Agrawal, 2015), (H.-M. Chen et al., 2015), (Nam et al., 2015), (Park et al., 2015), (Hung et al., 2016), (Ajimoko, 2017), (M.-K. Kim & Park, 2017), (Nguyen & Petersen, 2017), (Mahesh et al., 2018), (Yadegaridehkordi et al., 2018) (L. Wang et al., 2018), (Zanabria & Mlokozi, 2018)
7	Scalability	The need for innovative solutions for data models, algorithms and architectures must be designed providing the necessary scalability and flexibility for novel Big Data	(Motau & Kalema, 2016)

		analytics applications.	
8	Security	Security is protection of Big Data and privacy from malicious attacks and the misuse of data.	(Motau & Kalema, 2016), (Nguyen & Petersen, 2017)
9	Technology infrastructure	The internal technology ability to adopt new technology or the degree to which a firm has necessary technology infrastructure to adopt	(Malladi & Krishnan, 2013), (Yeh et al., 2015), (Motau & Kalema, 2016), (Olszak & Mach-Król, 2018)
10	Technology readiness/ Maturity	The maturity of the information technology within an organization and its information technology capabilities encourages the organization to apply information technology to achieve its strategic goals.	(Yeh et al., 2015), (L. Wang et al., 2018)

11	Trialability	The extent to which potential adopters can experiment with an innovation	(Ramdani et al., 2009)

Table 2 Studied Big Data Adoption Technological Factors.

2.10.4 Summary of Studied Organizational Factors Affecting Adoption of Big Data

The following adoption factors for Big Data are organizational specific. The focus of these factors are traits of the organization and its capabilities. The first factor surveyed is appropriateness adoption factor. The timing of the adoption of big data is advantageous for the organization (Sun et al., 2018). Business resources adoption factor is the firm's business resources that are adequate to the task of adopting big data (Sun et al., 2018). Business strategy orientation is the strategy that is oriented to business analytics and using big data for strategic decisions (Sun et al., 2018). Business IT alignment is the alignment of information systems capabilities with business goals (H.-M. Chen et al., 2015).

Centralization organizational adoption factor is the degree to which power and control are concentrated in the hands of relatively few individuals in an organization (Hameed, Counsell, & Swift, 2012). Change efficacy is the ability of an organization's members to easily handle the changes triggered by the adoption of big data (Sun et al., 2018). Complexity tolerance is the extent to which an

enterprise can tolerate the complexity in the technology and in its implementation process (H.-M. Chen et al., 2015). Data Environment is the extent to which data resources are managed in an organization (Joshi & Biswas, 2018; Verma & Bhattacharyya, 2017). Decision-making culture is top managers' decision making at the firm level (e.g., culture of evidence-based decision making, decision-making norms) (Sun et al., 2018). Digital Strategy adoption factor is the strategy and supported by all management levels for Big Data and digitization (Bremser, 2018).

Economic Inertia organizational adoption factor for Big Data is form of commitment to previously implemented IT solutions that do not pay off and create sunk costs, or through transition expenses which cause organizations to not adopt potentially better alternatives (Mikalef, Pappas, Krogstie, & Giannakos, 2018). Financial Support is the financial resources available to adapt Big Data (Bremser, 2018; Motau & Kalema, 2016; L. Wang et al., 2018). Fit with the business model factor is the compatibility of the IT innovation with the existing business model (H.-M. Chen et al., 2015).

Formalization organizational adoption factor is the degree to which an organization follows the rules and procedures on the role of performance of its members (Hameed et al., 2012). Information security culture factor is like the technological factor but addresses security from an organizational perspective. It is defined as the totality of patterns of behavior in an organization that contribute to the protection of information of all kinds (Salleh & Janczewski, 2018).

Innovator's dilemma factor speaks to the reluctance to adopt new innovations that impact existing mature business models (H.-M. Chen et al., 2015). IS strategy orientation factor being the firm's IS strategy prioritizes big data usage (Sun et al., 2018).

IT expertise organizational factor is the prior experience of IT employees in terms of skill and knowledge (K. Agrawal, 2015; Nam et al., 2015; Nguyen & Petersen, 2017; L. Wang et al., 2018). The IT Infrastructure organizational factor is defined in a wider scope than the technological factor. It includes the tangible resource comprising the physical IT infrastructure components, the human IT resources comprising the technical and managerial IT skills and the intangible IT-enabled resources such as knowledge assets, customer orientation, and synergy (K. P. Agrawal, 2013). Management ability is the capacity to increase efficiency of big data projects (K. P. Agrawal, 2013; L. Wang et al., 2018). Maturity is defined as the organization's capability to adopt big data (Olszak & Mach-Król, 2018). Maturity of data architecture is how the data is architected in the organization (harmonized vs fragmented) (Bremser, 2018).

Negative psychology is an adoption factor in the organization where some employees fear that the introduction of big data analytics and the corresponding technologies and tools for analyzing and visualizing data would render their skills as non-significant (Mikalef et al., 2018). Organizational absorptive capacity or human resources is an organizational adoption factor where it represents the

organizational ability of its members to utilize existing or pre-existing IT knowledge (K. Agrawal, 2015; K. P. Agrawal, 2013; Hung et al., 2016; O. Kwon et al., 2014; Mahesh et al., 2018; Sun et al., 2018). Organizational absorptive capacity is the seconds most cited organizational factor that was surveyed.

Organizational culture is another organizational adoption factor. It can be present at various levels (national, organizational, group) and can affect the success of IT (Joshi & Biswas, 2018; Puklavec, Oliveira, & Popovič, 2014). Organizational Innovation Process and Organizational innovativeness are two distinct organizational adoptions factors. Where organizational approaches to innovation and introducing new emerging IT innovations can take multiple approaches: top-down or bottom up; formal or informal (H.-M. Chen et al., 2015). Whereas, Organizational innovativeness is the willingness to take a risk and trying new solutions that have not been tried or tested before (Bremser, 2018; H.-M. Chen et al., 2015).

Organizational Learning Culture is an organizational adoption factor. It is the ability or processes in an organization that enables the acquisition, access and revision of organizational memory, which in turn leads to organizational actions (Salleh & Janczewski, 2018). Organizational readiness is one of the most studied factors in the factors surveyed. It is defined as the degree to which an organization has the awareness, resources, commitment, and governance to adopt IT (H.-M.

Chen et al., 2015; Joshi & Biswas, 2018; Park et al., 2015; Sun et al., 2018; Zanabria & Mlokozi, 2018).

Perceived cost and perceived financial readiness are organizational financial adoption factors. Perceived Cost is a factor in adopting Big Data as expenses of implementing necessary technologies in organizations and efforts devoted to organizational restructuring and process re-engineering (Verma & Bhattacharyya, 2017). Perceived financial readiness is defined as the financial resources available to pay for new technology innovation costs, for implementation of any subsequent enhancements, and ongoing expenses during usage (Nam et al., 2015).

Political Inertia is a form of organizations lock-in to their vendors that make adopting new technology more difficult (Mikalef et al., 2018). Project champion is an individual who performs the task of spreading knowledge of new technology within the organization about the new technology and advocates for it (Puklavec et al., 2014). Size of the organization is a factor in Big Data adoption. It is one of the most studied organizational factors. It is the size of the firm (i.e., the number of employees and annual revenue (K. Agrawal, 2015; Hung et al., 2016; Joshi & Biswas, 2018; Mahesh et al., 2018; Motau & Kalema, 2016).

Slack or organizational resources represent the organizational resources which are not committed to an existing business operation, and subsequently can be used in a discretionary manner (Boonsiritomachai, 2014; Hameed et al., 2012;

Nam et al., 2015; Park et al., 2015). Socio-Cognitive Inertia is the different mental models, use of language, and objectives caused conflicts that threatened and even greatly delayed big data implementation projects (Mikalef et al., 2018). Socio-technical Inertia can be defined as some employees' fear that decision-making would now reside in insight from analytics, therefore replacing them (Mikalef et al., 2018).

Technology Resources is the firm's technology resources adequate for the task of adopting Big Data (Mahesh et al., 2018; Sun et al., 2018; Zanabria & Mlokozi, 2018). Top management support is the most studied organizational factor. It is defined as the degree to which top management understands the importance of the technology and the extent to which it is involved in related initiatives (Hung et al., 2016; Motau & Kalema, 2016; Nguyen & Petersen, 2017; Park et al., 2015; Puklavec et al., 2014; Salleh & Janczewski, 2018; Verma & Bhattacharyya, 2017; L. Wang et al., 2018; Yeh et al., 2015; Zanabria & Mlokozi, 2018). Willingness to Explore is the willingness to explore based on a sense of urgency and/or sense of opportunity (Caesarius & Hohenthal, 2018).

Organizational Factors

No.	Factor	Definition	Sources
1	Appropriateness	The timing of the adoption of big data is advantageous for the organization	(Sun et al., 2018)
2	Business resources	The firm has business resources that are adequate to the task of adopting big data	(Sun et al., 2018)
3	Business strategy orientation	An organization strategy that is oriented to business analytics and using big data for strategic decisions.	(Sun et al., 2018)
4	Business-IT Alignment	Aligning information systems capabilities with business goals.	(H.-M. Chen et al., 2015)
5	Centralization	The degree to which power and control are concentrated in the hands of relatively few individuals in an organization	(Hameed et al., 2012)
6	Change efficacy	Organization members can easily handle the changes triggered by the adoption of big data.	(Sun et al., 2018),

7	Complexity Tolerance	The extent to which an enterprise can tolerate the complexity in the technology and in its implementation process.	(H.-M. Chen et al., 2015)
8	Data Environment	The extent to which data resources are managed in an organization.	(Verma & Bhattacharyya, 2017), (Joshi & Biswas, 2018)
9	Decision-making culture	Top managers' decision making at the firm level (e.g., culture of evidence-based decision making, decision-making norms).	(Sun et al., 2018),
10	Digital Strategy	Big Data or digitization is part of the strategy and supported by all management levels.	(Bremser, 2018)
11	Economic Inertia	The form of commitment to previously implemented IT solutions that do not pay off and create sunk costs, or through transition expenses which cause organizations to not adopt	(Mikalef et al., 2018)

		potentially better alternatives.	
12	Financial Support	The financial resources available to adapt Big Data.	(Motau & Kalema, 2016), (L. Wang et al., 2018), (Bremser, 2018)
13	Fit with Business Model	The compatibility of the IT innovation with the existing business model.	(H.-M. Chen et al., 2015)
14	Formalization	The degree to which an organization follows the rules and procedures on the role of performance of its members	(Hameed et al., 2012)
15	Information Security Culture	The totality of patterns of behavior in an organization that contribute to the protection of information of all kinds	(Salleh & Janczewski, 2018)
16	Innovator's Dilemma	The reluctance to adopt new innovations that impact existing mature business models.	(H.-M. Chen et al., 2015)
17	IS strategy orientation	The firm's IS strategy prioritizes big data usage	(Sun et al., 2018)

18	IT expertise	The prior experience of IT employees in terms of skill and knowledge	(K. Agrawal, 2015), (Nam et al., 2015), (Nguyen & Petersen, 2017), (L. Wang et al., 2018)
19	IT Infrastructure	The tangible resource comprising the physical IT infrastructure components, the human IT resources comprising the technical and managerial IT skills and the intangible IT-enabled resources such as knowledge assets, customer orientation, and synergy.	(K. P. Agrawal, 2013)
20	Management Ability	The ability to increase efficiency of big data projects.	(K. P. Agrawal, 2013), (L. Wang et al., 2018)
21	Maturity	Maturity assesses an organization's capability to adopt big data.	(Olszak & Mach-Król, 2018).
22	Maturity of data architecture	Harmonized vs. fragmented data architecture.	(Bremser, 2018)
23	Negative psychology	Some employees fear that the introduction of	(Mikalef et al., 2018)

		big data analytics and the corresponding technologies and tools for analyzing and visualizing data would render their skills as non-significant.	
24	Organizational absorptive capacity/Human Resources	Absorptive capacity of an organization is the ability of its members to utilize existing or pre-existing IT knowledge.	(K. P. Agrawal, 2013), (O. Kwon et al., 2014), (K. Agrawal, 2015), (Hung et al., 2016), (Sun et al., 2018), (Mahesh et al., 2018)
25	Organizational culture	Culture at various levels (national, organizational, group) can affect success of IT	(Puklavec et al., 2014), (Joshi & Biswas, 2018)
26	Organizational Innovation Process	The organizational process of introducing new emerging IT innovations: top-down or bottom up; formal or informal.	(H.-M. Chen et al., 2015)
27	Organizational innovativeness	Innovativeness is the willingness degree of taking a risk and trying new solutions that	(H.-M. Chen et al., 2015), (Bremser, 2018).

		not been tried or tested before	
28	Organizational Learning Culture	The ability or processes in an organization that enables the acquisition, access, and revision of organizational memory, which in turn leads to organizational actions.	(Salleh & Janczewski, 2018)
29	Organizational readiness	The degree to which an organization has the awareness, resources, commitment, and governance to adopt IT	(H.-M. Chen et al., 2015), (Park et al., 2015), (Sun et al., 2018), (Joshi & Biswas, 2018), (Zanabria & Mlokozi, 2018)
30	Perceived Cost	The expenses of implementing necessary technologies in organizations and efforts devoted to organizational restructuring and process re-engineering	(Verma & Bhattacharyya, 2017)
31	Perceived Financial Readiness	The financial resources available to pay for new technology innovation costs, for implementation	(Nam et al., 2015)

		of any subsequent enhancements, and ongoing expenses during usage.	
32	Political Inertia	The form of organizations lock-in to their vendors that make adopting new technology more difficult.	(Mikalef et al., 2018)
33	Project champion	An individual who performs the task of spreading knowledge of new technology within the organization.	(Puklavec et al., 2014)
34	Size	The size of the firm (i.e., the number of employees and annual revenue)	(K. Agrawal, 2015), (Hung et al., 2016), (Motau & Kalema, 2016), (Mahesh et al., 2018), (Joshi & Biswas, 2018)
35	Slack/ organizational resources	Those resources an organization has acquired which are not committed to an existing business operation, and subsequently can be used in a discretionary manner	(Hameed et al., 2012), (Boonsiritomachai, 2014), (Nam et al., 2015), (Park et al., 2015)

36	Socio-Cognitive Inertia	The different mental models, use of language, and objectives caused conflicts that threatened and even greatly delayed big data implementation projects.	(Mikalef et al., 2018)
37	Socio-technical Inertia	Some employee's fear that decision-making would now reside in insight from analytics, therefore replacing them.	(Mikalef et al., 2018)
38	Technology Resources	The firm's technology resources are adequate for the task of adopting big data.	(Sun et al., 2018), (Mahesh et al., 2018), (Zanabria & Mlokozi, 2018)
39	Top management support	The degree to which top management understands the importance of the technology and the extent to which it is involved in related initiatives	(Yeh et al., 2015), (Puklavec et al., 2014), (Park et al., 2015), (Hung et al., 2016), (Motau & Kalema, 2016), (Verma & Bhattacharyya, 2017), (Nguyen & Petersen, 2017), (Salleh & Janczewski, 2018),

			(L. Wang et al., 2018), (Zanabria & Mlokozi, 2018)
40	Willingness to Explore	The willingness to explore based on a sense of urgency and/or sense of opportunity.	(Caesarius & Hohenthal, 2018)

Table 3 Studied Big Data Adoption Organizational Factors.

2.10.5 Summary of Studied Environmental Factors Affecting Adoption of Big Data

Environmental factors affecting adoption of Big Data focus on external factors to the organization that affects it. Environmental factors do not fall under the direct control of the organization but affect the decision to adopt new innovations in general and Big Data in specific. Business Partners is an environmental adoption factor for Big Data. These partners are already working with components and data have a sizable effect on the Big Data adoption decision.

These partners can affect how the data is used and how they are managed (Zanabria & Mlokozi, 2018). One of the most studied environmental factors in Big Data is competitive pressure. Competitive pressure is the degree of pressure that the company faces from competitors within the industry (K. Agrawal, 2015; K. P. Agrawal, 2013; Boonsiritomachai, 2014; Bremser, 2018; H.-M. Chen et al., 2015; Mahesh et al., 2018; Malladi & Krishnan, 2013; Motau & Kalema, 2016; Nam et al., 2015; Nguyen & Petersen, 2017).

External pressure from suppliers and customers is another environmental factor determining the Big Data adoption (Motau & Kalema, 2016; Nam et al., 2015; Verma & Bhattacharyya, 2017). On the opposite end of external environmental factors is External support. It is where availability of support for implementing and using an information system (Hung et al., 2016; Nguyen & Petersen, 2017). There is also a fear factor. There is a fear of missing out adoption factor where the fear of missing a significant market opportunity or profitable investment or innovations which competitors are seeking can affect the Big Data adoption decision (H.-M. Chen et al., 2015). There is also the fear of uber effect where the fear of disruptive business models from others in one's market space can affect the adoption decision (H.-M. Chen et al., 2015).

Industry & market complexity as the degree and instability of change in a firm's environment is another environmental adoption factor (K. Agrawal, 2015). Not all industries have the same dynamics of Big Data adoption and that is another

adoption factor. Institutional based trust is the firm's belief that it will be safe to adopt big data (Sun et al., 2018). Investors' motivation for businesses to invest in providing services is another adoption factor (Zanabria & Mlokozi, 2018). Looking for "IS fashion", by observing organization's peers and perceived experts such as vendors and customers, is an environmental adoption factor (Bremser, 2018; H.-M. Chen et al., 2015; Sun et al., 2018).

Government plays a role in technology adoption in general and Big Data specifically. Legislation barriers of government policy, inadequate legal protection or business laws is a factor in adoption decisions (Gibbs & Kraemer, 2004). Market turbulence, where changes in customers' product preferences, demands, and needs in a big data environment, is a factor in adoption decisions (Sun et al., 2018). Media and press news about Big Data affect the adoption decision (Esteves & Curto, 2013).

Paradigm shift where the change in the basic assumptions or fundamental practices or paradigms in a discipline affects the decision to adopt Big Data (H.-M. Chen et al., 2015). Partners power measures the strength of the influence strategy (e.g., rewards and threats) used to exercise that potential power of technology adoption (Alrousan, 2015; Motau & Kalema, 2016; Park et al., 2015). Perceived industry pressure is the degree that the firm is affected by competitors and partners in the market (Nam et al., 2015). Political influence when presenting the adoption of the technology influences the adoption decision (Zanabria & Mlokozi, 2018).

Regulatory entities readiness for regulations related with national data policies is another adoption factor (Zanabria & Mlokozi, 2018). Regulatory environment adequacy of institutional frameworks and business laws governing the use of innovations/technology is an environmental adoption factor (K. Agrawal, 2015; Bremser, 2018; Sun et al., 2018).

Regulatory Support is the various types of incentives and assistance provided by the governments and related regulatory authorities (K. Agrawal, 2015; K. P. Agrawal, 2013; Mahesh et al., 2018). Risks in Outsourcing is the perceived degree of security and privacy risks associated with outsourcing (Salleh & Janczewski, 2018). Security and Privacy Regulatory Concerns is the level of concern organizations have towards the requirement to comply with security and privacy regulations (Nam et al., 2015; Salleh & Janczewski, 2018). Security, privacy and ethics of data collection from individuals causes individuals' security, privacy concerns (Sun et al., 2018). Social Influence of friends and/or colleagues' suggestion to adopt Big Data is an adoption factor (Esteves & Curto, 2013). Trading partners' readiness to adopt big data to follow partners and maintain the firm's internal balance with them is another adoption factor (Sun et al., 2018). Uncertainty/risk is the concerns regarding potential unexpected consequences related to Big Data adoption (K. Agrawal, 2015; K. P. Agrawal, 2013; Bremser, 2018; Esteves & Curto, 2013; Mahesh et al., 2018; Sun et al., 2018).

Environmental Factors

No.	Factor	Definition	Sources
1	Business Partners	Businesses already working with components make better use of the generated data.	(Zanabria & Mlokozi, 2018)
2	Competitive pressure	The degree of pressure that the company faces from competitors within the industry	(K. P. Agrawal, 2013), (Malladi & Krishnan, 2013), (Boonsiritomachai, 2014), (H.-M. Chen et al., 2015), (Nam et al., 2015), (K. Agrawal, 2015), (Motau & Kalema, 2016), (Nguyen & Petersen, 2017), (Mahesh et al., 2018), (Bremser, 2018)
3	Data Resources	The ability of enterprise, government, and society to improve data resources acquisition, promote data sharing and transactions.	(L. Wang et al., 2018)

4	External pressure	External pressure applied by suppliers and customers	(Nam et al., 2015), (Motau & Kalema, 2016), (Verma & Bhattacharyya, 2017)
5	External support	Availability of support for implementing and using an information system	(Hung et al., 2016), (Nguyen & Petersen, 2017)
6	Fear of Missing Out	The fear of missing a significant market opportunity or profitable investment or innovations which competitors are seeking.	(H.-M. Chen et al., 2015)
7	Fear of Uber Effect	The fear of disruptive business models from others in one's market space.	(H.-M. Chen et al., 2015)
8	Industrial Development	The cultivation of data talents, technological development, data trade platforms, data resources, to encourage the development of big data.	(L. Wang et al., 2018)

9	Industry & market complexity	The degree and instability of change in a firm's environment	(K. Agrawal, 2015)
10	Industry Type	The sector to which the business belonged.	(Verma & Bhattacharyya, 2017)
11	Institutional based trust	The firm's belief that it will be safe to adopt big data.	(Sun et al., 2018)
12	Investors	The motivation for businesses to invest in providing services.	(Zanabria & Mlokozi, 2018)
13	IS fashion	Information is obtained through external communication channels by focusing on an organization's peers and perceived experts such as vendors and customers	(H.-M. Chen et al., 2015), (Sun et al., 2018), (Bremser, 2018)
14	Legislation barriers	Government policy, inadequate legal protection, or business laws	(Gibbs & Kraemer, 2004)
15	Market turbulence	Changes in customers' product preferences, demands, and needs in a big data environment	(Sun et al., 2018)

16	Media	Media and press news about Big Data.	(Esteves & Curto, 2013)
17	Paradigm Shift	The change in the basic assumptions or fundamental practices or paradigms in a discipline.	(H.-M. Chen et al., 2015)
18	Partners	Enacted trading partner power measures the strength of the influence strategy (e.g., rewards and threats) used to exercise that potential power.	(Alrousan, 2015), (Park et al., 2015), (Motau & Kalema, 2016)
19	Perceived Industry Pressure	The degree that the firm is affected by competitors and partners in the market.	(Nam et al., 2015)
20	Political influence	The political influence when presenting the adoption of the technology.	(Zanabria & Mlokozi, 2018)
21	Regulatory entities readiness	The readiness of the entities in charge of regulations related with national data policies.	(Zanabria & Mlokozi, 2018)
22	Regulatory environment	The adequacy of institutional frameworks and business laws governing the use	(K. Agrawal, 2015), (Sun et al., 2018), (Bremser, 2018)

		of innovations/technology	
23	Regulatory Support	The various types of incentives and assistance provided by the governments and related regulatory authorities.	(K. P. Agrawal, 2013), (K. Agrawal, 2015), (Mahesh et al., 2018)
24	Risks in Outsourcing	The perceived degree of security and privacy risks associated with outsourcing.	(Salleh & Janczewski, 2018)
25	Security and Privacy Regulatory Concerns	The level of concern organizations has towards the requirement to comply with security and privacy regulations.	(Nam et al., 2015), (Salleh & Janczewski, 2018)
26	Security, privacy, and ethics	Data collection from individuals causes individuals' security, privacy concerns.	(Sun et al., 2018),
27	Social Influence	Friends and/or colleagues' suggestion to adopt Big Data.	(Esteves & Curto, 2013)

28	Trading Partners Readiness	Adopt big data to follow partners and maintain the firm's internal balance with them.	(Sun et al., 2018),
29	Uncertainty/ Risk	Concerns regarding potential unexpected consequences related to big data adoption	(K. P. Agrawal, 2013), (Esteves & Curto, 2013), (K. Agrawal, 2015), (Sun et al., 2018), (Mahesh et al., 2018), (Bremser, 2018)

Table 4 Studied Big Data Adoption Environmental Factors.

2.11 Summary

The potential of the Big Data technology is estimated to be substantial and compared to the oil impact on industry and world economy (Hirsch, 2014). Big Data is near the peak of the technology hype cycle (Hall, 2013). Yet, Big Data adoption is still at early stages (Hall, 2013; Nguyen & Petersen, 2017). Most organizations have not adopted Big Data in production, in some sectors around 20% of companies put Big Data into production use and Big Data projects had a failure rate of 55% (H.-M. Chen et al., 2015; Columbus, 2017). This paradox of high expectations, potential, excitement for Big Data and its low adoption needs to be studied further to offer better insights to solve it. Part of that effort is to identify

significant determinants affecting adoptions in multiple contexts (de Camargo Fiorini et al., 2018). Each organization with a data storage system is a potential adaptor of Big Data. This research investigates how to enable swaths of data storage providers to adopt Big Data.

DOI is one of the most cited technology theories in many fields including information Systems (Nguyen & Petersen, 2017; Everett M Rogers, 2003). It provides in depth understanding of adoption at the individual unit of adoption level and diffusion at the larger social and macro levels. The TOE framework expands the adoption factors by including organizational and environmental factors in addition to the technological factors that are in DOI. DOI and TOE have been used in tandem to study the adoption of many technologies and other phenomena. DOI and TOE will be used in this research to expand our understanding of factors affecting Big Data adoption.

This research intends to expand the current understanding of this gap by studying novel adoption factors that contribute to Big Data adoption. This research has captured current Big Data adoption factors. Some of these factors are too broad. Others are novel and have not been studied. While other factors have been limited studies on them. This research intends to expand the study of these factors.

Big Data adoption factors that were too broadly studied like computability (K. Agrawal, 2015; H.-M. Chen et al., 2015; Esteves & Curto, 2013; Mahesh et al., 2018; Salleh & Janczewski, 2018; Verma & Bhattacharyya, 2017). Compatibility

in technological context will be studied in more granular aspects of data storage systems' latency, ability to compute large data, and interface compatibility. Novelty will also arise from newly studied factors like open-source software and enterprise sourced software in organizational context. Perceived industry pressure for Big Data solution and services will also be studied from environmental context. This research will study less studied factors like cost (Verma & Bhattacharyya, 2017) in organizational adoption factors and legislation barriers (Gibbs & Kraemer, 2004), and market turbulence (Sun et al., 2018) for environmental context.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Research Objectives

To find the significant antecedent adoption factors of Big Data, one can resort to quantitative or qualitative research methods. Qualitative research methods add investigation, depth, complexity and meaning. On the other hand, quantitative research methods provide precision, rigor of hypotheses testing, larger sample size and generalization. Each has its own strengths and weaknesses. Pragmatic approach can utilize both research methods. Plenty of research literature advocate a strong case for pragmatic mixed methods approach (R. B. Johnson & Onwuegbuzie, 2004; Venkatesh, Brown, & Bala, 2013).

This research will utilize quantitative dominant mixed method sequential approach. The quantitative research objectives are carried out by starting with qualitative methods to inform later steps of quantitative research methods (pilot and large-scale surveys) (R. B. Johnson & Onwuegbuzie, 2004). Data collection and analysis will be carried out in three steps.

1- Phase I. Qualitative research of semi-structured phone interviews will provide insights and understanding of real-world experience from practitioners and academics in Big Data adoption. This will feed into the next phase by providing a preliminary version of the pilot survey.

2- Phase II. Mixed Method research of pilot survey will provide qualitative feedback and some limited quantitative data to test. This information will feed into the development of the next phase.
3- Phase III. Quantitative research of the final quantitative large-scale survey will confirm the hypothesized adoption factors.

These phases will be explained further in this chapter.

These steps are summarized in Figure-14.

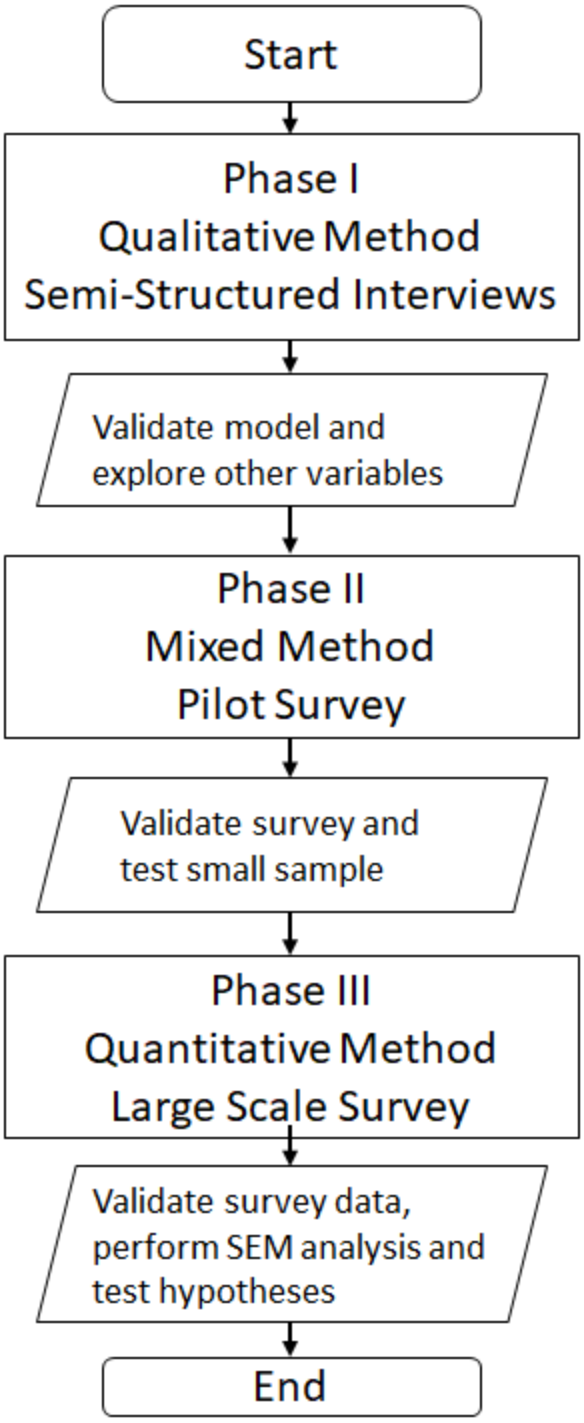

Figure 14 Research Flow Chart

3.2 Phase I. Qualitative Research Data Collection (Semi-Structured Phone Interviews)

3.2.1 Phase I Objectives

1- Validate Research Model

Qualitative research phase will be used to validate the research model of the selected Big Data adoption factors. Research collaboration with subject matter experts and practitioners to formulate, validate and explore research questions is advocated by Wagner in 1997 and others (Amabile et al., 2001; Parthasarathy, 2017; Wagner, 1997). There are multiple possible interaction methods to collaborate in research partnership. Phone or web conferencing interviews with industry practitioners is one of the recommended methods. Especially with COVID-19 challenge, web conferencing tools are becoming more popular and widely used. It provides audio, video and content sharing capabilities which allow for better collaboration and potentially more insight to the interviewer. Interviews allow you to gain in depth and real world understanding of the research in question.

These interviews in with data collected in literature review will validate the proposed factors

2- Capture Insights of Adoption Factors using Interpretive Phenomenological Analysis.

The interpretive phenomenological analysis (IPA) is one of the most "participant-oriented" qualitative research approaches (Alase, 2017). IPA captures the participant's opinion, perception and insights on the subject matter based on their real-world experiences. The researcher plays the role of interpreter of the participant's 'sense-making' of the experience based on the collective subjective data (Smith, Flowers, & Larkin, 2009). Since this method's unit of analysis is an individual, it is also called "idiographic approach" (Brocki & Wearden, 2006).

IPA is concerned with capturing the understanding, evaluation, and what each participant considers an important aspect of the event and how they experience it (Alase, 2017; Brocki & Wearden, 2006). It captures what is important to each participant. The researcher in IPA captures each of the participants' answers in inquisitive fashion, looks for common themes in the participants' responses, analyzes the responses in the context of the research and places the responses in appropriate contexts or brackets. These units of meaning, brackets or contexts are compared across the participants. Then, the researcher placement is based on the

observation of all the responses, literature review and based on the familiarity of the subject matter. Based on that, the researcher will interpret the patterns of these responses of these similar lived experiences (Alase, 2017).

3.2.2 Phase I - Interview Methodology

Interviews is a research methodology that has been used widely in qualitative and quantitative research. Interviews can produce valuable insights and information (Alvesson, 2003). Interviews allow the elicitation of information, meaning of observed behaviors and artifacts. Interviews enable the researcher to gain real-world experience insights and the interviewees' assessment on the subject matter (DiCicco-Bloom & Crabtree, 2006).

This research will utilize semi-structured interviews. Semi-structured interviews will give the focus needed to ask central questions (structured) that are germane to the research, and at the same time allow open ended (unstructured) questions to explore other aspects that may not be planned. This pragmatic approach of combining structured and unstructured interviews provide the advantages of both types of interviews that are likely to provide answers to the pertinent questions of this research and provide additional insights that may not be known to the researcher (Leedy & Ormrod, 2016).

Interviews can be conducted in multiple ways. It can be conducted face to face, by telephone or by video conferencing technology. Each interview method has its own advantages and disadvantages. Face to face interviews have the most potential of gaining the interviewees cooperation. On the other hand, it is the most time consuming. Phone interviews have lower cost financially and in time but may have lower response rate. Video conferencing is a newer technology that is like phone interviews. It can be considered the midway point between face to face and phone interviews (Leedy & Ormrod, 2016). It also can be the most time efficient method with busy practitioners that are geographically dispersed. This research will utilize the video conferencing technology to conduct semi-structured interviews.

The interview follows guidelines provided by McCracken (McCracken, 1988). This includes the ethical standard protocol, pre-conditioning, and estimations. Presenting the interviewees with detailed explanation of the purpose, scope, and instructions of the research to establish the boundaries and prevent scope creep.

3.2.3 Phase I - Participants' Selection for IPA Interviews

Selecting and recruiting participants for interviews is difficult. There are ways to improve the participation (Mapstone, Elbourne, & Roberts, 2007). One of

the effective ways to improve participation in interviews studies is word of mouth (Parthasarathy, 2017; Van Hoye, Van Hooft, & Lievens, 2009). This is especially true for this specific research that is specific to a special population. Thus, this research will use word-of mouth recruitment where the information of the researcher is shared to other participants (Alase, 2017).

The participants in phase I are selected either from industry practitioners who participated in the adoption of Big Data with existing data storage in an organization. Or academics who specialize in data storage and/or big data. The selection of participants with this expertise will provide a check on face validity, content validity for the instruments of the research. The researcher will solicit participation in phase I using professional social media contacts.

The number of participants in IPA interviews should not be large (in the hundreds) because it may cause the loss of subtle inflection of meaning. Smaller number of participants in IPA is the current consensus. The number of participants mentioned is between 5 to 10 who experienced similar events (homogenous) (Alase, 2017). This range of participants approximates a saturation point where no new insight is gained of adding new participants. Thus, 10 subjects will be the goal of this research (Brocki & Wearden, 2006; Parthasarathy, 2017).

3.2.4 Phase I - Interview Questions

The interview starts with an introduction about the research (Qu & Dumay, 2011). Then, the interviewer will ask the cover biographical information about the interviewee, role, industry, their organization size, and location. Then the researchers will ask open-ended questions regarding the challenges faced in the Big Data adoption. Then the researcher will ask about the 9 identified factors in the research. These questions will be open ended on how these factors affected the Big Data adoption. There will also be follow up questions to explore aspects of these factors and the adoption decision. For the full question details see (Appendix A.)

3.2.5 Phase I - Interview Data Analysis

Qualitative Content Analysis (QCA) method will be used to analyze the interviews (Cho & Lee, 2014; Glaser & Strauss, 1967). QCA can be used as inductive and deductive approaches. In this research, QCA will be used deductively to confirm the current hypotheses. It will also inductively (the open-ended portions) to generate new research questions for future research. QCA strength is in providing meanings, explanations of the phenomenon and insights on the why behind it. QCA provides categorization of themes of the interviews and other communications. These themes and categories then can be classified and

studied from the multiple interviews. It can be used as quantitative and qualitative methods (in this research it will be used as qualitative). It involves the classification of answers to each question to offer insight on common themes, context, and concerns. This will be used for both unstructured and structured questions. QCA will assist in providing interpretation of the common factors and concerns in Big Data adoptions. QCA provides a theory grounded in data collected (Cho & Lee, 2014).

3.3 Phase II. (Pilot Questionnaire) Mixed Research Data Collection

3.3.1 Phase II Objectives

Based on phase I of qualitative method output, a small-scale questionnaire is created to further refine the research questions of Big Data adoption factors. Small scale questionnaire provides input from multiple dimensions and previews for the large-scale survey in phase III. It can identify ambiguities and unforeseen problems. It provides estimates of time cost and logistics that can be extrapolated to large scale survey. It helps refine format, language, and clarity. It also provides face validity of the survey. Small-scale questionnaire provides a small sample of

survey data to assist in preliminary data analysis (Viechtbauer et al., 2015). This minimizes the need for correction plans for the large-scale survey.

3.3.2 Phase II - Pilot Questionnaire Methodology

Recruiting participants when there are no immediate benefits. In addition, finding participants who are knowledgeable about the subject to provide the proper context feedback can be challenging without first screening them (Viechtbauer et al., 2015). Thus, using the participants from phase I where the participants have been vetted for knowledge on the subject can add value and reduce the cost and time of finding new qualified participants. The participants for phase I of will be invited first to participate in the pilot questionnaire. These individuals already have the background in the subject and knowledge of the focus of this research. Their feedback and comments will be used to ensure that the questionnaire language, content, and instructions are clear. Others will be recruited to complete the survey to achieve an acceptable sample size for a pilot survey.

3.3.3 Phase II - Pilot Questionnaire Sample Size

The sample size for pilot survey varies in literature. In the 2005 paper, Julious argues that 12 subjects per group is needed for pilot survey (Julious, 2005).

Isaac and Michael suggest 10-30 participants in the pilot survey (Isaac & Michael, 1997). While Connelly proposes the pilot survey sample size to be 10% of the large-scale sample size (Connelly, 2008). The large-scale sample size calculation will be discussed in the upcoming sections. The large-scale survey sample size is between 300-500 participants. Thus, that will require 30-50 participants for the pilot survey in this instance according to Connelly. For this research, the pilot survey will be recruiting the participants from phase I.

3.4 Phase III. (Large Scale Survey)

Quantitative Research Data Collection

3.4.1 Phase III Objectives

The final phase of the data collection is conducting a large-scale survey. The questionnaire part of the survey is iteratively developed in the last 2 phases. The survey design is aimed to collect information regarding Big Data adoption factors in organizations with data storage systems. The questionnaire will be presented to many respondents. It captures the basic biographical information of the respondents and multiple-choice questions regarding Big Data adoption factors based on DOI and TOE in organizations with data storage systems. This phase is intended to capture enough responses that have adequate quantitative data to

achieve statistical significance to determine which factors are significant. In addition, the data captures which factors promote or prohibit the Big Data adoption for organizations with data storage systems.

3.4.2 Phase III - Survey Methodology

Online surveys have multiple advantages. It can reach a large sample in a shorter time (Wright, 2005). Many individuals, especially in the IT industry, are connected using the internet. That makes it a ubiquitous medium that IT practitioners and academics use frequently. Thus, it provides a wider reach, shorter response time, reduced geographical constraints and more economical compared to other survey modes (Simsek & Veiga, 2001). Online surveys on mobile devices have enabled an even wider reach, more accessibility and a faster response (Dillman, Smyth, & Christian, 2014). Web based surveys flow logic can enforce mandatory fields and adapt the survey flow to the specific responses.

This survey's questionnaire, delivery and data collection will be done using Qualtrics survey software tool that is provided by DePaul University. Qualtrics allows for an online delivery using a web-link and the capture of respondents' data online. It also allows for some data analysis, but other statistical methods and software will be used.

3.4.3 Phase III - Survey Data Collection Procedures

Survey's items were captured from published literature. In phase II these items will be reviewed by industry experts and academics. This will provide a face validity and content validity check of the survey's instruments. Upon completion of phase II, the survey's items will be included in the phase III large scale survey in Qualtrics survey company.

The population for phase III large scale survey will be elicited through a word-of-mouth recruitment strategy, asking participant to distribute the research literature and researcher contact info, reaching to social and professional networks of Big Data and through Qualtrics provided sample of the target audience. Qualtrics provides anonymous randomized sample for Big Data professionals through their panel aggregators.

The large-scale survey will be shared as a weblink. The landing page of that link will have a brief description of the research and its goals. It will also inform the respondent about the privacy protocol of this research and that no identifiable information will be shared. In addition to the voluntary nature of this study where participants can choose to abandon the survey. Also, it will qualify the respondent for the DePaul required information sheet that includes checks to be

over 18 years of age, speaks English, and has familiarity with Big Data solutions. The survey will ask respondents to voluntarily provide their email to have a chance to win a drawing and to share the results of this research. Respondents who do not meet the qualification above or choose not to complete the survey will be considered as nonresponses. Multiple tools will be used in this analysis. For the survey data collection, Qualtrics is used to create, distribute, capture, and analyze survey responses. Qualtrics is one of the leading survey platforms and is approved by DePaul University and available to its students and staff. SPSS will be used to perform missing data analysis and corrections when needed. For PLS-SEM, SmartPLS 3 is chosen since it is one of the more popular statistical tools among researchers.

3.4.4 Phase III - Participants' Selection for Survey

This research focuses on the evaluation of Big Data adoption factors according to individuals working within organizations that have data storage systems and intending to implement Big Data systems. In addition to that population, academics that are interested in this subject. Thus, the population targeted in this survey has certain knowledge and background. These individuals can be executives, managers, software engineers, IT professionals, storage

professionals, data scientists, consultants, project managers, academics among the roles that deal directly with adoption of Big Data.

There are organizations that deals with storage (Netapp, 2019) and Big Data like Storage Networking Industry Association (SNIA), Institute of Electrical and Electronic Engineers (IEEE), Association for Computing Machinery special interest group on management of Data (SIGMOD), Storage Management Interface Specification (SMI-S), Linux Foundation, OpenStack Foundation, Big Data Value Association (BVDA), Digital Analytics Association, Data Science Association, and Big Data & Analytics Association to name a few. There is no central repository of the population of the above professions who deal with aspects of storage and Big Data. Compiling a population of the above is also a daunting task that is not in the scope of this research. There is a high-level estimate of 2 to 3 million Big Data professionals worldwide, but these are based on secondary data (Qualtrics, 2020). There is no way to ensure that each group of the targeted population is represented in the sample. Thus, non-probability convenience sampling is preferred in this situation (Leedy & Ormrod, 2016).

3.4.5 Phase III - Calculating Minimum Sample Size for Large Scale Survey

There are multiple methods of calculating the minimum required sample size that will provide statistical power to examine the research hypotheses. The rule of thumb is the larger sample size results in higher power and better representation of the population. Conversely, limitation on time, budget and scope necessitates constraining the sample size to make the study feasible. This research will explore multiple methods of calculating the minimum sample size required and compare the value of each method. This research will use structural equation modeling (SEM). Specifically, this research will use Partial Least Square Structural Equation Modeling PLS-SEM (will discuss the rational later in this chapter.)

There are varying methods and thus varying recommendations for the minimum required sample size. One of the references cites to have 5 to 10 observations per observable variables with minimum of 300 participants (Comrey & Lee, 1992; Yong & Pearce, 2013). While in the published book "Using Multivariate Statistics", the range of sample size recommendations are: 50 very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good and 1,000 as excellent (Nguyen & Petersen, 2017; Tabachnick & Fidell, 2007).

Using Bartlett, et al. (2001) table to determine the minimum returned sample size requires determining few parameters (Bartlett, Kotrlik, & Higgins,

2001). The model contains categorical data. The assumption for this research will be the prevalent p-value of 0.5 and 95% confidence level the t value of 1.96 in Information Systems research. The minimum sample size needed from the table is 370. Another method is Naing et al., (Naing, Winn, & Rusli, 2006) with prevalence p-value of 0.5, precision d value of 0.05, the minimum required sample size is 384. Whereas Qualtrics online survey tool for sample size calculator (the same survey tool that will be used in conducting the survey) recommends 385 sample size based on 95% confidence interval, 5% margin of error and population of 3,00,000 (the higher limit of number of Big Data professionals) (https://www.qualtrics.com/blog/calculating-sample-size/) (Qualtrics, 2020). The consensus of the multiple methods above is coalescing around 380 respondents for a sample size. That is the minimum sample size that will be targeted for this large-scale survey.

3.4.6 Phase III - Validity of the Survey Instruments

Survey's validity is the extent that the instrument measures what it is intended to measure. Validity covers face validity, content validity, criterion validity and construct validity (Leedy & Ormrod, 2016). The following is the discussion of each of these aspects of validity and how it relates to this research.

Face validity is the extent which, on the surface, an instrument looks like it is measuring the intended characteristics (Leedy & Ormrod, 2016). Face validity for the survey instrument is checked in phase II of the research by asking the participants in the pilot survey, specifically participants in phase I of the semi structured interview who continued to phase II, if the survey is measuring what is intended. The researcher will follow up with emails and phone calls to ensure their feedback is captured.

Content validity is the extent that a measurement instrument is a representative sample of the domain being measured (Leedy & Ormrod, 2016). In other words, the extent that the questions are relevant to answer the research question. That can be answered by subject matter experts in both industry and academia (the participants from phase I and in phase II) can provide feedback if the questions are relevant to the research. Emails and phone calls to the participant will be conducted to collect their feedback.

Criterion validity is the extent that the results of a measure with another related measure (Leedy & Ormrod, 2016). Since many of these measurements are novel or scarcely studied, it will be difficult to relate to other measurements.

Construct validity is the extent that an instrument measures a characteristic that cannot be measured directly (latent variable) but can be measured through

some other measurable characteristics (observable variables). All the observable variables have been taken from existing literature (albeit they have been extended for this research's specifics.) For example, there has been multiple research papers on compatibility from technology perspective (K. Agrawal, 2015; H.-M. Chen et al., 2015; Esteves & Curto, 2013; Mahesh et al., 2018; Salleh & Janczewski, 2018; Verma & Bhattacharyya, 2017), but that has not been explored to more granular aspects of compatibility. However, (K. Agrawal, 2015; Lin, 2008) papers have studied compatibility as an observable variable with multi measurement items. In both papers, compatibility found to have a significant positive effect. This compatibility observable variable and its measurement are extended to explore compatibility of data storage latency, compute-ability, and interface with Big Data. Case and point, compatibility measurement item that was used in (K. Agrawal, 2015) "The changes introduced by BDA are consistent with existing practices" is modified in this research as "Big Data storage latency requirements are consistent with existing practices." All the measurement items used in this research and their original wordings can be found in appendix B.

3.4.7 Phase III - Reliability of the Survey Instruments

Reliability is the extent that an instrument is consistent with what it measures. In surveys, internal consistency reliability is considered a measure of the reliability where all items within a single construct produce similar results (Leedy & Ormrod, 2016). Cronbach's alpha coefficient is one of the most often used measurements for internal consistency reliability. A value of 0.7 or higher of Cronbach's alpha coefficient is considered to indicate high internal consistency (Streiner, 2003). The Cronbach's alpha coefficient will be used in phase III with the large-scale survey to test the reliability of the survey instruments and adjust the measurement if needed.

3.4.8 Phase III - Survey Questions

The survey's questionnaire will cover 5 main sections.

1- The first section will introduce the correspondents to the research interest, the motivation for the study, privacy notice and brief instructions on how to proceed with the survey.
2- The second section will focus on the biographical information of the survey's correspondents.

3- The Third section will capture the correspondents' data storage environment and if they have implemented or looking to implement a Big Data solution.
4- The fourth section will capture the correspondents' evaluation of the DOI/TOE Big Data factors construct identified from the previous chapter by answering their corresponding measurement item (Appendix B). These measurement items are reflective and not formative.
5- The fifth section will capture the correspondents' interest in receiving the results of this study when completed to their email addresses and a thank you for participating. See (Appendix C) for the full questionnaire.

3.5 Survey Data Analysis - Statistical Methods

3.5.1 Background on Selecting Statistical Technique

Quantitative data is provided by the large-scale survey. The survey's theory is based on TOE-DOI. Most of the variables are independent (adoption factors), whereas the dependent variable is the adoption intention. This type of data opens multiple statistical analysis possibilities. Multivariate statistical analysis is needed where multiple variables are analyzed simultaneously. Multivariate statistical methods can be categorized into primarily exploratory and primarily confirmatory

(although it is not a clear-cut distinction) (Afthanorhan, 2013; Joseph F Hair Jr, Hult, Ringle, & Sarstedt, 2016). Confirmatory approach is used when there are well established theories and concepts. Exploratory approach is used when there is no or little knowledge on how the variables are related (Joseph F Hair Jr et al., 2016).

First generation of multivariate analysis techniques were used through the 1980s. Statistical analyses such as cluster analysis, and exploratory factor analysis are considered as first-generation exploratory analysis. Analysis of variance, logistic regression, multiple regression and confirmatory factor analysis are considered first-generation exploratory analysis (Chin, 1998a; Joseph F Hair Jr et al., 2016). First generation multivariate analyses have multiple limitations that were overcome by the second-generation statistics. These include the inability to include latent variables measured indirectly by observable variables and limitation of incorporating measurement error in the model. There are other limitations as well (Chin, 1998b).

Second-generation multivariate analysis techniques started to gain popularity in the 1990s. They are collectively called Structural Equation Modeling (SEM). SEM enables the inclusion of latent variables that are measured by observed variables or items and their paths (structure). SEM is part of the logical positivist tradition (Akter, Fosso Wamba, & Dewan, 2017). IS research has largely adopted second generation SEM analysis (Gerow, Grover, Roberts, & Thatcher,

2010). The initial SEM that was used extensively earlier is Covariance Based SEM (CB-SEM) that is primarily confirmatory (Joseph F Hair Jr et al., 2016). The other type is Partial Least Squared SEM (PLS-SEM). PLS-SEM is used primarily exploratory and has gained popularity lately (Joseph F Hair, Risher, Sarstedt, & Ringle, 2019). SEM-PLS can be used for confirmatory testing as well (Afthanorhan, 2013; J. F. Hair, Ringle, & Sarstedt, 2014; Joseph F Hair et al., 2019; Lowry & Gaskin, 2014). Relatively new variant of PLS-SEM, that used in confirmatory testing is Consistent Partial Least Square SEM (PLSc-SEM) (Becker, Ringle, & Sarstedt, 2018; Theo K. Dijkstra & Schermelleh-Engel, 2014). Summary of the multivariate statistical analysis discussion can be summarized in figure 15 (from (Joseph F Hair Jr et al., 2016))

	Primarily Exploratory	**Primarily Confirmatory**
First-generation techniques	• Cluster analysis • Exploratory factor analysis • Multidimensional scaling	• Analysis of variance • Logistic regression • Multiple regression • Confirmatory factor analysis
Second-generation techniques	• Partial least squares structural equation modeling (PLS-SEM)	• Covariance-based structural equation modeling (CB-SEM)

Figure 15 Multivariate Statistical Analysis (Joseph F Hair Jr et al., 2016).

CB-SEM is based on covariance measurement only. Whereas PLS-SEM is based on total variance (composite). CB-SEM uses a common factor model where the analysis is based on common variance among the variables in the data (specific variance and error variance are ignored). PLS-SEM uses a composite model where all the variance (common, specific and error) in the analysis. The CB-SEM objective is to estimate the model parameters that minimize the difference between observed sample covariance matrix (that is calculated earlier) and the covariance matrix estimated after confirming the theoretical model. The PLS-SEM objective is to maximize the variance explained by dependent variables (Joe F. Hair Jr, Matthews, Matthews, & Sarstedt, 2017).

SEM includes 2 models. The first is the measurement model of the SEM or outer model. The second is the structural model, or inner model. Measurement models consist of observable variables (indicators, items, or manifest variables), latent variables (unobserved variables), and unidirectional arrow between each observed variable and latent variable. The structural model describes the relationships between the latent variables constructs of the model. The measurement and structural models are illustrated in figure 16 below (from (K. K.-K. Wong, 2013)).

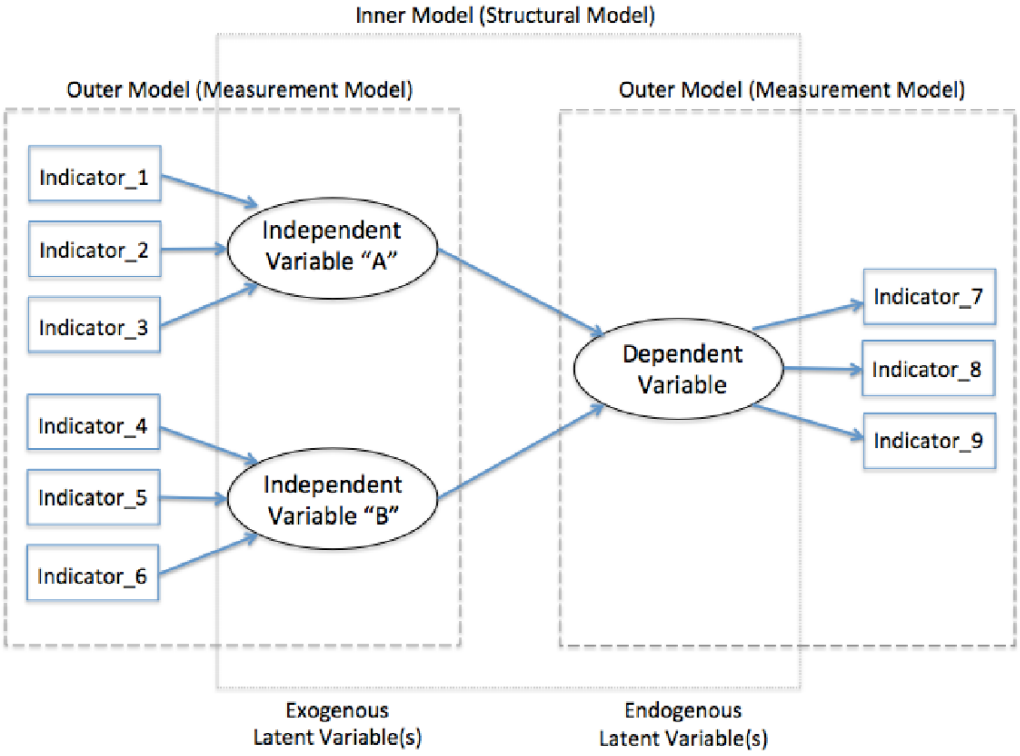

Figure 16 Measurement and Structural models (K. K.-K. Wong, 2013).

In the measurement model, observable variables are variables that are directly observed and measured (in this case with a survey instrument). Latent variables are variables that can only indirectly be measured through their effect or inferred relationships among observable variables (Leedy & Ormrod, 2016). There are 2 types of relationships between observable and latent variables. They are either reflective or formative. Reflective measurement model relation is where it is assumed that the latent variables "cause" the observable variables. Thus, this is

specified with an arrow pointing from latent variable to observable variable(s). Formative measurement model is where observable variables "form" or "cause" the latent variable. This is specified with an arrow from observable variable(s) to latent variable (Joseph F Hair Jr et al., 2016). Both SEM versions support reflective measurement models. Whereas, only PLS-SEM can perform formative measurement models (Joe F. Hair Jr et al., 2017).

SEM has also a structural model, or inner model. That is where parameters of the paths among latent variables (constructs) are calculated. Exogenous constructs (independent variables) explain other constructs. Thus, exogenous constructs have arrows pointing out of them. Exogenous constructs changes are generally not captured by the SEM model. The other is endogenous constructs (dependent variables) that are explained by other constructs. The arrows in this case are pointing toward the endogenous constructs (Joseph F Hair Jr et al., 2016). Exogenous constructs can affect other exogenous constructs as well. Linear relationship between latent variables is called recursive. Whereas circular latent variables relationship is non-recursive (Sarstedt, Ringle, & Hair, 2017).

SEM is the preferred choice for this type of research. The next question is which version of SEM will be better suited for this research. Both versions of SEM (PLS and CB) have different requirements and abilities. CB-SEM is a better suited explanation. Whereas PLS-SEM can perform explanation, prediction, and exploratory research. CB-SEM only captures common variance whereas PLS-SEM

factors in total variance. Both versions support reflective measurements whereas only PLS-SEM can accommodate formative measurements. CB-SEM does require normality of the data whereas PLS-SEM does not. CB-SEM requires a larger sample size than PLS-SEM. PLS-SEM achieves higher statistical power at all the sample sizes than CB-SEM (Joe F. Hair Jr et al., 2017).

This research has both elements of exploratory (extending existing Big Data adoption factors) and confirmatory (testing existing Big Data adoption factors.) In addition, all the observable variables are reflective and the causal effect in the structural model is recursive. PLS-SEM seeks to maximize the explained variance of the dependent latent variable and minimize the error. SEM-PLS is more robust especially when CB-SEM assumptions are violated (Joe F Hair, Ringle, & Sarstedt, 2011). Thus, PLS-SEM seems like a better fit for this research. Thus, the rest of this section will expand on the proposed use of PLS-SEM.

3.5.2 Sample Size

The sample size has been discussed earlier in details in the phase III large scale section under calculating minimum sample size for large scale survey. Based on the model with 9 adoption factors as independent latent variables and 1 dependent latent variable as the intention to adopt, the sample size should be ~ 400.

3.5.3 Missing, Anomalous and Outlier Survey Data

Missing data, suspicious response pattern, invalid, outliers, and unusual distribution in questionnaire survey data can cause issues for PLS-SEM. These data anomalies need to be examined and mitigated in certain situations (Joseph F Hair Jr et al., 2016). They are common problems for questionnaire survey data especially for online surveys. They can lead to difficulties in multivariate analysis specifically in PLS-SEM and can produce biased results (Durdyev, Ihtiyar, Banaitis, & Thurnell, 2018; J. F. Hair et al., 2014). There are detection and mitigation techniques for each of these issues.

Missing data is a common issue where respondents to the questionnaire survey may skip one or more questions. The missing data presents a balancing challenge between discarding respondent's data will reduce the sample size, while retaining the missing data or replacing it may skew the results (Joseph F Hair Jr et al., 2016). There are 2 aspects of missing data that needs to be detected and mitigated. From the respondent perspective and from observable variables perspective (item, indicator, question).

If the response from a respondent is missing 15% of the questionnaire, then this observation should be discarded. From the observable variable perspective, there is a need to make sure that there is no missing data that can be indicative of the unwillingness of respondents to answer that. This should be covered in phase II of the research with the pilot survey and feedback from its participants. This can

be a sensitive subject that some respondents are not willing to share their perspective on. This may need further investigation. If there is less than 5% missing data an observable variable, then use mean imputation (or median). Otherwise, case wise (listwise) deletion where the respondent data with that missing data is removed. A careful examination is needed to ensure that the case wise deletion does not discard a particular group of respondents (Joseph F Hair Jr et al., 2016). Others have suggested that using Maximum Likelihood Estimator (MLE) proved to produce better results than deletion (listwise or pairwise) (Allison, 2003; Enders & Bandalos, 2001).

Response data can be suspicious. Straight lining (where all the answers are the same), diagonal lining (where the responses follow diagonal line) and alternating extreme pole responses (where responses alternate between the minimum and maximum values) are examples of suspicious response data. These suspicious data responses are invalid and need to be removed from the data set.

Outliers is an extreme response to a question or extreme responses to all questions. Outliers are extreme values that can influence the results (Sarstedt & Mooi, 2014). Outliers can be caused by entry error. For example, a question with a Likert scale response from 1 to 7 has 9 as a response can be interpreted as an entry error since 9 is an invalid value for that scale. Another example is straight lining or Christmas trees where the responses appear visually like a straight line or zig zagging between extremes. Extreme values over the Likert scale are guarded with

Qualtrics software but the other types need human intervention per Qualtrics suggestions in the following link here https://www.qualtrics.com/uk/experience-management/research/survey-data-cleaning/. The range will also be tested in phase II of the research. An outlier can be just an extreme case of reality, but it is still valid. Another form of outliers is a combination of responses that are rare (multivariate outliers) (Lowry & Gaskin, 2014). The first step is to identify outliers (Joseph F Hair Jr et al., 2016).

The outlier detection can be classified as univariate outlier detection or bivariate outlier detection (Sarstedt & Mooi, 2014). Univariate outlier detection can be done by analyzing SPSS box plot data and identifying outliers' values that are 3 times the interquartile range (Joseph F Hair Jr et al., 2016; Sarstedt & Mooi, 2014). Bivariate outlier detection can be done using scatter plot between 2 observed variable data (Sarstedt & Mooi, 2014). Others have suggested using Mahalanobis squared distance D^2 to identify outliers (IBM, 2013).

After outliers are detected, a determination is needed to remove the outliers or not. If there is a valid explanation for that extreme value, then the data needs to be kept. One of the main issues removing outliers is the risk of discarding a valid subgroup. Subgroups can be identified from prior knowledge (ex. biographical profile) or latent class techniques on the collected data. Latent class techniques allow the researchers to identify and treat unobserved heterogeneity (Joseph F Hair Jr et al., 2016). For example, the survey data may contain subgroups that are more

concerned with different aspects of the survey questions. These subgroups will answer differently (individuals, groups, etc.). The inability to identify these subgroups can be a threat to validity (Joe F Hair Jr, Sarstedt, Matthews, & Ringle, 2016). These techniques identify the subsegments in the data and allow the division of data into multiple groups that can create a model for each subgroup. Examples of latent class techniques are FIMIX-PLS (Joseph F Hair Jr et al., 2016; Joe F Hair Jr et al., 2016), PLS-GAS and PLS-POS (Joseph F Hair Jr et al., 2016).

Although normal distribution is not required in PLS-SEM, substantially deviated from normal distribution can distort SEM analysis. Skewness and kurtosis are needed to examine the normality of the data. Skewness is the measure of the symmetry of the data distribution. Kurtosis is the measure of how peaked the data distribution is. Values close to zero for skewness and kurtosis indicate close to normal. While values that are greater than +1 or smaller than -1 indicate departure from normality (Joseph F Hair Jr et al., 2016).

3.5.4 Non-Response Bias

Although the online survey instrument provides quality check on the respondent, respondents may not be qualified to answer the survey questions correctly. The minimum requirement is that this survey should be completed by practitioners and academics who have or will be implementing Big Data solutions. On the other hand, one cannot stop participants to mis-represent their qualifications.

These factors can lead to respondents complete the survey hastily, incomplete or enter random responses. Non-response bias will introduce non-representative sample that can affect the sample frame thus affects the generalizability of the results (Dillman et al., 2014). To ensure the uniformity of the responses, statistical tests of independence like t-test or chi-squared test needs to be conducted. The sample size needed is around 400. Thus, the researcher will split the sample randomly into 2 separate groups containing 200 each then compare these responses (Mikalef & Pateli, 2017).

3.5.5 Statistical Techniques

SEM, as discussed earlier, consist of two models that work together. They are measurement model (outer model) and structural model (inner model). The models need to be examined in sequence to ensure each are meeting the required criteria. First, the researcher needs to assess the measurement model's reliability and validity. Second, the researcher needs to assess the structural mode. This process applies for both CB-SEM and PLS-SEM with differences. The following sections will explore measurement model evaluation then structural model evaluation for PLS-SEM with comparison to CB-SEM.

3.5.5.1 Measurement Model Evaluation

CB-SEM relies heavily on goodness of fit indices since it confirms existing theory. The fit indices measure the fit between hypothesized model and observed data. Each of these fit indices has cutoff criteria a researcher can determine if the observed data fit the model (the cutoff values are still being debated). There are multiple fit indices and researchers rely on plurality of them and not a single measure. The following are commonly used chi-square ($\chi 2$), Root Mean Square Error of Approximation (RMSEA), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Normed-Fit Index (NFI), Non Normed-Fit Index (NNFI), which also is called the Tucker Lewis Index (TLI) and Comparative Fit Index (CFI) (Afthanorhan, 2013; Ainur, Sayang, Jannoo, & Yap, 2017).

PLS-SEM, on the other hand, does not have a widely agreed on goodness of fit indices or cutoff criteria guidelines that have been comprehensively tested (Joseph F Hair et al., 2019). The efforts to develop goodness of fit indices are ongoing. PLS-SEM algorithm is based on maximizing the explained variance (R^2 value) in contrast to minimizing the divergence between the model and observed data. Coefficient of determination (R^2) measures the amount of variance in a dependent (endogenous) variable explained by the independent (exogenous) variables that have paths to it (Mathai, 2019). Coefficient of determination (R^2) will be explored further in the assessment of the structural model. Covariance

goodness of fit indices may not be completely transferable to variance based PLS-SEM.

PLS-SEM measurement model evaluation does have a distinction between reflective and formative measurement models. Reflective measurement model is where items are caused by the latent variable. This suggests a high degree of correlation among the items (Sarstedt et al., 2017). Whereas in the formative model where the items cause the latent variable, the degree of correlation among the items may not be high (Joe F Hair et al., 2011). Thus, items' reliability, discriminant and convergent validities can be measured for the reflective model. Formative measurement models have their own sets of measures of collinearity, significance of the outer weights (loadings) that need to be examined. Loadings that are not significant need to be removed from the model and will be discussed in further details.

3.5.5.2 Reflective Measurement Model Evaluation

Reflective measurement model relation is where it is assumed that the latent variables cause or effect the observable variables (indicators). Thus, this is specified with an arrow pointing from latent variable to observable variable(s)(Joseph F Hair Jr et al., 2016; Sarstedt et al., 2017). Observable items may have strong correlation if indicators are from the same domain (Sarstedt et al., 2017). Reflective measurement model is evaluated using indicator reliability,

internal consistency reliability, convergent validity and discriminant validities (Joseph F Hair et al., 2019; Sarstedt et al., 2017).

Indicator reliability is measured by indicator loading or outer loading. Outer loading is defined as each indicator (item) contribution to an assigned construct (Mathai, 2019). Indicator loading of 0.708 is recommended as it indicates that more than 50% of the indicator's variance is explained (Joseph F Hair et al., 2019). Indicator loadings that are less than 0.40 needs to be removed, and its impact to content validity needs to be examined. Loadings with values between 0.4 and less than 0.7, need to be examined to determine if removal of those loadings affect the internal consistency reliability negatively. This needs to be done one observable variable at a time. If the deletion increases the internal consistency reliability, then delete that observable variable with that loading and consider the impact on content validity. If the deletion of the observable variable with the low loading and does not increase the internal, then keep the observable variable (Joseph F Hair Jr et al., 2016).

Internal consistency reliability is the measure of multiple indicators to agree in measuring a latent variable (Sarstedt & Mooi, 2014). Internal consistency reliability for reflective measurement models is assessed using multiple measures. Cronbach's alpha is the most used measure for internal consistency reliability. Cronbach's alpha is a measure that provides an estimate of reliability based on the intercorrelation among observed variables (Joseph F Hair Jr et al., 2016).

Cronbach's alpha measures all items without their individual relative loading (weight). Cronbach's alpha measurement range is from 0 to 1. Higher values indicate higher levels of reliability. Acceptable range of Cronbach's alpha in internal consistency reliability starts at 0.60 for exploratory research and 0.70 for non-exploratory research. Maximum acceptable value is 0.95 to avoid indicator redundancy which lowers content validity. Cronbach's alpha produces lower values and less precise measurement values thus considered a lower bound value than composite reliability for internal consistency reliability measures (Joseph F Hair et al., 2019).

The second measure of internal consistency reliability is composite reliability. Composite reliability is a measure of reliability that calculates observable variables according to their outer loadings (Joseph F Hair Jr et al., 2016). Like Cronbach's alpha, Composite reliability value ranges from 0 to 1 and higher values indicate higher levels of reliability. Composite reliability calculates the weight of the indicator in its calculation thus produces higher results than Cronbach's alpha. Between the two values of Cronbach's alpha on the lower end and composite reliability of the higher end the internal consistency reliability is located. As an alternative, reliability coefficient ρ_a was introduced as a middle ground solution to capture more precisely the internal consistency reliability (Theo K Dijkstra & Henseler, 2015; Joseph F Hair et al., 2019). Internal consistency reliability value of over 0.95 is considered problematic since it indicates either some

items are redundant or other problems like straight lining. Bootstrapping confidence intervals can be used to check if the internal consistency is reliable as well (Joseph F Hair et al., 2019).

The next step is convergent validity of each construct measure. Convergent validity measures the extent that latent variables converge to explain the variance of its items (observable variables) which is measured by Average Variance Extracted (AVE). Acceptable minimum value for AVE of .50 or higher is acceptable that indicates the construct explains at least 50% of the variance of its items (Joseph F Hair et al., 2019).

Discriminant validity is explored next for the reflective measurement model which is the empirical extent that a construct is distinct from other constructs. Fornell-Larcker criterion, cross-loadings, and the newer HeteroTrait-MonoTrait (HTMT) ratio of correlations are measures of discriminant validity (Joe F. Hair Jr et al., 2017). Recent research however indicated that HTMT is better suited for discriminant validity (Joseph F Hair et al., 2019; Henseler, Ringle, & Sarstedt, 2015).

HTMT ratio of correlation is the mean value of items correlation across constructs relative to the geometric mean of average correlation of items measuring the same construct (Joseph F Hair et al., 2019). HTMT value ranges from 0 to 1 (Henseler et al., 2015). High HTMT value indicates low discriminant validity among constructs. For a conceptually similar construct, the maximum value of 0.9

is acceptable. Whereas, HTMT maximum value of 0.85 is acceptable for conceptually different constructs (Joseph F Hair et al., 2019).

3.5.5.3 Formative Measurement Model Evaluation

Formative measurement model is where observable variables (indicators) form or cause the latent variable. This is specified with an arrow from observable variable(s) to latent variable (Joseph F Hair Jr et al., 2016). There are 2 types of formative indicators: causal indicators and composite indicators. Causal indicators include error terms since there is an understanding that not all causes are captured and the error captures other "causes". Composite indicators are where the assumption that the indicators capture the model in full. Thus, the error from all other causes is set to zero. Indicators may not necessarily have high correlation since there are multiple contributing causes of the latent variable (Sarstedt et al., 2017).

PLS-SEM is the preferred method (compared to CB-SEM) for evaluating formative measurement models. Assessing formative measurement models in PLS-SEM includes convergent validity, indicator collinearity (or multi-collinearity if compared to multiple indicators), statistical significance and relevance of the indicators' weights.

Convergent validity is the extent a measure and its alternative are related. This can be done with redundancy analysis (Chin, 1998a; Joseph F Hair et al.,

2019). This is where formative measures and alternative reflective measures are compared in their correlation. The convergent validity needs to be at 0.7.0 or higher.

If there is a high correlation (thus collinearity) among formative measures, then some of the measures are redundant and may not be needed (not significant). That is unlike the formative measures which are considered interchangeable. Variance inflation factor (VIF) is often used to measure collinearity. VIF values of 5 and higher indicate collinearity issues. Collinearity issues can still exist for VIF value of 3. VIF value that is lower than 3 is recommended (Joseph F Hair et al., 2019). One can consider removing indicators if that solves the issue of collinearity taking into consideration the rest of the indicators capture the construct content (Joseph F Hair Jr et al., 2016).

The third step in formative measurement evaluation is outer weights, outer loadings, and each of their significance. Outer weights are the primary criterion to assess each indicator's relative importance in formative measurement models (relative contribution). Outer weights values range from -1 to 1. Higher or lower values indicate abnormal results. Zero outer weight value indicates a weak relationship between the indicator and the construct. Whereas outer weight values close to 1 or -1 indicate strong positive or negative relationship (Joseph F Hair et al., 2019).

Outer loading value determines each indicator's absolute contribution to its assigned construct (absolute contribution). Or it is the bivariate relationship between the indicator and the construct. The significance of outer loadings and outer weights are determined using bootstrapping from the collected data. Outer loadings of 0.5 and higher are considered significant (Joseph F Hair Jr et al., 2016; Mathai, 2019).

Statistical significance of formative indicator's coefficient outer weight and outer loadings are generated using nonparametric bootstrapping methods. The recommendation is either to use a Bias-Corrected and accelerated (BCa) bootstrap confidence interval when bootstrap distribution of indicator weights is skewed. Otherwise, percentile method to construct bootstrap-based confidence interval. If this confidence interval includes zero, then this indicates the outer weight is not statistically significant and can be a candidate for removal. Even in this case, outer loading value needs to be checked before removing that indicator. If outer loading is also not statistically significant then the indicator can be removed (Joseph F Hair et al., 2019).

3.5.6 Structural Model Evaluation

Structural model evaluation investigates the relationships between exogenous (independent) latent variables and endogenous (dependent) latent variables in the model. These relationships are evaluated for collinearity (VIF),

coefficient of determination (R^2), total effect, effect size (f^2), blindfolding-based cross validation redundancy measure or predictive relevance (Q^2), effect size (q^2), relevance and statistical significance of path coefficient, and out of sample predictive power using PLSpredict procedure that uses mean absolute error (MAE) and root mean squared error (RMSE) (Joseph F Hair et al., 2019).

Collinearity needs to be examined like formative measurement model evaluation. Collinearity in structural models can bias the regression results. The calculation in structural models uses the latent variable scores of the predictor constructs in partial regression to calculate VIF values. The VIF values cut off is the same as the formative measurement model recommendations where values over 5 indicate collinearity issues. VIF values of 3 to 5 can also have collinearity problems. VIF value should be less than 3 ideally. If collinearity is a problem, a common solution is to create a higher order model that can be supported by theory (Joseph F Hair et al., 2019).

Coefficient of determination (R^2) measures the variance that is measured in endogenous (dependent) variables explained by all the exogenous (independent) variables that have a path to it, therefore measures the model's explanatory power. R^2 is also known as (in-sample) predictive power. R^2 value ranges from 0 to 1 where higher value indicates a greater explanatory power. R^2 values of 0.75, 0.50, 0.25 are considered substantial, moderate, and weak respectively. High R^2 values

of 0.90 and higher are indicative of overfitting, where random noise is included in the R^2 (Joseph F Hair et al., 2019; Mathai, 2019).

Total effect also needs to be calculated to help in evaluating and constructing the full effect on another. Total effect is the sum of all the direct effects (standardized path coefficients) and indirect effect (effect of latent variable on a certain endogenous latent variable mediated through one or more additional latent variables) (Joseph F Hair et al., 2019; Mathai, 2019).

Since PLS-SEM is a dynamic process where independent latent variables effects are evaluated, some of these independent variables may be considered for removal as discussed earlier. Removal effect of independent latent construct on the dependent latent variable R^2 value needs to be evaluated using the f^2 effect size. The rule of thumb for the f^2 effect size value needs to be higher than 0.02, 0.15 and 0.35 to be regarded as small, moderate and strong respectively (Joseph F Hair et al., 2019; Mathai, 2019).

Predictive relevance or blindfolding-based cross validation redundancy measure (Q^2) is used to combine aspects of (out of sample) predictions and in sample explanatory power. It can be defined as the ability of the structural model to predict original observed values (Mathai, 2019). This measurement was first suggested by Stone and Geisser in 1974 and thus called Stone-Geisser's Q^2 value (Geisser, 1974; Stone, 1974). This is done by removing a single data point, imputes the removed point with the mean and then estimates the model parameters. Q^2

value should be greater than 0 for a specific endogenous construct to indicate predictive accuracy of the structural model for that construct. Q^2 values higher than 0, 0.25 and 0.50 indicate small, medium and large predictive relevance of the structural model (Joseph F Hair et al., 2019).

Effect size q^2 measures the predictive relevance of the inner (structural model) to the endogenous (dependent) latent variable. Effect size q^2 value needs to be higher than 0.02, 0.15 and 0.35 to be regarded as weak, moderate, and strong respectively. Negative or close to zero values indicate exogenous constructs are not relevant to the prediction of a given endogenous construct. It is calculated as follows:

$$q^2 = \frac{Q^2\ included\ -\ Q^2\ excluded}{1 - Q^2\ included}$$

(Joseph F Hair et al., 2019; Joseph F Hair Jr et al., 2016; Mathai, 2019).

Another option for measuring the predictive power of the structural model is to use PLSpredict procedure. Part of the sample (training sample) is used to generate the model then verify its results on the other part of the sample (analysis sample or holdout sample) (Galit Shmueli, Ray, Velasquez Estrada, & Chatla, 2016). This is also available in SmartPLS and R software (Joseph F Hair et al., 2019). PLSpredict runs k-fold cross validation. The value (k) is the number of equally divided subgroups from the total randomized sample. Once the sample subgroups are set, each subgroup is kept as a holdout sample then compared to the

model produced by the other subgroups. Then the process is repeated k times making each subgroup used as a holdout sample. The value of k is not set but recommended values are 5 or 10. Dividing the total sample into a random group can create a sample with extreme or abnormal values. It is recommended to run the k-fold process multiple times (Joseph F Hair et al., 2019).

There are multiple statistics that measure the predictive power of the PLSpredict process above. One measure is mean absolute error (MAE) that measures the average magnitude of error in prediction without considering the direction of the error. Also, there is the mean absolute percentage error (MAPE) where the prediction error is presented in percentage format. Another metric that is used is root mean square error (RMSE) which is the square root of the average squared root of the difference between prediction and actual observations. RMSE magnifies the larger errors because of its calculation whereas MAE and MAPE give equal weight to prediction errors. RMSE can give a more pessimistic evaluation of the prediction error while MAE and AMPE can be less sensitive to extreme predictive errors. Recent research suggests the combination of MAE and RMSE reliably select the model that balances model fit and predictive power. MAPE is not recommended. The focus when comparing these values is on the main endogenous construct (Joseph F Hair et al., 2019; Galit Shmueli et al., 2016).

Smaller values for MAE, MAPE, RMSE indicate higher prediction power. Since these measures are not scaled, the threshold for the predictive power becomes

arbitrary. (G. Shmueli et al., 2019) suggest using a naïve benchmark of simple indicator level average of the dependent variable from the training sample as a prediction of the variables in the holdout sample. This benchmark is like blindfolding based Q^2 that was discussed earlier thus calling this one as $Q^2_{predict}$. In other terms, $Q^2_{predict}$ is one minus the quotient of the PLS model's sum of squared prediction errors in relation to the mean value's sum of the squared prediction errors. Positive value of $Q^2_{predict}$ indicates prediction error is smaller than the prediction error given by the most (naïve) benchmarks (G. Shmueli et al., 2019).

Another measure using linear regression model (LM) to generate prediction for the observable variables of the dependent (endogenous) latent variables' indicators on the indicators of the independent (exogenous) latent variables in the PLS path model. This adds the PLS path model that $Q^2_{predict}$ does not consider. The PLS path model's predictive power is at least equal or greater than the naïve LM benchmark (G. Shmueli et al., 2019).

The method to interpret the PLSpredict results is to first evaluate the $Q^2_{predict}$ value. If the $Q^2_{predict}$ value is less than or equal to 0, then the predictive relevance is NOT confirmed. If the $Q^2_{predict}$ value is greater than 0, then we need to evaluate the predictive error. If the prediction errors are highly symmetrically distributed, then use RMSE since it will emphasize a higher degree of error. If prediction error is not highly symmetrically distributed, then use MAE. Then check each of the dependent indicators PLS-SEM value to LM. If none of the indicators PLS-SEM

value is smaller than LM, then the predictive relevance is not confirmed as well. Then depending on the relative number of indicators having PLS-SEM value less than LM we can determine the predictive power. If a minority of the dependent variable indicators have a value of PLS-SEM less than LM then, it has low predictive power. If the majority of the dependent variable indicators have a value of PLS-SEM less than LM then, it has medium predictive power. If all the dependent variable indicators have a value of PLS-SEM less than LM then, it has high predictive power. This can be summarized in Figure 16 from (G. Shmueli et al., 2019).

Figure 17 Guidelines for Interpreting PLSpredict Results (G. Shmueli et al., 2019).

Having confirmed the model's explanatory and predictive powers, the next step is to examine statistical significance and relevance of the path coefficients. Path coefficient is the hypothesized relationship value that ranges from -1 to +1.

Where the value of -1 indicates a strong negative effect on the independent latent variable on the dependent latent variable. Whereas +1 value indicates the strongest positive effect. Values close to zero indicate weak relationships. The significance of the path coefficient is evaluated using non-parametric bootstrapping technique. Significant level of 0.05 and lower, the t-statistics needs to be greater than or equal 1.96 (Mathai, 2019).

3.6 Proposed Model

The proposed model consists of 9 exogenous constructs that represent each of the independent variables. These variables are hypothesized to affect the Big Data adoption (dependent variable) that is represented as the endogenous construct. Each of these constructs has several indicators (observable variables).

Each of the following independent variables is hypothesized to be significant. Each of them is hypothesized to have positive or negative correlation with the dependent variable of Big Data adoption. The independent variables are divided into 3 groups per the TOE framework. From a technological perspective, data storage latency compatibility, data storage compute computability and data storage interface compatibility are hypothesized to be significant factors and have positive correlation to Big Data adoption at the organizational level. From an organizational perspective, open-source availability, enterprise source availability

and perceived cost of Big Data are hypothesized to be significant with positive correlation except that cost will have negative correlation. From an environmental perspective, perceived industry pressure, legislation barriers, and market turbulence are hypothesized to be significant and negatively correlated with Big Data adoption except for perceived industry pressure to be positively correlated. This can be summarized in Table 5.

Perspective	Big Data Adoption Factor	Correlation
Technology	Data Storage Latency Compatibility	Significant & Positive
Technology	Data Storage Compute Compatibility	Significant & Positive
Technology	Data Storage Interface Compatibility	Significant & Positive
Organization	Open-Source Software Availability of Big Data	Significant & Positive
Organization	Enterprise Source Software Availability of Big Data	Significant & Positive
Organization	Perceived Cost of Big Data	Significant & Negative
Environment	Perceived Industry Pressure	Significant & Positive
Environment	Legislation Barriers of Big Data	Significant & Negative
Environment	Market Turbulence of Big Data	Significant & Negative

Table 5 Hypothesized Big Data Adoption Factors and Their Correlations

The proposed model represented in SEM format is summarized in Figure 17 below.

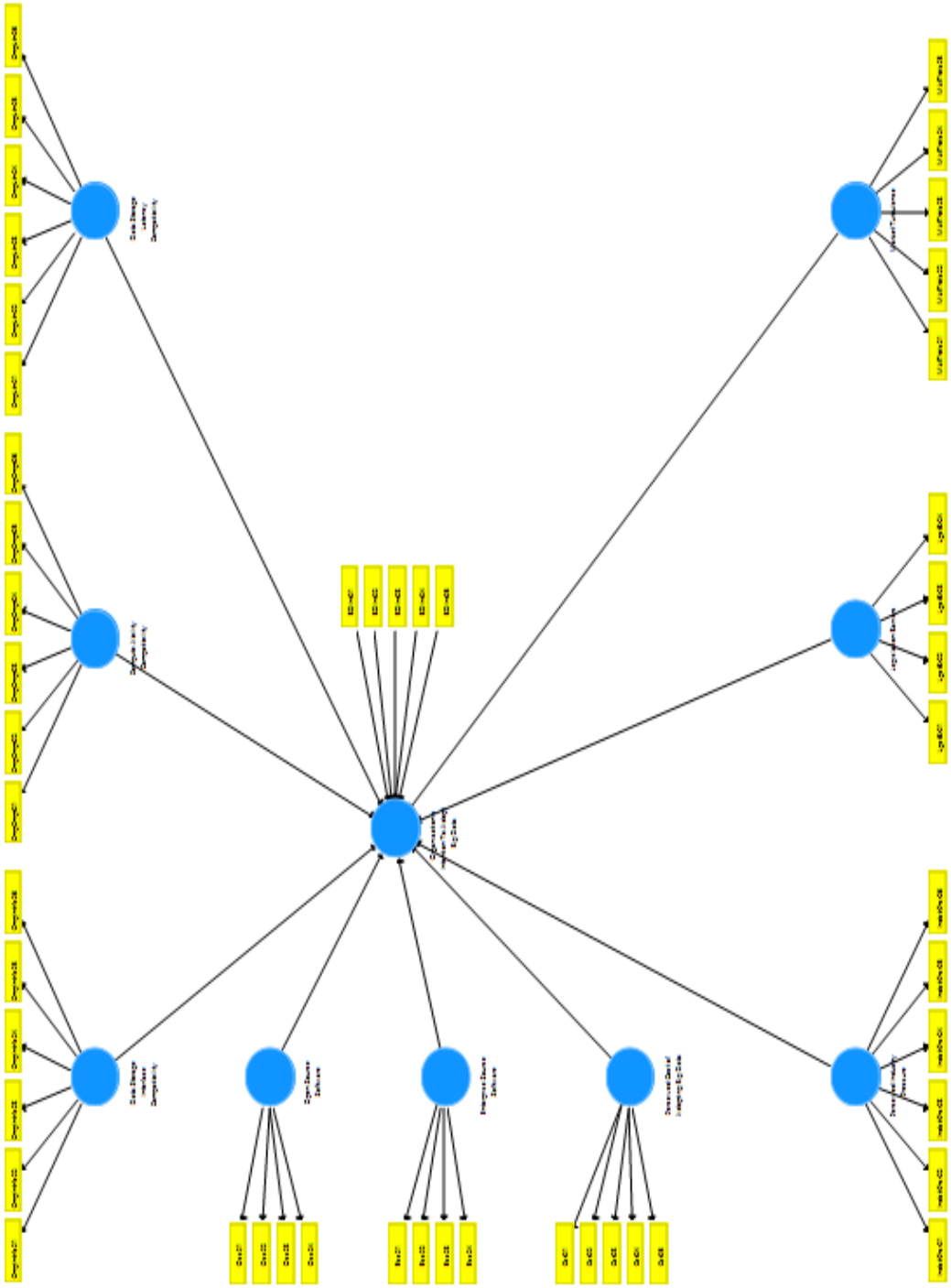

Figure 18 Proposed Model of Big Data Adoption Based on DOI theory and TOE Framework.

3.7 Institutional Review Board (IRB) Approval

This research involves interacting with human subjects in all the three stages of interviews, pilot survey and large-scale survey. This requires the researcher and the faculty sponsor to pass CITI training. The training is completed for both. This human interaction also requires a formal application and approval from DePaul Institutional Review Board (IRB). Since this research does not involve special subjects' categories and involves interview and survey methods, it qualifies for exempt review level (University, 2019). The IRB application is currently pending. The IRB letter of consent will be used to inform the participant of phase I and phase II. The IRB consent letter is captured in Appendix D. Participants in phase III of the large-scale survey will have similar but shorter wording in the survey's landing page.

3.8 Summary

This research is set to explore factors that affect the adoption decision of Big Data. These adoption factors are based on diffusion of innovation (DOI) theory

and technological, organizational, and environmental (TOE) framework. Based on literature and with the TOE framework, 9 adoption factors were identified that need further exploration. This research is based on the recommended mix method, where both qualitative and quantitative research methods are used. This pragmatic combination of diverse research methods aims at harnessing the strength of both methodologies and reducing their weaknesses (R. B. Johnson & Onwuegbuzie, 2004).

This research will be conducted into 3 phases. The first phase is qualitative research methods with semi structured interviews to gain insights from Big Data adopters and validate the adoption factors extracted from literature. The second phase is a mixed method where a pilot questionnaire is developed based on phase I. This pilot questionnaire with limited audience will ensure clarity and unforeseen problems and provide a preview of the data to be collected. Phase III of the research is quantitative where a large-scale questionnaire is sent to Big Data practitioners and academics to identify what are the significant factors that affect Big Data and if they affect positive or negative on the adoption decision.

Statistical methods on the phase III qualitative research method are selected to be structural equation modeling (SEM). The different types of SEM are explored with their advantages and disadvantages. The SEM partial least squared (SEM-PLS) method is chosen to test the significance of the identified factors.

CHAPTER 4: RESULTS OF THE STUDY

4.1 Overview

This research was conducted in three distinctive successive phases. The first phase is a qualitative method that uses a semi-structured interview with IPA to explore and validate then structured interview to validate Big Data adoption factors. Phase I utilizes interviews of individuals in organizations with data storage that adopted Big Data. The exploratory part applies semi-structured interviews with Interpretative phenomenological analysis (IPA). The structured interviews are used to confirm the importance of the identified factors based on the literature. Phase II is a mixed method of filling and reviewing the generated pilot survey to validate the proposed survey and test a small sample. Phase III is a quantitative large-scale survey to validate and test the hypotheses using SEM analysis. This chapter will present the data collected from these phases and the findings from each phase.

4.2 Phase I. Qualitative Research Data Collection (Semi-Structured Phone Interviews)

4.2.1 Phase I Demographics

IPA research is conducted on a small number of participants using purposeful sampling (Gauci, 2019; Smith & Shinebourne, 2012). The sampling focuses on selecting participants who lived similar experiences, which differs from other sampling methods (Smith & Shinebourne, 2012). The number of participants mentioned is between five and ten who experienced similar events (homogenous) (Alase, 2017). This range of participants approximates a saturation point where no new insight is gained from adding new participants (Brocki & Wearden, 2006; Parthasarathy, 2017). IPA involves a detailed analysis of each case and across cases; thus, having a large number of cases is difficult.

For this research, the purposeful sampling focused on individuals within organizations involved in the adoption decision and the implementation of Big Data on their existing data storage systems. Selecting and recruiting participants for interviews is difficult. One of the effective ways to improve participation in interview studies is word of mouth (Parthasarathy, 2017; Van Hoye et al., 2009).

Thus, this research used word-of-mouth recruitment to share the researcher's information with other participants (Alase, 2017).

In this study, nine individuals were interviewed using the IPA method who participated in adopting Big Data within their organizations' existing data storage systems. Seven of the nine participants worked in large organizations (>250 employees). One in medium-size organizations (50 to 250 employees) and one in small organizations (<50 employees). See Table 6.

No.	Role in Organization	Industry	Location	Org's Size
1	Founder	IT	USA	Medium
2	Engineer	IT	USA	Small
3	Manager	Automotive	USA	Large
4	Architect	Insurance	USA	Large
5	Architect	IT	USA	Large
6	Manager	Cloud	USA	Large
7	Architect	Media and Entertainment	USA	Large
8	Architect	IT	USA	Large
9	Manager	Cloud	USA	Large

Table 6 Phase I Participants' Information

4.2.2 Phase I Data Analysis and Results

4.2.2.1 Exploratory Semi-Structured IPA Analysis

IPA data analysis consists of six steps. The first step is, to read and re-read each transcript multiple times. The goal is to immerse oneself in the data. One of the goals of this step is to make the participant's voice the focus of the analysis. Another goal is to slow down the analysis and allow more absorption and reflection. It also provides for the ordering of ideas as the researcher moves forward (Smith et al., 2009).

The second step is the initial noting of the transcript. The researcher keeps an open mind and explores and notes the text of the transcript with a focus on noting anything of interest. The goal is to produce a comprehensive note (not word by word but by relevance). This step can be mainly descriptive and describe the participant's explicit meaning, but the focus will change in the following steps. The notes at this step can be descriptive, linguistic, or conceptual. This free textual analysis can be one of the most time-consuming steps (Smith et al., 2009). The researcher used NVivo 12 qualitative analysis software for this step and the rest of the data analysis steps (Brown, Smith, Arduengo, & Taylor, 2016).

The third step is to develop emergent themes for each of the transcripts. The researcher uses more notes taken in the previous step to find an emergent theme. If done correctly, the notes should closely relate to the main text. This is where the "Interoperative" part of IPA is applied. The researcher uses the parts of the

interview scripts to develop a theme that can group these ideas together. The researcher can be described as doing "double hermeneutics" because they are trying to make sense of the participant trying to make sense of the experience. This is done by producing concise statements of what is essential in a direct line with the statements of the transcript (Smith et al., 2009).

The fourth step is to search for connections across emergent themes. Since not all ideas in an interview are chronological, there is a need to discover themes that can group these ideas. Not all emergent themes need to be incorporated. The goal is to connect the emergent themes to point to the participants' most interesting and compelling accounts. Move the themes around and see if any clustering or connections can be drawn, even if it is not chronological. These themes can also be connected using abstraction, subsumption, polarization, contextualization, numeration, or function (Smith et al., 2009).

The fifth step is to repeat that process on the next participant's transcript (Smith et al., 2009). The last step is to look for patterns across cases. This is where the researcher explores common themes across all participants. This is where one can find which themes are more common, potent, or mentioned (Rivituso, 2014; Smith et al., 2009).

4.2.2.2 Structured Interview Data Analysis

Since the structured interviews are captured on a five-point Likert scale, descriptive statistics are used. It was encoded as follows for each factor (-2 Very Insignificant, -1 Insignificant, 0 Neutral, 1 Significant, 2 Very Significant.) Then the count was tabulated in Table 7.

Likert Value	-2	-1	0	1	2	
Adoption Factors	**Very Insignificant**	**Insignificant**	**Neutral**	**Significant**	**Very Significant**	**Total**
Data storage latency compatibility		1		7	1	9
Ability to compute large amounts of data compatibility		1	2	6		9
Data storage interface compatibility			1	6	2	9
Open-source software				7	2	9
Enterprise sourced software				6	3	9
Cost		1		6	2	9
Perceived industry pressure		1	2	5	1	9
Legislation barriers		2		6	1	9
Market turbulence		1	1	6	1	9

Table 7 Phase I – Structured Interview Summary

Descriptive statistics then was applied (count, min, and max). See Table 8.

Adoption Factor	Count (N)	Min	Max
Data storage latency compatibility	9	-1	2
Ability to compute large amounts of data compatibility	9	-1	1
Data storage interface compatibility	9	0	2
Open-source software	9	1	2
Enterprise sourced software	9	1	2
Cost	9	-1	2
Perceived industry pressure	9	-1	2
Legislation barriers	9	-1	2
Market turbulence	9	-1	2

Table 8 Phase I – Structured Interview Descriptive Statistics

4.2.2.3 Phase I Data Validity

Exploratory Semi-Structured IPA Validity

Qualitative research has a different set of evaluation criteria than quantitative research. One qualitative research validity method is by Yardley (Yardley, 2000). IPA research methodology extends Yardley's qualitative research validity work (Smith et al., 2009; Yardley, 2000). Others have used other

qualitative validity approaches (Creswell & Miller, 2000; Rivituso, 2014). Yardley's validity evaluation is based on four principles: sensitivity to context, commitment and rigor, transparency, coherence, impact, and importance.

Per Yardley's work, sensitivity to context can be described as the researcher's "awareness of different perspectives and complex arguments that can be brought to bear on the subject provide the researcher with the scholastic tools to develop a more profound and far-reaching analysis" (Yardley, 2000). This research is part of a larger dissertation thesis that thoroughly examined the literature, current and previously used methodologies, participants, interview questions, and reporting for the Big Data adoption. That is also demonstrated in the purposeful sampling that pursued individuals who participated in Big Data adoption and captured their lived experiences individually (Smith et al., 2009).

Commitment and rigor are the other criteria for evaluating validity. Commitment can be described as the "degree of attentiveness to the participant during data collection and the care with which the analysis of each case is carried out" (Smith et al., 2009). Close attention was given to all aspects of the interview and data analysis. The interview transcripts were given to the participants for review, and no major issues were found. The data presented include the perspectives of all participants. On the other hand, Rigor refers to the "thoroughness of the study" (Smith et al., 2009). This is manifested in selecting the sample, questions, interviews, and completeness of the analysis. The sample

was chosen carefully to represent instances where Big Data was adopted in various industries, industries, and job titles. The interview was semi-structured and interactive, where the researcher asked probing questions and asked for clarification and characterization from the participants. The analysis was ideographic, where each interview script was read, noted, and went beyond description to interpret and highlight the important aspects of individual interviews and the shared themes.

Transparency and coherence are described as clarity and cogency of the persuasiveness of the description and arguments presented (Yardley, 2000). Transparency is the clarity in which the IPA method's steps, participant selection, interview, and data analysis are given. Coherence is a way to describe how the arguments are received and how the research adheres to the method's principal.

Impact and importance are the final principal of Yardley's qualitative research validity. The research should leave the reader with the main themes and conclusions that enrich and influence. Previous studies have presented empirical and theoretical work; this study explores the inductive human element of Big Data adoption. The experiences of individuals who adopted Big Data captured in this research can enlighten the human perceptions of these adoption factors. The intent is for these insights to inform and assist IT practitioners and academics in what the factors are and how they can address them to enable further adoption of Big Data.

Structured Interviews Validity

The structured interview methodology is also qualitative and follows the validity and reliability of case studies (Yin, 2003). Construct validity can be achieved using multiple sources of evidence, which is the use of a literature review. Internal validity can be attained using pattern matching, which was done in the data analysis section. External validity can be accomplished by replicating multiple cases with various interviewees. Reliability can be achieved by defining the protocol and developing a case study database using the procedures above and the NVivo 12 software (Yin, 2003).

4.2.2.4 Phase I – Findings

Exploratory Semi-Structured IPA Findings

The Challenge of Big Data Value – New Insights

Big Data has been developed and used since 2005 (Melby, 2013); eight out of nine participants indicated Big Data value is a major adoption factor. Specifically, Big Data value perception is still a challenge. The main features of Big Data are data volume, data velocity, and data variety which was defined in 2001 (Russom, 2011). IDC added Value as another V (M. Chen et al., 2014; Gantz & Reinsel, 2011). Value is economically extracted from the other 3Vs. Value is a factor in adopting big data (Lamba & Dubey, 2015)

Perceived value remains a challenge to many, as 0ne participant stated: ***"The ability to link the data value to the business goals, I think, is probably one of the biggest challenges."***

Several other participants repeated this challenge: "***it's just some people don't see the actual value of the Big Data.", "A lot of smaller companies, where there will be a big value for them, it's difficult for them to see that."*** *and* ***"We have to convince our customers that our value proposition."***

That is echoed again with this participant's excerpt:

"The biggest factor in terms of why we adopt or not adopt [is] leadership doesn't quite see the overall benefit of doing that."

This challenge of perceiving the value of Big Data has put many organizations in a dilemma about whether to adopt Big Data. Big Data is still relevant technology with many organizations that implemented it and successfully extracted value (Eggers & Hein, 2020). Conversely, many organizations cannot discern how to extract value from Big Data. As one participant described

"A lot of companies decided to look and see. Is this something we can do with our data because there's always value in data and in processing it?"

Another participant articulated: ***"Managers, directors, VPs think it's a trendy topic, and they want to explore it, but they don't really know the benefit of it, they can't quantify it, and they have a hard time understanding why you need Big Data?"***

And another: ***"Until you actually run those jobs on the data. You don't know how much value you're going to get, so it's kind of a chicken and egg thing."***

And the question remains, ***"How do we link the benefits that we get from having to manage big data sets?"***

Some organizations attempt to be on the adoption path or do limited adoption but fail to realize the value (Barham, 2017). Some organizations are keen on storing data with the understanding the data has value. The following participants articulated that as follows: ***"They like to be able to collect more data because they feel like it might be valuable."***, ***"I think a lot of organizations have this idea that they would like to be able to extract some value from the data."***

Even with the data stored, is it in a form that is a value can be extracted? ***"The question is your data in the form [and] type Where you know [which] processing would give you the most value from it."***

Having data is necessary but not sufficient to build a viable Big Data (Data storage + data + ability to analyze) (Nguyen & Petersen, 2017). Other organizations started to query the data but with no clear use case. Having the ability to query the data opens the door to extracting value. Conversely, a longer, iterative journey is needed to refine these queries.

"They'll run queries. They'll be able to go to link datasets together with no problem, right? But then what comes out of it is not, as it needs a lot of refinement."

That journey of refinement to find value is unceasing

"You always have to prove the value. The thing that I always struggle with on the data science piece."

Hirsch compared Big Data to oil in terms of value and other aspects (Hirsch, 2014). This comparison can be extended to various uses and processes to achieve them. There are over 6,000 products made of petroleum (Abutu, 2014). Similarly, Big Data products/use cases can be as diverse as the organizations (Eggers & Hein, 2020). To realize Big Data value, identify the needs of the organization and design Big Data solutions that support that use cases.

The Challenge of Security (Old, New, and Unique)

Eight of nine participants mentioned security as an essential Big Data adoption factor consistent with the literature (Motau & Kalema, 2016; Nguyen & Petersen, 2017). As one participant stated, "***Concern number one was security.***" Other participants echoed similar sentiments like "***There's a lot of focus on security***" and "***There can be reservation from the perspective of the security aspect.***"

The participants for security, however, raise some novel points. One issue is that Big Data can access private information or violate the security policy. Structured data (ex., databases, or tables) can be secured by limiting access to the type of data to be analyzed. Big Data can access semi-structured or non-structured,

where secure access can be more obscured. That can be problematic as one participant described, "***I get a Social Security in a table then you can easily classify. Imagine if the data is buried in a Word document.***"

Analyzed data can contain restricted access information that may not be correctly classified. As one participant explained, ***"There are some customer, external facing system that is going to interact with that. That is where security comes in."***

Big Data can interact with data-at-rest (data locality) (Xiaoqiang et al., 2017) or move data and deal with data-in-transit (Y. Liu & Katramatos, 2019). This requires software tools that ensure correct access, encryption, cryptography, and others to support security. As a participant portrayed it, "***Especially if it's dealing with any security algorithms or encryption or cryptography, they look for meeting certain standards.***"

The Burden of Regulations

Regulation is a major factor affecting Big Data adoption (K. Agrawal, 2015; Bremser, 2018; Mahesh et al., 2018; Sun et al., 2018). One participant described Big Data as a liability in light of regulations "***Just the existence of Big Data is almost in some ways a liability. Especially if you're talking about GDPR.***" One aspect that the participant repeated is the inconsistencies across regulations. One participant asserted, "***The data may be proprietary or may have different privacy***

regulations governing it." Another participant confirmed the sentiment: "***We cannot have any information that is customer identifier on a record for more than 45 days.***" That participant chose the 45 days as the lower limit to simplify his requirements (instead of having different retention times for different data).

Participants found regulations hard to implement as one participant confirmed, "***Data retention [from] Legal sort of compliance was a big [and] very difficult to get.***" That difficulty is increased for consumer data. "***There is a lot more regulation now as far as the consumer data***." In addition to interfacing with external entities, "***There are certain legal aspects of it because we have a lot of external data.***"

Big Data Needs Big Network (More Refined Compatibility Is Needed)

One of Big Data's main features is processing large volumes of data (Russom, 2011). Some data needs to be transferred even with data locality (Xiaoqiang et al., 2017). This transfer of data adds additional utilization to the network bandwidth. That utilization can be significant. One participant pointed out that "**That Switch when you're processing terabytes or petabytes of data, then transferring the data from the storage system to the processing servers becomes the bottleneck.**"

Another participant reiterated that concern "**As far as dealing with my customers, [the] biggest factor for them on the technical side has always been**

bandwidth." Another mentioned, "**The number of pipes that are required to put data into a shared storage are going to be problematic.**"

Higher bandwidth utilization can turn into longer processing time "**To download and then do analysis on that data. That can be very time consuming.**" And that is not only for Big Data jobs but other applications running on the same data storage systems, as one participant stressed to ensure that "**allocation for that bandwidth [is] not going to hinder any of the other part of their apps.**" This adoption factor is not well studied in the literature.

Structured Interviews Findings

The following factors based on the exploratory structured interviews seem to be significant and need to be investigated further in future research:

Data storage system latency compatibility – New Factor

This factor has not been studied before. Eight of the nine participants report this as a significant factor in adopting Big Data.

The data storage system's ability to compute a large amount of data – New Factor

This factor has not been studied before. Six of the nine participants report this as a significant factor in adopting Big Data.

Data storage system interface compatibility – New Factor

This factor has not been studied before. Eight of the nine participants report this as a significant factor in adopting Big Data.

Open-source software – New Factor

This factor has not been studied before. All the nine participants report this as a significant or very significant factor in adopting Big Data.

Enterprise sourced software – New Factor.

This factor has not been studied before. All the nine participants report this as a significant or very significant factor in adopting Big Data.

Cost

Eight out of the nine participants report this as a significant factor in adopting Big Data.

Perceived industry pressure

Eight out of the nine participants report this as a significant factor in adopting Big Data.

Legislation barriers

Seven out of the nine participants report this as a significant factor in adopting Big Data.

Market turbulence

Seven out of the nine participants report this as a significant factor in adopting Big Data.

4.3 Phase II. (Pilot Questionnaire) Mixed Research Data Collection

4.3.1 Phase II Demographics

In phase II, a survey is created as a pilot survey with a limited audience. The goals of the pilot survey are to test the survey on a small-scale audience, solicit the input on clarity, language use, face validity of the questions and provide estimates for the logistics of running the survey. The same participants for phase I participated in phase II.

4.3.2 Phase II Results

Participants in phase II provided input on multiple aspects of the pilot survey incorporated in the phase III survey.

1- Landing page Input

Participants provided feedback to change the reference style, including the references on the page, word choice improvements, instructions on how to fill in the survey (pointing to the arrow), and instructions on how to fill in questions in case of not sure (to choose neutral), add IRB link and updated the time estimates of completion to 15 minutes. All these changes are incorporated in the phase III survey. Table 9 below highlights the changes included on the landing page of the pilot survey. Red text highlights the changes.

Phase II (Initial)	Phase III (Final)
Empirical Assessment of Big Data Technology Adoption Factors for Organizations with Data Storage Systems	Big Data Adoption Factors for Organizations with Data Storage Systems
What?	What?
This research explores factors that affect the adoption of Big Data technology in organizations with data storage systems.	This research explores factors that affect the adoption of Big Data technology in organizations with data storage systems.
Motivation	Motivation
An estimated 9.3 Zettabytes of data was created, captured, and replicated in 2016 (1ZB = ~1 billion Terabytes) (Westervelt, 2017). Data storage systems are evolving (Yianilos & Sobti, 2001) but not all of them are able to adopt Big Data (Ajimoko, 2017; H.-M. Chen, Kazman, & Matthes, 2015; Dubey, Gunasekaran, Childe, Wamba, & Papadopoulos, 2016). Big Data enables further insight and value from stored data. Big Data are data storage systems with the additional ability to analyze large volumes, with fast velocity and large variety of data (Russom, 2011). Most organizations have not adopted Big Data in production. Only around 13% of companies put Big Data into production use and Big Data projects had failure rates of 55% (H.-M. Chen et al., 2015). There is a need to identify the key determinants affecting Big Data adoptions and identify their multiple contexts of adoption (de Camargo Fiorini, Seles,	An estimated 9.3 Zettabytes of data was created, captured, and replicated in 2016 (1ZB = ~1 billion Terabytes) [1]. Data storage systems are evolving [2] but not all of them are able to adopt Big Data [3,4,5]. Big Data enables further insight and value from stored data. Big Data are data storage systems with the additional ability to analyze large volumes, with fast velocity and large variety of data [6]. However, most organizations have not yet adopted Big Data in production. Only around 13% of companies put Big Data into production use and Big Data projects had failure rates of 55% [4]. There is a need to identify the key determinants affecting Big Data adoptions [7]. This survey is part of research that address factors that enables or hinders Big Data adoption for organizations with data storage systems who have implemented or planning to implement Big Data.

Jabbour, Mariano, & de Sousa Jabbour, 2018).	
Requirements	Requirements
We are asking individuals to participate in the research who have work or academic experience in Big Data technology and data storage systems. You must be age 18 or older to be in this study. We hope to enroll up to 500 people into this phase of the research. This survey will take about 10 minutes to complete.	We are asking individuals to participate in the research who have work or academic experience in Big Data technology and data storage systems. Participants must be at least 18 years old in order to participate in this survey. The target is to enroll up to 500 people into this phase of the research. This survey will take about 15 minutes to complete.
In Scope: Organizations implementing their own Big Data solution on data storage system (on-premises or cloud).	Scope: Organizations implementing their own Big Data solution on existing data storage system (on-premises or cloud).
Out of Scope: Organizations buying Big Data feature from cloud providers.	Out of Scope: Organizations buying Big Data as a feature from cloud providers.
What's in it for me?	What's in it for me?
A raffle from the entered emails will be drawn and three winners will be awarded a brand-new Apple iPad Air https://www.apple.com/ipad-air/. The winners will be notified by email and will be sent to the address of his/her choosing free of charge.	A drawing from the entered emails will be drawn randomly. Three winners will be awarded a brand new Apple iPad Pro https://www.apple.com/ipad-pro/ each. The winners will be notified by email and the iPad Pro will be shipped to the address of his/her choosing free of charge.
Questions or Concerns	Questions or Concerns
Contact the principal investigator Ahmad Alnafoosi aalnafoo@mail.depaul.edu	Contact the principal researcher Ahmad Alnafoosi by email aalnafoo@depaul.edu.
What else?	What else?
The data collected from this survey will NOT have identifiable information except for your email (OPTIONAL).	The data collected from this survey will NOT have identifiable information except for your email (OPTIONAL).
Your email address will be removed for data analysis and WILL NOT BE	Your email address will be removed for data analysis and WILL NOT BE SHARED

SHARED. Your email address will only used by the researcher	WITH ANYONE. Your email address will only be used by the researcher to:
-to notify the winner of the raffle; and	-notify the winners of the drawing; and
-to share the survey results if requested.	-share the survey results if requested by the participant.
	Note: All questions must be answered. Should you feel an answer is Not Applicable, then select Neutral.
Fine Print	Fine Print
Please see the following document for more information regarding this survey.	Please see the following document for more information regarding this survey.
Institutional Review Board (IRB) Information Sheet	Institutional Review Board (IRB) Information Sheet
	https://1drv.ms/w/s!AkVCepUsHerVuWGe-w8EbefqRi94?e=qCIja1
By completing the survey you are indicating your agreement to be in the research.	By completing the survey you are indicating your agreement to be in the research.
	Please click the ARROW in the BOTTOM RIGHT of this page to continue.
	References
	[1] Westervelt, R. (2017). IDC White Paper: Information-Centric Security: Why Data Protection Is the Cornerstone of Modern Enterprise Security Programs, March 2017.
	[2] Yianilos, P. N., & Sobti, S. (2001). The evolving field of distributed storage. IEEE Internet Computing, 5(5), 35-39.
	[3] Ajimoko, O. J. (2017). Exploring the Cloud-Based Big Data Analytics Adoption Criteria for Small Business Enterprises. Colorado Technical University,
	[4] Chen, H.-M., Kazman, R., & Matthes, F. (2015). Demystifying big data adoption: Beyond IT fashion and relative advantage. Paper presented at the Proceedings of Pre-

	ICIS (International Conference on Information System) DIGIT workshop.
	[5] Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2016). The impact of big data on world-class sustainable manufacturing. The International Journal of Advanced Manufacturing Technology, 84(1-4), 631-645.
	[6] Russom, P. (2011). Big Data Analytics, TDWI best practices report. Fourth quarter, 1-35.
	[7] de Camargo Fiorini, P., Seles, B. M. R. P., Jabbour, C. J. C., Mariano, E. B., & de Sousa Jabbour, A. B. L. (2018). Management theory and big data literature: From a review to a research agenda. International Journal of Information Management, 43, 112-129.
	[8] Macharis, C., Lebeau, P., Van Mierlo, J., & Lebeau, K. (2013, November). Electric versus conventional vehicles for logistics: A total cost of ownership. In 2013 World Electric Vehicle Symposium and Exhibition (EVS27) (pp. 1-10). IEEE.

Table 9 Phase II – Pilot Survey Landing Page updates

2- Questions Updates

Participants also provided feedback on questions clarity and face validity and the feedback was incorporated in to phase III survey. Table 10 below highlights the changes incorporated in the survey questions/instruments from the pilot survey. Red text highlights the changes.

Phase II (Initial)	Phase III (Final)
Q2 - Does your organization have data storage system?	Q2 - Does your organization have a data storage system?

Q3 - Have your organization adopted or planning to adopt BD?	Q3 - Has your organization adopted BD?
	Q69 - Is your organization planning to adopt BD?
Q4 - BD storage latency requirements do not contradict the current internal Information Systems' applications.	Q4 - BD storage latency requirements align the current internal Information Systems' applications at my organization.
Q5 - BD storage latency requirements are supported by the existing Information Systems' infrastructure.	Q5 - BD storage latency requirements are supported by the existing Information Systems' infrastructure.
Q6 - BD storage latency requirements are supported by the organizational IT human resources.	Q6 - BD storage latency requirements are supported by the organizational IT department resources.
Q7 - My organization adopts BD open source software wherever possible.	Q7 - My organization adopts BD open source software wherever possible.
Q8 - Given a choice, my organization prefers to use BD open-source software in the near future.	Q8 - Given a choice, my organization prefers to use BD open-source software in the near future.
Q9 - My organization is likely to adopt BD open-source software in the near future.	Q9 - My organization is likely to adopt BD open-source software in the near future.
Q10 - BD is requested by important business partners.	Q10 - BD is requested by important business partners.
Q11 - BD is requested by majority of business partners.	Q11 - BD is requested by majority of business partners.
Q12 - Important competitors using or soon to be using BD.	Q12 - Important competitors using or soon to be using BD.
Q13 - Majority of competitors using or soon to be using BD.	Q13 - Majority of competitors using or soon to be using BD.
Q14 - Adoption of BD will be a strategic weapon of the organization to enhance the company’s competitive advantage.	Q14 - Adoption of BD will be a strategic goal of the organization to enhance the company’s competitive advantage.
Q15 - My organization will invest more resources (e.g., human, hardware, and finical resources) in the adoption of BD.	Q15 - My organization will invest more resources (e.g., human, hardware, and financial resources) in the adoption of BD.
Q16 - The adoption of BD will be important business strategy in the near future.	Q16 - The adoption of BD will be an important business strategy in the near future.

Q17 - BD computing requirements do not contradict the current internal Information Systems' applications.	Q17 - BD computing requirements align the current internal Information Systems' applications at my organization.
Q18 - BD computing requirements are supported by the existing Information Systems' infrastructure.	Q18 - BD computing requirements are supported by the existing Information Systems' infrastructure.
Q19 - BD computing requirements are supported by the organizational IT human resources.	Q19 - BD computing requirements are supported by the organizational IT department resources.
Q20 - My organization cannot afford the cost of adopting BD.	Q20 - My organization can not afford the cost of adopting BD.
Q21 - Adopting BD is expensive.	Q21 - Adopting BD is expensive.
Q22 - BD adoption can result in a high level of total cost of ownership in my organization.	Q22 - BD adoption can result in a high level of total cost of ownership in my organization.
Q23 - Business laws do not support BD.	Q23 - Business laws do not support BD.
Q24 - Inadequate legal protection for BD.	Q24 - There is inadequate legal protection for BD.
Q25 - Tight, inconsistent, or changing laws related to BD.	Q25 - Business laws related to BD are tight, inconsistent, or changing.
Q26 - BD storage interface requirements do not contradict the current internal Information Systems' applications.	Q26 - BD storage interface requirements align the current internal Information Systems' applications at my organization.
Q27 - BD storage interface requirements are supported by the existing Information Systems' infrastructure.	Q27 - BD storage interface requirements are supported by the existing Information Systems' infrastructure.
Q28 - BD storage interface requirements are supported by the organizational IT human resources.	Q28 - BD storage interface requirements are supported by the organizational IT department resources.
Q29 - My organization adopts BD enterprise source software wherever possible.	Q29 - My organization adopts BD enterprise source software wherever possible.
Q30 - Given a choice, my organization prefers to use BD enterprise source software in the near future.	Q30 - Given a choice, my organization prefers to use BD enterprise source software in the near future.
Q31 - My organization is likely to adopt BD enterprise source software in the near future.	Q31 - My organization is likely to adopt BD enterprise source software in the near future.

Q32 - Competition in our market is cutthroat.	Q32 - Competition in our market is cutthroat.
Q33 - BD in our industry is changing rapidly.	Q33 - BD in our industry is changing rapidly.
Q34 - Customers tend to look for new products all the time.	Q34 - Customers tend to look for new products all the time.
Q35 - What is your title?	Q35 - What is your title?
Q35_16_TEXT - Other. Please specify:	Q35_16_TEXT - Other. Please specify:
Q36 - What industry do you work in?	Q36 - What industry do you work in?
Q36_18_TEXT - Other. Please specify:	Q36_18_TEXT - Other. Please specify:
Q37 - How many employees are in your organization?	Q37 - How many employees are in your organization?
Q38 - Where are you located?	Q38 - Where are you located?

Table 10 Phase II – Pilot Survey Questions updates

4.4 Phase III. (Large Scale Survey)

Quantitative Research Data Collection

4.4.1 Phase III Demographics

Nine hundred eighty-one participants viewed the landing page of the survey. However, not all respondents completed the survey. Five hundred seventeen respondents completed all the Likert scale questions (survey questions measuring the adoption factors and adoption questions). Five hundred ten respondents also completed the demographic information (at the end of the survey). Twenty-two respondents completed 52% - 90% of the survey questions. Two hundred eleven participants completed 1% to 51% of the survey (which is not usable for PLS-SEM) (Joseph F Hair Jr et al., 2016). See Table 11.

No. of Likert Scale Questions Answered (31 Total)	Percentage of Likert Scale Questions Answered	No. of Respondents
31	100%	517
28	90%	3
25	81%	4
22	71%	6
19	61%	4
16	52%	5
15 or less	1% - 51 %	211
Did not answer any questions	0%	231

Table 11 Phase III – Survey Participants' Completion

Five hundred ten did complete the survey and the demographic information. Participants spanned the globe but were mainly in North America (45%), Asia (the Asia Pacific and South Asia) (41%), the Middle East (8%), Europe (4%), and smaller numbers in the rest of the world. Some participants chose others for location; eighteen referred to India, thus moving them to South Asia count and removing them from Other Count. See Table 12. Participants' organization size covered all the organization size ranges (large, medium, and small). The organization sizes represented are mainly large (62%), then medium (22%), and small is at (19%). See Table 13.

Location	Count	Percentage
North America	230	45.10%
Asia Pacific	127	24.90%
South Asia	82	16.08%
Middle East & Africa	39	7.65%
Europe	20	3.92%
Central and South America	6	1.18%
Other. Please specify	4	0.78%
Australia & New Zealand	2	0.39%
Grand Total	510	100.00%

Table 12 Phase III – Survey Participants' Locations

Organization Size	Count	Percentage
250 Employees or larger	314	61.57%
50 to 249	110	21.57%
Less than 50	86	16.86%
Grand Total	510	100.00%

Table 13 Phase III – Survey Participants' Organization Sizes

Participants' industry selections covered many industries. IT is the highest selection, accounting for (25%). Financial Services came second at (11%). Academia came third at (10%). Cloud services and Health care had the same number of participants at 8% each. Consulting came after that at (7%) then telecommunications at (5%). The rest of the industries came in at less than 5% of

the respondents. They are "Business Services, Retail, Wholesale, Distribution," Manufacturing, Media, Other, "Government Federal, State or Local," Storage Services, "Advertising / Marketing / PR," Transportation, Software, "Oil & Gas," social media, Non-Profit, "Travel and Leisure," and Utilities. Twenty-one respondents chose other for the industry. The following were notable enough to report as separate categories from the other category and removed from the other category count. The non-profit industry has two respondents. The oil & Gas industry has three respondents. The software industry has four respondents. See Table 14 for details.

Industry	Count	Percentage
IT	128	25.10%
Financial Services	58	11.37%

Academia/Education	52	10.20%
Cloud Services	43	8.43%
Health care	43	8.43%
Consulting	36	7.06%
Telecommunication	25	4.90%
Business Services	19	3.73%
Retail, Wholesale, Distribution	17	3.33%
Manufacturing	12	2.35%
Media	12	2.35%
Other. Please specify:	12	2.35%
Government Federal, State or Local	9	1.76%
Storage Services	9	1.76%
Advertising / Marketing / PR	8	1.57%
Transportation	7	1.37%
Software	4	0.78%
Oil & Gas	3	0.59%
Social Media	3	0.59%
Non-Profit	2	0.39%
Travel and Leisure	2	0.39%
Utilities	2	0.39%
Grand Total	510	100.00%

Table 14 Phase III – Survey Participants' Industries

Participants' professions covered a wide variety of professions related to Big Data. The top profession that responded to the survey were data engineers (29%). Followed by architects (9%), then consultants (8%), managers (7%), data scientists (7%), developers (6%), students (5%) then IT engineers (5%). Then multiple other professions at lower response rates like storage engineer, system integrator, quality assurance, VP, security engineer, Big Data admin, "CEO/Founder," director, network engineer, "CIO/CTO," and academics. The

following were notable enough to report as separate categories and removed from the other category count from the other category. Big Data admin title has eight respondents. The Director title has seven respondents. The academic title has four respondents. See Table 15 for further details.

Profession	Count	Percentage
Data Engineer	146	28.63%
Architect	48	9.41%
Consultant	39	7.65%
Manager	36	7.06%
Data Scientist	34	6.67%
Developer	33	6.47%
Student	27	5.29%
IT Engineer	25	4.90%
Other. Please specify:	22	4.31%
Storage Engineer	15	2.94%
System integrator	13	2.55%
Quality Assurance	12	2.35%
VP	11	2.16%
Security Engineer	9	1.76%
Bigdata Admin	8	1.57%
CEO/Founder	8	1.57%
Director	7	1.37%
Network Engineer	7	1.37%
CIO/CTO	6	1.18%
Academics	4	0.78%
Grand Total	510	100.00%

Table 15 Phase III – Survey Participants' Titles

4.4.2 Phase III Analysis

4.4.2.1 Phase III – Missing, Anomalous, Outliers Data, and Sample Evaluation

Five hundred seventeen completed surveys do not have any missing Likert scale data. The number of respondents who have 15% (Joseph F Hair Jr et al., 2016) or less missing data is only three respondents. Because of enough completed surveys and the small number of usable respondents, the 517 completed surveys will be used for the analysis. See Table 11 above.

The researcher detected six straight-lining anomalous data where the standard deviation for their answers was zero (indicating all their responses are the same). These respondents were deleted from the data, thus still having 511 respondents to investigate. The diagonal lining was also checked, and no response had that pattern repeated. The data was also checked for extreme pole responses; no response had that pattern repeated multiple times.

There are no outliers since all the Likert scale questions survey design forces the respondents to only select one of the following options and nothing else.

Strongly disagree (Value of -2)

Somewhat disagree (Value of -1)

Neutral (Value of 0)

Somewhat agree (Value of 1)

Strongly agree (Value of 2)

All the values were also examined using the MAX function. The maximum value was 2. The data was also examined with the MIN function. The minimum value was -2. That confirms that the data has no outliers.

The remaining 511 responses are randomized and then divided into two groups. The first group is the training set containing 411 responses (training Set). The second group is the hold-out set that includes 100 responses. This will address the observable measurement instruments' PLS-SEM analysis robustness by creating the model using the training set and holdout set to assess the predictive validity of the PLS path model (Hair J.F, 2012).

4.4.2.2 Phase III – Measurement Model Evaluation, Validity & Reliability

Survey instrument validity consists of four types: face validity, content validity, criterion validity, and construct validity (Leedy & Ormrod, 2016). Face validity is examined by the participants in phases I & II of the research to the extent that an instrument looks like it is measuring the intended characteristics. Content validity is examined in phases I & II of the research to the extent that a measurement instrument is a representative sample of the domain being measured. These participants provided feedback that was incorporated into the phase III survey. See Table 10 above for their contribution to the face and content validity of the survey. Criterion validity is difficult to examine since many of these measurements are novel or scarcely studied. Construct validity is the extent to which an instrument

measures a characteristic that cannot be measured directly but can be measured through other measurable characteristics. The observable measurement instruments taken from other research papers are discussed and refined by phases I & II feedback.

Survey data in PLS-SEM has two models. The first model is the measurement or an outer model that studies observable and latent variables. The second model is the structural or outer model that studies the relationships between the latent variables. First, the measurement model's (reflective model used in this research) reliability is studied by indicator reliability and internal consistency reliability. Validity is examined using convergent validity and discriminant validity.

PLS-SEM algorithm for the measurement model was run using Smart PLS (V 3.3.3). Initial runs suggested reliability issues for Cost and Legislative Barrier indicators (observable variables). The PLS-SEM algorithm was run with a factor weighting scheme of 500 maximum iterations and a stop criterion value of 7. PLS-SEM offers flexibility to address these potential issues. There were intermediate interactive steps in evaluating and modifying the measurement model to achieve validity and reliability of the model. These steps involved removing observable variables. These intermediate steps are detailed in Appendix H of this document. The following model will be called the final measurement model.

Evaluating model measurement model reliability and validity for the final model shows all reliability and validity measurements are within an acceptable range. Cronbach's alpha values are all within (0.6 – 0.95). Composite reliability values are all within (0.7 - 0.95). Finally, AVE values are all greater than 0.5. See Table 16.

Construct	Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
BD Adoption	0.799	0.882	0.714
Compute Compatibility	0.864	0.936	0.88
Cost	0.652	0.827	0.71
Enterprise Source SW	0.893	0.933	0.824
Industry Pressure	0.821	0.881	0.65
Latency Compatibility	0.848	0.906	0.763
Legislation Barrier	0.717	0.862	0.76
Market Turbulence	0.707	0.837	0.631
Open-Source SW	0.884	0.928	0.812
Storage Interface Compatibility	0.903	0.939	0.838

Table 16 Phase III – Final Model Construct Reliability and Validity

Checking the indicator reliability for the final measurement model for reliability/outer loading values showed all values are above the threshold. The lowest value is 0.702, which is higher than the 0.40 threshold. See Table 17.

Indicator	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
Adp1	0.897									
Adp2	0.801									
Adp3	0.833									
Cmp1		0.947								
Cmp2		0.929								
Cst2			0.702							
Cst3			0.963							
ESS1				0.895						
ESS2				0.921						
ESS3				0.907						
IPr1					0.781					
IPr2					0.803					
IPr3					0.811					
IPr4					0.831					
Inf1										0.901
Inf2										0.925
Inf3										0.92
Lat1						0.892				
Lat2						0.897				
Lat3						0.83				
Lgs1							0.955			
Lgs2							0.779			
Mkt1								0.818		
Mkt2								0.804		
Mkt3								0.759		
OSS1									0.899	
OSS2									0.889	
OSS3									0.915	

Table 17 Phase III – Final Model Indicator Reliability

Discriminant validity was examined for the final measurement model using three methods. The first is cross-loadings, where the indicators of each construct

should be the highest cross-loading across all indicators. That is shown in Table 18, with the highest values highlighted and bolded.

Indicator	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
Adp1	**0.897**	0.242	0.225	0.358	0.595	0.222	-0.125	0.424	0.29	0.25
Adp2	**0.801**	0.199	0.147	0.324	0.491	0.141	0.022	0.365	0.311	0.185
Adp3	**0.833**	0.199	0.199	0.322	0.557	0.189	-0.172	0.346	0.187	0.223
Cmp1	0.253	**0.947**	-0.066	0.026	0.193	0.668	-0.38	0.095	0.049	0.748
Cmp2	0.221	**0.929**	-0.094	-0.012	0.198	0.667	-0.359	0.085	0.037	0.734
Cst2	0.093	-0.147	**0.702**	0.023	0.087	-0.105	0.3	0.123	0.111	-0.176
Cst3	0.245	-0.048	**0.963**	0.234	0.183	-0.067	0.23	0.199	0.186	-0.119
ESS1	0.347	-0.013	0.152	**0.895**	0.315	-0.004	0.092	0.364	0.16	-0.016
ESS2	0.351	0.011	0.164	**0.921**	0.272	0.001	0.033	0.305	0.087	0.024
ESS3	0.381	0.026	0.217	**0.907**	0.261	-0.013	0.033	0.3	0.14	0.003
IPr1	0.505	0.166	0.138	0.286	**0.781**	0.115	-0.043	0.243	0.19	0.142
IPr2	0.505	0.141	0.12	0.28	**0.803**	0.128	-0.039	0.373	0.252	0.108
IPr3	0.538	0.182	0.156	0.214	**0.811**	0.224	-0.147	0.389	0.228	0.187
IPr4	0.548	0.18	0.151	0.226	**0.831**	0.174	-0.1	0.423	0.261	0.196
Inf1	0.224	0.716	-0.156	0.002	0.205	0.618	-0.416	0.071	0.005	**0.901**
Inf2	0.242	0.733	-0.145	0.004	0.162	0.638	-0.4	0.077	-0.026	**0.925**
Inf3	0.25	0.72	-0.116	0.005	0.178	0.662	-0.417	0.097	-0.05	**0.92**
Lat1	0.195	0.603	-0.04	0.024	0.163	**0.892**	-0.439	0.072	-0.095	0.587
Lat2	0.227	0.646	-0.071	0.008	0.193	**0.897**	-0.3	0.093	-0.017	0.626
Lat3	0.138	0.62	-0.138	-0.069	0.167	**0.83**	-0.442	0.058	-0.085	0.632
Lgs1	-0.121	-0.377	0.232	0.044	-0.123	-0.445	**0.955**	-0.003	0.177	-0.43
Lgs2	-0.057	-0.306	0.293	0.067	-0.03	-0.281	**0.779**	0.018	0.148	-0.347
Mkt1	0.372	0.092	0.174	0.342	0.382	0.047	0.032	**0.818**	0.217	0.103
Mkt2	0.37	0.052	0.134	0.277	0.393	0.095	-0.064	**0.804**	0.148	0.057
Mkt3	0.327	0.087	0.17	0.22	0.277	0.067	0.048	**0.759**	0.209	0.052
OSS1	0.298	0.043	0.16	0.164	0.304	-0.067	0.198	0.266	**0.899**	-0.025
OSS2	0.249	0.011	0.173	0.091	0.213	-0.066	0.132	0.193	**0.889**	-0.043
OSS3	0.287	0.067	0.171	0.124	0.257	-0.053	0.168	0.184	**0.915**	-0.007

Table 18 Phase III – Final Model Cross Loadings Criterion

Second, discriminant validity was examined using the Fornell-Larcker method, where values are calculated for each construct and tabulated against all other constructs, expecting their value to be the highest for each construct. That is the case for all constructs, as shown in Table 19, and the highest values are highlighted and bolded. The last discriminant validity examination method is HTMT, examined in Table 20, and all values are under the 0.9 threshold. That completes the final measurement model verifications. The final measurement model is presented in Figure 19.

Construct	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
BD Adoption	**0.845**									
Compute Compatibility	0.254	**0.938**								
Cost	0.228	-0.084	0.842							
Enterprise Source SW	0.397	0.009	0.197	**0.908**						
Industry Pressure	0.651	0.208	0.176	0.31	**0.806**					
Latency Compatibility	0.22	0.711	-0.087	-0.006	0.2	**0.873**				
Legislation Barrier	-0.112	-0.394	0.28	0.057	-0.104	-0.436	**0.872**			
Market Turbulence	0.449	0.097	0.2	0.355	0.445	0.088	0.004	**0.794**		
Open Source SW	0.31	0.046	0.186	0.142	0.289	-0.069	0.186	0.24	**0.901**	
Storage Interface Compatibility	0.261	0.79	-0.151	0.004	0.198	0.699	-0.449	0.09	-0.027	**0.915**

Table 19 Phase III – Final Model Fornell Larcker Criterion

Construct	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
BD Adoption										
Compute Compatibility	0.303									
Cost	0.269	0.152								
Enterprise Source SW	0.469	0.025	0.196							
Industry Pressure	0.801	0.247	0.214	0.365						
Latency Compatibility	0.256	0.831	0.141	0.049	0.237					
Legislation Barrier	0.162	0.491	0.474	0.079	0.113	0.542				
Market Turbulence	0.595	0.124	0.278	0.445	0.577	0.11	0.076			
Open Source SW	0.368	0.051	0.228	0.157	0.335	0.089	0.229	0.302		
Storage Interface Compatibility	0.305	0.894	0.224	0.019	0.229	0.802	0.547	0.112	0.04	

Table 20 Phase III – Final Model HTMT

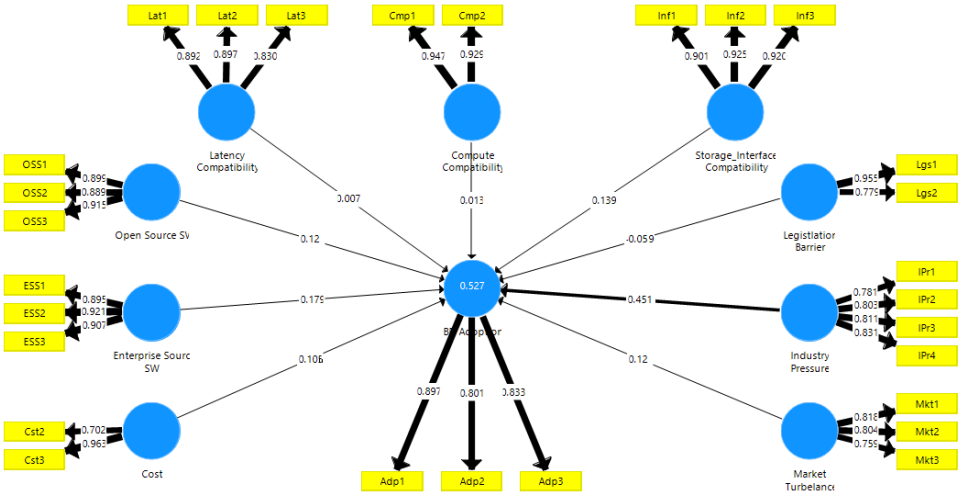

Figure 19 Phase III – Final Measurement Model

4.4.2.3 Phase III – Structural Model Evaluation

First, collinearity issues were assessed by calculating the variance inflation factor (VIF) for both outer (indicators) and inner models (constructs) using SmartPLS software. VIF values were calculated from the earlier run using the PLS algorithm in the measurement model. VIF value of 5 and more is considered a critical level of collinearity. None of the VIF values for both outer and inner models have any value over 3.4, indicating no critical collinearity issues in the current model. See Table 21 for outer model collinearity VIF values and Table 22 for inner model collinearity VIF values.

Indicator	VIF
Adp1	2.096
Adp2	1.579
Adp3	1.74
Cmp1	2.371
Cmp2	2.371
Cst2	1.305
Cst3	1.305
ESS1	2.529
ESS2	3.081
ESS3	2.557
IPr1	1.845
IPr2	1.982
IPr3	2.018
IPr4	2.163
Inf1	2.667
Inf2	3.151
Inf3	2.923
Lat1	2.25
Lat2	1.968
Lat3	2.008
Lgs1	1.454
Lgs2	1.454
Mkt1	1.439
Mkt2	1.39
Mkt3	1.332
OSS1	2.315
OSS2	2.535
OSS3	2.784

Table 21 Phase III – Collinearity VIF values for Outer Final Model.

Construct	BD Adoption
BD Adoption	
Compute Compatibility	3.135
Cost	1.186
Enterprise Source SW	1.207
Industry Pressure	1.441
Latency Compatibility	2.368
Legislation Barrier	1.435
Market Turbulence	1.367
Open-Source SW	1.2
Storage Interface Compatibility	3.097

Table 22 Phase III – Collinearity VIF values for Inner Final Model.

The value and significance of path coefficients were evaluated by a bootstrapping sampling method using 5,000 subsamples from the 411 respondents' data for two-tailed tests at a 0.05 confidence level. This calculation determines the path coefficients (strength of the relationship) of each adoption factor constructs to the Big Data adoption constructs, the direction of these relationships (positive or negative), and the significance of these relationships based on the p-value.

Path coefficient is essentially a standardized regression coefficient. The standardization is intended to enable a comparison between the different regression coefficients. In the context of path analysis, a unit increase in the independent variable's standard deviation will result in a "path coefficient" change in the dependent variable's standard deviation (Benitez, 2020). Path coefficient is used as a measure to represent the strength of the relationship between the independent

(exogenous) variables and the dependent (endogenous) variables. The sign of the path coefficient indicates a positive or negative change.

A p-value of < 0.05 indicates significant relationship. An original sample-path coefficient positive value indicates positive correlation, whereas a negative value indicates negative correlation. The original sample-path coefficient value indicates the contribution of that construct to the Big Data adoption construct. See Table 23 for the details of the path coefficients. Significant paths are highlighted by green color. Path coefficients with significant p-values are evaluated further in the phase III results summary.

Path	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	P-Value
Industry Pressure -> BD Adoption	0.451	0.447	0.053	8.581	0
Enterprise Source SW -> BD Adoption	0.179	0.177	0.043	4.191	0
Storage Interface Compatibility -> BD Adoption	0.139	0.136	0.06	2.298	0.022
Market Turbulence -> BD Adoption	0.121	0.123	0.047	2.563	0.01
Open-Source SW -> BD Adoption	0.12	0.122	0.037	3.209	0.001
Cost -> BD Adoption	0.106	0.108	0.043	2.461	0.014
Compute Compatibility -> BD Adoption	0.013	0.012	0.061	0.209	0.835
Latency Compatibility -> BD Adoption	0.007	0.01	0.052	0.127	0.899
Legislation Barrier -> BD Adoption	-0.059	-0.062	0.042	1.405	0.16

Table 23 Phase III – Path Coefficients for Final Model.

4.4.2.4 Phase III – PLS Model Predictive Power

The coefficient of determination R^2 is used to measure the model's in-sample predictive power. R^2 value ranges from 0 to 1, where a higher value indicates a greater explanatory power. R^2 values of 0.75, 0.50, and 0.25 are considered substantial, moderate, and weak, respectively. Values of 0.90 and higher indicate overfitting, where random noise is included in the R^2 (Joseph F Hair et al., 2019). The coefficient of determination for the endogenous construct (dependent variable) of "BD Adoption" is 0.527, which is moderate for this model. It has a p-value of 0, thus significant. See Table 24.

	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	P Values
BD Adoption	0.527	0.542	0.041	12.778	0

Table 24 Phase III – Coefficients of Determination R^2 Final Model.

Predictive relevance Q^2 also needs to be evaluated to measure the accuracy of predicting data not used in model estimation. This is done by removing a single data point, imputing the removed point with the mean, and estimating the model parameters. In other words, Q^2 shows how well the data collected empirically can be reconstructed with the help of the model and the PLS parameters (Akter, D'ambra, & Ray, 2011). The Q^2 value should be greater than 0 for a specific endogenous construct to indicate the predictive accuracy of the structural model for

that construct. Q^2 values higher than 0, 0.25, and 0.50 show the structural model's small, medium and large predictive relevance (Joseph F Hair et al., 2019). Q^2 is obtained using the SmartPLS blindfolding procedure with an omission distance value of 7. In the endogenous construct of "BD Adoption," the value of Q^2 is 0.362, which falls between medium and large predictive relevance. See Table 25 for more details.

Construct	Q^2 (=1-SSE/SSO)
BD Adoption	0.362

Table 25 Phase III – Predictive relevance Q^2 Final Model

The removal effect of the independent latent construct on the dependent latent variable R^2 value needs to be evaluated using the f^2 effect size. This reduction of variance in the endogenous variable can assist in estimating the effect of that removal. The rule of thumb for the f^2 effect size value needs to be higher than 0.02, 0.15, and 0.35 to be regarded as a small, medium, and large effect, respectively (Joseph F Hair et al., 2019; Mathai, 2019). These paths were identified to have surpassed the total effect threshold of significance and listed in descending order of significance (industry pressure, enterprise source SW, open-source SW, Market Turbulence, and cost). In the same order, they have effect size of 0.298 (medium), 0.056 (small), 0.025 (small), 0.023 (small), and 0.02 (small) respectively. The expectation is that not all latent constructs will have a large effect size. Not all

paths are expected to play large roles, just like not all actors will have a lead role in a movie (Benitez, 2020). See Table 26 for details.

Path	f^2
Industry Pressure -> BD Adoption	0.298
Enterprise Source SW -> BD Adoption	0.056
Open-Source SW -> BD Adoption	0.025
Market Turbulence -> BD Adoption	0.023
Cost -> BD Adoption	0.02
Storage Interface Compatibility -> BD Adoption	0.013
Legislation Barrier -> BD Adoption	0.005
Compute Compatibility -> BD Adoption	0
Latency Compatibility -> BD Adoption	0

Table 26 Phase III – Effect Size f^2 Final Model.

Effect size q^2 measures the predictive relevance of the inner (structural model) to the endogenous (dependent) latent variable. Effect size q^2 value needs to be higher than 0.02, 0.15, and 0.35 to be regarded as weak, moderate, and strong, respectively. Negative or close to zero values indicate exogenous constructs are not relevant to predicting a given endogenous construct. SmartPLS does not calculate q^2 directly, but by removing each exogenous construct from the model and calculating that Q^2 as Q^2 Excluded (for that excluded construct), then calculate it q2 compared to the original Q^2 using the formula mentioned in section 3.5.6 of this document. The exogenous construct enterprise source software has a value of 0.028 (weak) and industry pressure has a value of 0.155 (moderate). The rest of the exogenous constructs are less than 0.02. Similar to effect size, not all constructs

are expected to have large roles (Joseph F Hair et al., 2019). See Table 27 for details.

Construct	Q2 Included	Q2 Excluded	q^2
Industry Pressure	0.362	0.263	0.155
Enterprise Source SW	0.362	0.344	0.028
Open-Source SW	0.362	0.355	0.011
Market Turbulence	0.362	0.356	0.009
Cost	0.362	0.357	0.008
Storage Interface Compatibility	0.362	0.359	0.005
Legislation Barrier	0.362	0.362	0.000
Compute Compatibility	0.362	0.363	-0.002
Latency Compatibility	0.362	0.364	-0.003

Table 27 Phase III – Predictive relevance q^2 Final Model.

PLSPredict is another method to evaluate the predictive power of the model. It uses multiple folds of a holdout sample from the current sample to determine the predictive power within the given sample. With the 411 responses sample, the PLSPredict calculation in SmartPLS was run with ten folds and ten repetitions. The prediction between the PLS model and "naive" regression LM (Linear Model) are compared in 2 aspects.

The first aspect of the PLSPredict evaluation is the $Q^2_{Predict}$ for the VM (items). $Q^2_{Predict}$ values for PLS are above zero for all the items. That indicates the PLS model outperforms the LM model (Joseph F Hair et al., 2019; G. Shmueli et al., 2019). See Table 28.

The second is evaluating the error aspects using RMSE and MAE for the Manifest Variables (MV) (indicator or item). The comparison is between the PLS

and LM in these two error indicators in which lower values indicate better predictive power. RMSE values are lower in the PLS model than in the LM. Thus, the PLS model has high predictive power. For MAE, all indicators have a lower PLS value than in ML, indicating high predictive power for PLS (Joseph F Hair et al., 2019; G. Shmueli et al., 2019). See Table 29.

Indicator	Q^2_predict
Adp1	0.427
Adp3	0.339
Adp2	0.284

Table 28 Phase III – PLSPredict $Q^2_{Predict}$ Final Model.

PLS				LM		
	RMSE	MAE			RMSE	MAE
Adp1	0.684	0.519		Adp1	0.713	0.537
Adp3	0.702	0.536		Adp3	0.73	0.543
Adp2	0.859	0.633		Adp2	0.899	0.677

Table 29 Phase III – PLSPredict MV Error Final Model.

An essential method for predictive validity assessment of the PLS model using an out of sample holdout data (Cepeda Carrion, 2016; Hair J.F, 2012). The holdout sample of 100 responses that are not used in the training sample above will be used in this eight-step evaluation. Step 1 is to split the sample. This is done as stated earlier in 4.4.2.1 Missing, Anomalous, Outliers Data and Sample Evaluation. Four hundred and eleven respondents' data are used as the training sample, and 100 respondents' data are used as the holdout sample. Step 2 is to estimate the model based on the training sample. This is done by using the PLS path model parameters

based on path coefficients (table 23) and the outer weights for the items to latent variables (table 30 below).

Indicator	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
Adp1	0.437									
Adp2	0.362									
Adp3	0.382									
Cmp1		0.569								
Cmp2		0.496								
Cst2			0.309							
Cst3			0.814							
ESS1				0.354						
ESS2				0.359						
ESS3				0.389						
IPr1					0.298					
IPr2					0.299					
IPr3					0.318					
IPr4					0.324					
Inf1										0.342
Inf2										0.369
Inf3										0.382
Lat1						0.397				
Lat2						0.461				
Lat3						0.28				
Lgs1							0.755			
Lgs2							0.357			
Mkt1								0.438		
Mkt2								0.435		
Mkt3								0.385		
OSS1									0.397	
OSS2									0.331	
OSS3									0.382	

Table 30 Phase III –Outer Weights Final Model.

Step 3 is to standardize the holdout sample by subtracting the mean of each indicator from each response and then dividing it by the standard deviation of the indicator. See Appendix I for the Table. Step 4 is to create construct scores for the holdout sample. This is done by creating a linear equation of each latent variable construct based on the outer weights of the observable variables that contribute to it. This can be summarized as the following formulas, and the actual values and calculations table is in Appendix I:

Lat (latency Compatibility)	= 0.397 Lat1 + 0.461 Lat2 + 0.28 Lat3
Cmp (Compute Compatibility)	= 0.569 Cmp1 + 0.496 Cmp2
Inf (Storage Interface Compatibility)	= 0.342 Inf1 + 0.369 Inf2 + 0.382 Inf3
OSS (Open-Source Software)	= 0.397 OSS1 + 0.331 OSS2 + 0.382 OSS3
Ess (Enterprise-Source Software)	= 0.354 Ess1 + 0.359 Ess2 + 0.389 Ess3
Cst (Cost)	= 0.309 Cst2 + 0.814 Cst3
Ipr (Industry Pressure)	= 0.298 IPr1 + 0.299 IPr2 + 0.318 Ipr3 + 0.324 IPr4
Lgs (Legislation Barrier)	= 0.755 Lgs1 + 0.357 Lgs2
Mkt (Market Turbulence)	= 0.438 Mkt1 + 0.435 Mkt2 + 0.385 Mkt3
Adp (Big Data Adoption)	= 0.437 Adp1 + 0.362 Adp2 + 0.382 Adp3

Step 5 is to standardize the construct scores for the holdout sample. This is like the earlier standardization, subtracting the mean from each of the values and dividing it by the standard deviation. The actual table is in Appendix I. Step 6 is to create prediction scores for each endogenous (dependent) constructs as a linear equation of the exogenous (independent) variables based on the path coefficient values from Table 23. This equation is represented as follows, and the actual table of these scores is included in Appendix I.

$$Adoption_{Predicted} = 0.007*Lat + 0.12*Oss + 0.451*Ipr + 0.013*Cmp + 0.106*Cst - 0.059*Lgs + 0.39*Inf + 0.179*Ess + 0.121*Mkt$$

Step 7 determines the proportion of explained variance (R^2) as the squared correlation of the prediction scores and the construct scores for each endogenous construct of the holdout sample. R^2 value is 0.482, which is considered moderate (Joseph F Hair et al., 2019). Step 8 compares R^2 to the training sample's R^2 (in-sample predictive power), 0.527 (also moderate). As expected, both the training sample and hold-out sample R2 values are similar. This substantiates how the statistical analysis is generalized and how well the predictive model performs in practice.

This same procedure for steps 6, 7, and 8 was done with only the significant paths of cost, enterprise source software, industry pressure, market turbulence, open-source software, and storage interface compatibility. The R^2 value for only significant paths was slightly higher at 0.494, which is moderate explanatory power.

4.4.3 Phase III Results

The findings of the PLS-SEM analysis on the model have resulted in a modification to the measurement model by removing an observable variable for cost and another one for legislative barriers. With that change, the measurement model passed the validity and reliability test. The structural model did not change

and showed multiple significant exogenous constructs (independent variables) that affect the endogenous construct (dependent variable) of the Big Data adoption. The model can be displayed as related to the hypotheses presented earlier in figure 20, where hypothesis 1a is represented as H1a and so on.

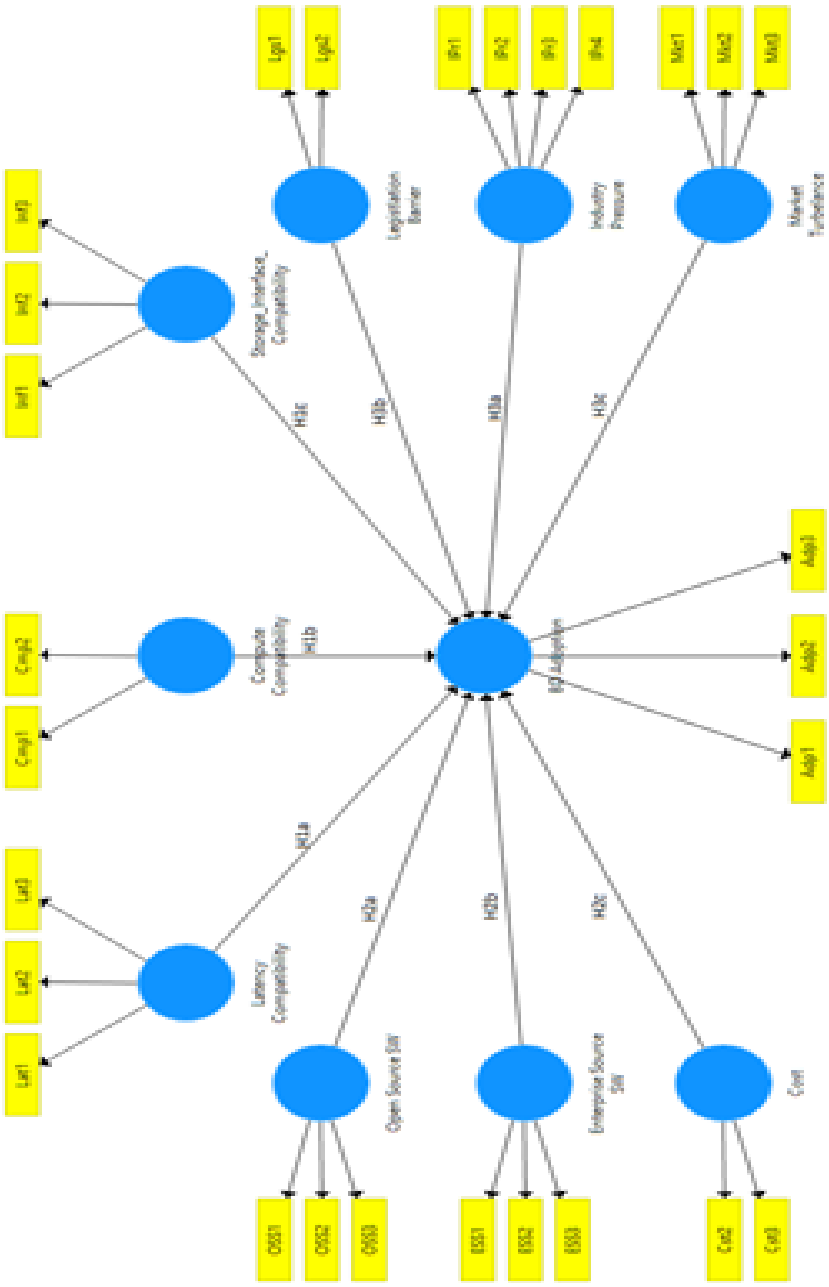

Figure 20 Phase III –PLS Model and Hypotheses.

PLS-SEM focuses on maximizing the explained variance and path predictive power estimates (Joe F. Hair Jr et al., 2017). Thus, the structural model can be interpreted by examining the Coefficient of determination R^2 values of the dependent variable (Big Data adoption) of 0.527, which is considered a moderate effect. See Table 24 above.

The second area of focus in the structural model is to test the stated hypotheses by testing the significance of the path coefficients. Path coefficients were tested, and their t-values and p-values were calculated using PLS bootstrapping method. See Table 23 above. There are six significant factors from the nine that are tested. The six significant factors have a positive correlation to Big Data adoption.

The path between industry pressure and Big Data adoption has (path = 0.451, t = 8.581, p = 0), enterprise-sourced software to Big Data adoption has (path = 0.179, t = 4.191, p = 0), storage interface compatibility to Big Data adoption has (path = 0.139, t = 2.298, p = 0.022), market turbulence to Big Data adoption (path = 0.121, t = 2.563, p = 0.01), open-sourced software to Big Data adoption (path = 0.12, t = 3.209, p = 0.001), and cost to Big Data adoption (path = 0.106, t = 2.461, p = 0.014). Thus, this study supports the significance of six of the hypotheses of factors influencing Big Data adoption at an organizational level. The direction of the correlation is positive for all the significant paths to the dependent adoption variable. Despite their significance, the hypotheses' direction did not match

enterprise source software, market turbulence, and cost. The significance and direction agree with the hypotheses for data storage interface, industry pressure, and open-source software for adopting Big Data. However, compute, and latency compatibilities and legislation barriers factors were not found to be significant factors in adopting Big Data in this study. Table 31 shows a summary of the hypotheses and their corresponding finding.

No.	Aspect	Hypothesis	Significance Finding	Direction Finding	Findings
H1a	Technical Factors	Data storage latency compatibility will positively correlate with Big Data adoption.	Not Significant	Positive	Not Supported
H1b		Data storage compute compatibility will positively correlate with Big Data adoption.	Not Significant	Positive	Not Supported
H1c		Data storage interface compatibility will positively correlate with Big Data adoption.	**Significant**	**Positive**	***Supported***
H2a	Organizational Factors	Open-source software availability of Big Data solutions will positively correlate with Big Data adoption.	**Significant**	**Positive**	***Supported***
H2b		Enterprise source software availability of Big Data solutions will negatively correlate with Big Data adoption.	**Significant**	**Positive**	*Not Supported*
H2c		Perceived cost of Big Data will negatively correlate with Big Data adoption.	**Significant**	**Positive**	Not Supported
H3a	Environmental Factors	Perceived industry pressure for Big Data will positively correlate with Big Data adoption.	**Significant**	**Positive**	***Supported***
H3b		Legislation barriers of Big Data will negatively correlate with Big Data adoption.	Not Significant	Negative	Not Supported
H3c		Market turbulence of Big Data will negatively correlate with Big Data adoption decisions.	**Significant**	**Positive**	*Not Supported*

Table 31 Phase III –Hypotheses and Findings Summary.

4.5 Summary of the Results

The three phases of this research incrementally added more data, insights, refinement, and verification to the area of Big Data adoption for organizations with data storage systems. Taking a pragmatic approach, qualitative, mixed, and quantitative research methodologies were used to explore, validate, and test the proposed hypotheses.

Phase I added to the subject by qualitatively exploring the Big Data adoption factors from nine practitioners who have implemented Big Data on their data storage systems. First is the challenge of finding value from the data and linking it to a business case was one of the findings for these organizations. Second is the security challenge that addresses Big Data's diversity of data and the variety of security risk factors in the large data sets from various sources. The third is the challenge of regulation as organizations try to chart a new balanced path between value and regulations. In addition is the compatibility of network and infrastructure as the data volume, variety, and velocity put pressure on current systems that may not be able to handle it.

In addition, phase I confirmed that the nine factors proposed needed further explorations with a broader audience. Most phase I participants assessed that the nine factors identified in this research are worth investigating further. Third, phase

I recruited, in part, the participants for phase II to take and validate the pilot survey instrument.

Phase II validated the survey instruments by taking and reviewing the pilot survey. The input from participants improved clarity and word choices and provided face and content validities. Phase II participants also tested the survey instruments and provider and first investigated the logistics and time it takes to take the survey. They also provided valuable insights and feedback for the landing page and helped improve the survey questions. The researcher also tested Qualtrics, the survey provider, and their features for coding and downloading the data.

Phase III has provided the quantitative methodology to test the proposed hypotheses of which factors are significant in adopting Big Data for organizations with data storage systems. The number of respondents to the phase III survey was over 511, much higher than the minimum of 380 required to have statistically significant results. It also allowed us to randomize and then split the collected sample into training and holdout samples to test out of sample validity and predictive validity. Phase III demographics were diverse regarding titles, industries, organizations, sizes, and location in the information technology space. Few anomalous data were detected and removed from the sample.

The PLS-SEM model consists of measurement and structural models assessed in part by SmartPLS 3 software. The initial measurement model resolved a reliability issue by removing low reliable items per the PLS-SEM procedure. That

was corrected, but there was a discriminant validity issue in the HTMT test between the compatibility construct and the storage interface compatibility construct. Once that is fixed, the modified model passes the measurement model reliability and validity tests. Discriminant validity of the measurement model was also tested and verified to meet the acceptance criteria in the three methodologies of cross loading, Fornell Larcker criterion, and HTMT. Thus, that model was designated as the final model and was used for the rest of this research.

The phase III structural model was evaluated with SmartPLS 3 as well. First, collinearity issues were assessed by calculating the variance inflation factor (VIF) for outer and inner models. No collinearity issues were found for inner and outer models. Bootstrapping was done next to determine the path coefficient for the inner model and their significance using T statistics and p-values. Six of the nine paths were significant (p-value < 0.05). They are storage interface compatibility (technical factor), open-source software, enterprise source software, cost (organizational factors), industry pressure, and market turbulence (environmental factors).

The next phase was to determine the model's predictive power using multiple techniques. The coefficient of determination (R^2) was 0.527, which is moderate and does not exceed the threshold of 0.9 for overfitting. The following technique was effect size (f^2) which identified industry pressure as a medium effect and enterprise source software, open-source software, market turbulence, and cost

as a small effect. That is not to say these are not significant; it is just to say that the lead role in this model is for industry pressure.

The other technique used in predictive relevance (Q^2) is calculated using the blindfolding procedure in SmartPLS. Q^2 needed to be larger than 0 and was found to be 0.362, which is medium predictive relevance. The predictive relevance of the inner structural model can also be measured using effect size q^2 to the endogenous (dependent) latent variable. Industry pressure has a value of 0.155 (moderate), and enterprise source software has a value of 0.028 (weak). Similar to effect size, not all constructs are expected to have large roles (Joseph F Hair et al., 2019). Another technique used to evaluate the prediction of the model is PLSPredict. That also showed that the PLS-SEM model is better than the linear model in all indicators, indicating high predictive power.

All the previous predictive model evaluation techniques used the same sample data. The following technique uses a holdout sample (in this case, 100 respondents) to evaluate the predictive model. Manually calculating the holdout sample coefficient of determination (R^2) was done through the eight-step process (Cepeda Carrion, 2016; Hair J.F, 2012). The R^2 value of 0.494 is moderate and, like the original model's R2 value, 0.527, which indicates the model's predictive power.

CHAPTER 5: SUMMARY

5.1 Summary and Recommendations

This research aims to find factors that affect organizations with data storage systems to adopt Big Data. Through literature review and qualitative research in phase I, novel factors that were not studied or were understudied were found. A phase III qualitative large-scale survey was done to determine the significance of the proposed adoption factors. It also found the direction of that correlation to the adoption for each of these factors (positive or negative). This chapter discusses the findings, their importance, implications, limitations, and future research recommendations.

5.2 Key Findings and Importance

The research contributes to the literature by presenting a current state of studied factors of Big Data adoption, main research theories, and frameworks. It introduces novel adoption factors that have not been studied in this space, such as open-source, enterprise source software and storage interface compatibility. The quantitative research showed these significant factors in the Big Data adoption decision. Also, storage latency and compute compatibilities are other novel factors identified in the qualitative part of this research, which were not confirmed to be

significant in this research. Understudied factors verified to be significant and positive are industry pressure, market turbulence, and cost. The study also demonstrated that under-studied factors and legislation barriers are not significant in Big Data adoption.

5.3 Theoretical and Practical Implications

There are multiple theoretical contributions to this research. First, the use of semi-structured interview IPA qualitative methodology to find insights that have been argued is needed in the IS field in general (Mingers, 2001). It produced novel insights on Big Data adoption that were not explored previously, such as the challenge of linking Big Data to value and the need for more granular compatibility. These findings should be quantitatively studied further to verify their significance. IPA or other qualitative methodologies can be used to explore other gaps in Big Data adoption.

The quantitative part of the research has verified the significance of three novel Big Data adoption factors that have not been previously studied. These factors positively correlate with the Big Data adoption decision. The other factors that were understudied in this field were industry pressure, market turbulence, and cost, which are confirmed to be significant and positively correlated with Big Data adoption. It was studied earlier once for industry pressure factor and found

significant (Nam et al., 2015). This study replicated that finding of significance for industry pressure. It also replicated the significance of the market turbulence effect on Big Data adoption and found a positive correlation (Sun et al., 2018).

These research findings have multiple implications for the practice of Big Data adoption. It provides new insights and adoption factors to practitioners and academics on enabling the more extensive adoption of Big Data. Some insights are concerns and challenges that need to be understood and mitigated to enable Big Data adoption, such as extracting value with Big Data, Big Data specific security, and regulation compliance. While others are factors that can be utilized and enhanced to enable further Big Data adoption, such are the significant factors found in this research. PLS-SEM usage allowed the simultaneous evaluation of nine adoption factors simultaneously. It also shows the relative effect size of each factor on the Big Data adoption decision.

The qualitative findings show the need to link Big Data to business goals, and in many instances, the value is found by the process of trials and refinements of the queries to Big Data. Hirsch compared Big Data to oil in terms of value and other aspects (Hirsch, 2014). In their raw form, both may have limited value. This comparison can also be extended to the various uses and processes to achieve them. There are over 6,000 products made of petroleum (Abutu, 2014). Similarly, Big Data products/use cases can be as diverse as the organizations that use them (Eggers & Hein, 2020). To realize Big Data value es identifying the needs of the

organizations and their customer and defining these as use cases with specific business and monetary objectives. Then design a Big Data solution that supports these use cases.

Another finding from the qualitative part is security, which has nuanced new challenges for Big Data. One definition of Big Data is the ability to process data characterized by volume, variety, and velocity (Russom, 2011). As Big Data technology processes a variety of data formats (text, table, document, image, video, audio, etc.), it is faced with securing unexpected restricted, sensitive, or individually identifiable data that it parses. This information may come in a not-expected format or in an identifiable way that may not be apparent. This also manifests as a legislation burden since the data becomes a liability to the company that can negatively impact the organization in how it retains and processes the data.

From a quantitative perspective, this research showed that industry pressure is the highest contributing significant factor for Big Data adoption from the examined factors, and it is positively correlated. The novel finding of this research is that open-source software, enterprise-source software, and storage interface compatibility are significant factors in Big Data adoption. They are also positively correlated to Big Data adoption. Open-source software and enterprise-source software may seem like opposing factors. Still, from this research (qualitatively and verified quantitatively), they are significant and work together to enable Big Data adoption. There is a need to have an open-source community that drives

innovation and new features. There is also a need to have enterprise-sourced software with stable and dependable support.

Market turbulence is also a significant factor positively correlated to Big Data adoption. As the market becomes less predictable, organizations can adopt and use Big Data to navigate that challenge. The cost was also found to be a significant factor but also positively correlated with Big Data adoption.

5.4 Limitations and Future Research Opportunities

With a low adoption rate of Big Data in production and a high failure rate of adopting Big Data, there is a need to study Big Data adoption factors (H.-M. Chen et al., 2015; de Camargo Fiorini et al., 2018). The literature review provided a baseline list of the current studied factors and the number of studies for each factor. That list of factors can be expanded. Un-studied or under-studied factors can be explored by future research.

This research used a pragmatic mixed-methods approach in multiple phases (R. B. Johnson & Onwuegbuzie, 2004; Venkatesh et al., 2013). This research used semi-structured interviews with IPA, structured interviews, a pilot survey (qualitative), and a large-scale survey (quantitative). Future research can use other

combinations of qualitative and quantitative research methods to investigate Big Data adoption factors.

The large-scale survey method had over 500 respondents, practitioners, and academics who participated in their organizations' decision to adopt Big Data. The research was structured with an exploratory outlook. It will be interesting if the results can be replicated, and confirmatory research is conducted in the future.

This exploratory study covered diverse locations, organization sizes, industries, and professions related to Big Data adoption. Future studies can be exploratory to investigate other aspects and populations or focus more on specific populations (CIOs, architects, etc.)

Compatibility is one of the most studied factors, possibly because it is one of the initial DOI theory factors (K. Agrawal, 2015; H.-M. Chen et al., 2015; Esteves & Curto, 2013; Mahesh et al., 2018; Salleh & Janczewski, 2018; Verma & Bhattacharyya, 2017). In contrast, more granular aspects of compatibility adoption factors are least studied. There is a gap for a more detailed understanding of compatibility and its various aspects on Big Data adoption. That was also verified in phase I of the semi-structured interview IPA method and the structured interview process. Yet, phase III of this research using a large-scale quantitative survey did find only storage interface compatibility to be significant (latency and compute ability were not). The three compatibility constructs had discriminant validity issues due to their similarities. Each aspect of compatibility may need to be studied

separately to reduce the discriminant validity issues. Other aspects of compatibility can be explored in future research, and discriminant validity issues need to be considered when studying more than one of these factors.

The perceived cost was studied qualitatively and identified as the main factor in Big Data adoption (Verma & Bhattacharyya, 2017). The cost was also a main factor in phase I of this research in the structured interview process. Phase III of this research using a large-scale quantitative survey did find the cost to be a significant factor in adopting Big Data but positively correlated with Big Data adoption. This can be explored further in future research.

5.5 Conclusion

64.2 ZB of digital data was created in 2020, despite the pandemic and all the downward repercussions associated with that, according to International Data Corporation (Rydning, 2021). Compare that to 9.3 ZB in 2016 (Westervelt, 2017). That amount of data is the highest generated in human history, with numbers expected to rise even more in 2021 and beyond (Holst, 2021). The importance of data cannot be overstated. As Hirsch declared data to be more valuable than oil (Hirsch, 2014), Parkins stated data to be the most valuable resource on the planet (Parkins, 2017). Yet many organizations are not able to capture that value from the data they own on their data storage systems using Big Data (Ajimoko, 2017; H.-M.

Chen et al., 2015; Dubey et al., 2016). As Rogers described in his seminal work "Diffusion of Innovations" with multiple examples, great ideas may take centuries to be adopted or not adopted at all (Everett M Rogers, 2003). Researchers and practitioners need to take a closer look at the factors than enable or prohibit the adoption of Big Data from empowering an adoption of a technology or an innovation and extracting additional value from the stored data.

In the context of data, data storage, and Big Data, this research endeavored to explore the significant Big Data adoption factors for organizations that own their data storage system. Since there are many dimensions or factors, the researcher used a TOE framework combined with DOI theory (DePietro, 1990; Everett M Rogers, 2003). Succinctly, it attempted to answer the following research questions:

1) What are the significant technical factors affecting Big Data adoption for organizations with data storage systems?
2) What are the significant organizational factors affecting Big Data adoption for organizations with data storage systems?
3) What are the significant environmental factors affecting Big Data adoption for organizations with data storage systems?

To answer the above questions, mixed-method research was conducted in three phases. The initial two phases were qualitative using a semi-structured IPA

interview methodology, structured interviews, and a pilot survey review. Phase I provided novel insights and factors for Big Data adoption based on interviewees who have experienced the phenomenon of adopting Big Data in their organizations. IPA provided the following factors: the challenge of linking value to data, security challenges specific to Big Data, the burden of regulations, and Big Data need for a big network. The structured interviews provided validation that the nine factors identified as un-studied or under-studied need further research. Phase II provided corrections and validation of the survey instruments. Phase III provided the quantitative methodology to test the hypotheses of which factors identified are significant and what are their correlation direction to Big Data adoption.

The first question of identifying significant factors is answered in part in the three phases of the research. Phase I semi-structured IPA showed that regulations and compatibility are important factors that need further study. Phase I structured interviews confirmed that compatibility and regulations are important and identified the other factors. The technical factors of compatibility (storage latency, compute-ability, and interface compatibilities) are identified as novel and important factors. Organizational factors of open-source software, enterprise-source software, and cost are also important. Environmental factors of industry pressure, regulations, and market turbulence are important to investigate further.

Phase II provided feedback and validity to the survey instruments (to all the research questions). Phase III provided a quantitative test of significance of the

three aspects of compatibility (TOE) and the nine identified factors. In the technological adoption factors context, data storage latency compatibility is a significant factor in adopting Big Data and correlates positively to adoption. The other two compatibility aspects identified (compute-ability and interface compatibilities) were not significant. These compatibility factors or other compatibility factors may need to be examined in future research since the phase I data suggests compatibility in general, which is important to the Big Data adoption decision.

For the second question of organizational adoption factors, the identified open-source software, enterprise-source software (which are novel and have not been studied before), and cost (which has been studied once qualitatively before) were examined in the research phases. Phase I validated that they are all important factors in the structured interview process. Phase III confirmed the significance of open-source and enterprise-source software on Big Data adoption and their positive correlation. The cost was also confirmed as a significant adoption factor and positively correlated with Big Data adoption.

This research's phases examined the third question of environmental adoption factors (industry pressure, legislation barriers, and market turbulence). Phase I IPA method identified legislation as an important factor. Phase I structured interview confirmed all three factors to be important. Phase III confirmed Industry pressure and market turbulence to be significant and positively correlated to Big

Data adoption. However, the legislation barrier was not confirmed based on the survey data.

Overall, this research has identified six significant factors contributing to the positive correlation of Big Data adoption. In organizational factors, open-source software and enterprise-source software are novel significant factors that have not been studied in this context before. The cost has also been significant and positively correlated to adoption. In environmental factors, industry pressure and market turbulence are significant factors that were under-studied in earlier research.

Storage interface compatibility and open-source software are significant and positive enablers of Big Data adoption that come as no major surprise. Conversely, the other factors can be expensive, and some think of them as detracting from Big Data adoption. Organizations that adopted Big Data realized additional value because of these factors. Big Data found a value positively correlated with cost, enterprise-sourced software, industry pressure, and even market turbulence. In other words, the value of data extracted by Big Data is expensive (cost) and requires more technical and financial commitments (enterprise sources software). Yet, Big Data captured value that makes these organizations respond positively to the (industry pressure) and even overcome (market turbulence). These findings are supported quantitatively and qualitatively as participants expressed that Big Data enables their organizations to thrive, compete, and succeed.

As CIOs and leaders of organizations ponder adopting Big Data, one of the most important questions they need to answer is how Big Data will unlock additional value from the data being stored in their organizations. Like crude oil, stored raw data has limited value. Refinement and distillation of crude oil produce over 6,000 oil products with various uses and values. I would say that refinement and distillation of raw data have a much higher number of products and more value. Big Data is a major tool to accomplish that at a scale that no other technology can offer. The challenge is to know what data product is needed to enable your organization to capture that value.

These new findings of significance and correlation direction of the identified adoption factors of Big Data should assist researchers and practitioners in enabling further Big Data adoption enablement and value creation for organizations from the data stored in their data storage systems. This research also identified other research opportunities to be examined in future research.

APPENDICES

APPENDIX A: INTERVIEW QUESTIONS

1- Demographic Information

   a. Name

   b. Email

   c. Role in Organization

   d. Industry

   e. Location (Country)

   f. Organization's Number of Employees (Size)

2- Technical Background Information

   a. Does the organization have a storage system?

   b. Have you adopted or are you planning to adopt Big Data?

3- TOE Unstructured Questions:

   a. What are technical factors that affect your Big Data adoption decision?

   b. What are Organizational factors that affect your Big Data adoption decision?

   c. What are Environmental factors that affect your Big Data adoption decision?

4- TOE Structured Questions:

a. Technical

    i. Describe how “Data Storage Latency Compatibility” has impacted your organization’s Big Data adoption? Positive or negative?

    ii. Describe how “Ability to Compute Large Amount of Data” have impacted your organization’s Big Data adoption? Positive or negative?

    iii. Describe how “Data Storage Interface Compatibility” has impacted your organization’s Big Data adoption? Positive or negative?

b. Organizational

    i. Describe how “Open-Source Software” have impacted your organization’s Big Data adoption? Positive or negative?

    ii. Describe how “Enterprise Source Software” have impacted your organization’s Big Data adoption? Positive or negative?

    iii. Describe how “Cost of Implementing Big Data” has impacted your organization’s Big Data adoption? Positive or negative?

c. Environmental

i. Describe how “Industry Pressure” has impacted your organization’s Big Data adoption? Positive or negative?

ii. Describe how “Legislation Barriers” have impacted your organization’s Big Data adoption? Positive or negative?

iii. Describe how “Market Turbulence” have impacted your organization’s Big Data adoption? Positive or negative?

5- Follow-up Questions

a. Would you like to receive a copy of this research once it is published?

b. Are you willing to evaluate the survey questionnaire pilot study and provide feedback?

Thank you for participating in the interview.

APPENDIX B. CONSTRUCTS, ITEMS AND THEIR SOURCES

Domain	***Construct***	***Ref #***	***Item (Question)***	***Source***	***Original Wording in Source***
Technological	Data Storage Latency Compatibility.	CmpLtn01	BD storage latency requirements align the current internal Information Systems' applications at my organization.	(Lin, 2008)	Implementation of e-business does not contradict the current internal IS applications.
		CmpLtn02	BD storage latency requirements are supported by the existing Information Systems' infrastructure.	(Lin, 2008)	Implementation of e-business is supported by the existing IS infrastructure.
		CmpLtn03	BD storage latency requirements are supported by the organizational IT department resources.	(Lin, 2008)	Implementation of e-business is supported by the organizational IT human resources.
		CmpCmpt01	BD computing requirements align the current internal Information Systems' applications at my organization.	(Lin, 2008)	Implementation of e-business does not contradict the current internal IS applications.
		CmpCmpt02	BD computing requirements are supported by the existing Information	(Lin, 2008)	Implementation of e-business is supported by the existing

			Systems' infrastructure.		IS infrastructure.
		CmpCmpt03	BD computing requirements are supported by the organizational IT department resources.	(Lin, 2008)	Implementation of e-business is supported by the organizational IT human resources.
		CmpIntrfc01	BD storage interface requirements align the current internal Information Systems' applications at my organization.	(Lin, 2008)	Implementation of e-business does not contradict the current internal IS applications.
		CmpIntrfc02	BD storage interface requirements are supported by the existing Information Systems' infrastructure.	(Lin, 2008)	Implementation of e-business is supported by the existing IS infrastructure.
		CmpIntrfc03	BD storage interface requirements are supported by the organizational IT department resources.	(Lin, 2008)	Implementation of e-business is supported by the organizational IT human resources.
Organizational	Open-Source Software	Oss01	My organization adopts Big Data open-source software wherever possible.	(Y. Li, Tan, Teo, & Siow, 2005)	Our organization adopts open-source software wherever possible.
		Oss02	Given a choice, my organization prefers to use Big Data	(Y. Li et al., 2005)	Given a choice, my organization prefers to use open-source

			open-source software in the near future.		software in the near future (i.e., within one year).
		Oss03	My organization is likely to adopt open-source software in the near future.	(Y. Li et al., 2005)	Our organization is likely to adopt open-source software in the near future (i.e., within one year).
	Enterprise Source Software	Ess01	My organization adopts Big Data enterprise source software wherever possible.	(Y. Li et al., 2005)	Our organization adopts open-source software wherever possible.
		Ess02	Given a choice, my organization prefers to use Big Data enterprise source software in the near future.	(Y. Li et al., 2005)	Given a choice, my organization prefers to use open-source software in the near future (i.e., within one year).
		Ess03	My organization is likely to adopt enterprise source software in the near future.	(Y. Li et al., 2005)	Our organization is likely to adopt open-source software in the near future (i.e., within one year).
	Perceived Cost of Adopting Big Data	Cst01	My organization cannot afford the cost of adopting Big Data	(Lee & Kozar, 2008)	I cannot afford the cost of adopting anti-spyware software
		Cst02	Adopting Big Data is expensive	(Lee & Kozar, 2008)	Adopting anti-spyware software is expensive
		Cst03	Big Data adoption can result in a high level of total	(Hanafizadeh & Zare Ravasan, 2018)	ITO can result in a high level of total cost of

			cost of ownership in my organization.		ownership in our business.
Environmental	Perceived Industry Pressure	IndstrPrsr01	Big Data is requested by important business partners.	(Nam et al., 2015)	Requested by important business partners.
		IndstrPrsr02	Big Data is requested by the majority of business partners.	(Nam et al., 2015)	Requested by majority of business partners.
		IndstrPrsr05	Important competitors using or soon to be using Big Data.	(Nam et al., 2015)	Important competitors using or soon to be using ValuNet.
		IndstrPrsr06	Majority of competitors using or soon to be using Big Data.	(Nam et al., 2015)	Majority of competitors using or soon to be using ValuNet.
	Legislation Barriers	LgstBr01	Business laws do not support Big Data.	(Gibbs & Kraemer, 2004)	Business laws do not support e-commerce.
		LgstBr02	Inadequate legal protection for Big Data.	(Gibbs & Kraemer, 2004)	Inadequate legal protection for internet purchases.
		LgstBr03	Tight, inconsistent, or changing laws related to Big Data.	(Siepmann & Nicholas, 2018)	Tight, inconsistent, or changing laws related to the organic certification.
		LgstBr04	Lack of Government guidelines for Big Data.	(S. Wong & Gray, 2019)	Lack of Government guidelines.
		MktTrbls021	Competition in our market is cutthroat.	(G. Wang, Dou, Zhu, & Zhou, 2015)	Competition in our market is cutthroat.
		MktTrbls02	BD in our industry is changing rapidly.	(G. Wang et al., 2015)	The technology in our industry is changing rapidly.

		MktTrbls03	Customers tend to look for new products all the time.	(Jaworski & Kohli, 1993)	Customers tend to look for new products all the time.
		BDInt01	Adoption of BD will be a strategic goal of the organization to enhance the company's competitive advantage.	(RUI, 2007)	The active application and development of e-commerce will be a strategic weapon of the company to enhance the company's competitive advantage.
		BDInt02	My organization will invest more resources (e.g., human, hardware, and financial resources) in the adoption of BD.	(RUI, 2007)	Your company will invest more resources (e.g., human, hardware, and financial resources) in the application and development of e-commerce.
		BDInt03	The adoption of BD will be an important business strategy in the near future.	(RUI, 2007)	The active application and development of e-commerce will be your company's important business strategy in the near future.

Table 32 Constructs, items, and their sources

APPENDIX C. SURVEY'S QUESTIONS

As an example, we illustrate with a simple case having four constructs and fifteen items developed as potential measures for each.

1- Demographic Information

   a. Name

   b. Email

   c. Role in Organization

      i. Academic

      ii. Architect

      iii. CIO/CTO

      iv. Consultant

      v. Developer

      vi. IT Engineer

      vii. Manager

      viii. Network Engineer

      ix. Quality Assurance

      x. Security Engineer

      xi. Storage Engineer

      xii. Student

xiii. System Integrator

xiv. VP

xv. Other. Please specify

d. Industry

i. Advertising /Marketing /PR

ii. Business Services

iii. Cloud Services

iv. Education/Academia

v. Financial Services

vi. Government Federal, State or Local

vii. Health Care

viii. IT Consulting

ix. Manufacturing

x. Media

xi. Retail, Wholesale, Distribution

xii. Social Media

xiii. Storage Services

xiv. Telecommunication

xv. Travel and Leisure

xvi. Transportation

xvii. Utilities

xviii. Other. Please specify

e. Location:

i. Asia Pacific

ii. Australia & New Zealand

iii. Central and South America

iv. Europe

v. Middle East and Africa

vi. North America

vii. South Asia

viii. Other. Please specify

f. Organization's Size (Number of Employees)

i. Less than 50

ii. 50 to 249

iii. 250 Employees or more

2- Technical Background Information

a. Does the organization have a storage system?

b. Have you adopted or are you planning to adopt Big Data?

3- TOE Model Questions: (See Appendix B)

4- Follow-up Questions

a. Would you like to receive a copy of this research once it is published?

Thank you for participating in the survey.

APPENDIX D: IRB CONSENT FORM

INFORMATION SHEET FOR PARTICIPATION IN RESEARCH STUDY

Empirical Assessment of Big Data Technology Adoption Factors for Organizations with Data Storage Systems

Principal Investigator: Ahmad B. Alnafoosi, College of Computing/CDM/Ph.D. Candidate

Institution: DePaul University, USA

Faculty Advisor: Theresa Steinbach, Ph.D. College of Computing/CDM

We are conducting a research study because we are trying to learn more about factors that affect the adoption of Big Data technology in organizations with non-cloud data storage systems.
We are asking you to participate in the research because you have worked or have academic experience in Big Data technology and data storage systems. You must be age 18 or older to be in this study. This study is not approved for the enrollment of people under the age of 18.

Phase 1- Interview
If you agree to be in this study, you will be asked to complete an interview on the phone or teleconferencing (skype or zoom). The interview will include questions about factors affecting the decision to adopt Big Data technology on data storage systems. The interview will also collect some personal information about you such as name, email. The personal information will not be shared and will not be published. Other biographical information such as job title, industry, country, and the number of employees in the organization will also be collected. If there is a question you do not want to answer, you may skip it. This conversation is being recorded for research purposes. Please let me know now if you do not agree to being recorded. You may request that the recording stop at any time. This audio recording will be used to assist the researcher to take notes. The audio file will be stored on password protected at DePaul storage system and will be accessed only by the researcher and the faculty sponsor. The audio files will be destroyed after 36 months. The study should take about 60 minutes to complete.

Phase 2- Pilot Survey
If you agree to be in this study, you will be asked to complete an online survey and provide feedback on its clarity and validity. The survey will include questions about factors affecting the decision to adopt Big Data technology on data storage systems. The survey will also collect some personal information about you such as name, email. Personal information will not be shared or published. Biographical information such as job title,

industry, country, and the number of employees in the organization will also be collected. If there is a question you do not want to answer, you may skip it. The study should take about 30 to 60 minutes to complete.

Phase 3- Large Scale Survey
If you agree to be in this study, you will be asked to complete an online survey. The survey will include questions about factors affecting the decision to adopt Big Data technology on data storage systems. The interview will also collect some personal information about you such as name, email, and job title. Personal information will not be shared or published. Biographical information such as job title, industry, country, and the number of employees in the organization will also be collected. If there is a question you do not want to answer, you may skip it. The study should take about 30 to 45 minutes to complete.

Research data collected from you will be de-identified right after collection. Data analysis will use unidentifiable data. When you first give us your information it will be linked to you with a code number, and we will have a key that tells us who that code number belongs to. So, for a period, it is possible to link this information to you. However, we have put some protections in place, such as storing the information in a secured computer under password protection and with encrypted files. After the study is completed (in about 36 months), we will remove all the identifiers and make the data de-identified. The data will be kept for an undetermined period in the de-identified way, since there should be no risk to you should someone gain access to the data.

Your participation is voluntary, which means you can choose not to participate. There will be no negative consequences if you decide not to participate or change your mind later after you begin the study.
You can withdraw your participation at any time prior to submitting your survey. If you change your mind later while answering the survey, you may simply exit (or not hand in) the survey. You have a choice NOT to include your email in the survey. Once you submit your responses to me directly (or online), I will be unable to remove your data later from the study because all data is anonymous, if you choose not to include your email, and I will not know which survey response belongs to you."
You can withdraw your participation at any time, by contacting me at: Ahmad Alnafoosi aalnafoo@mail.depaul.edu or call me at (847) 920-4987. Since the information you gave me is still identifiable and linked to your email (or other direct identifier), I can remove your data from the research at any time.]

If you have questions, concerns, or complaints about this study or you want to get additional information or provide input about this research, please contact Ahmad Alnafoosi aalnafoo@mail.depaul.edu or call me at (847) 920-4987 or my faculty sponsor Theresa Steinbach tsteinbach@cdm.depaul.edu.

If you have questions about your rights as a research subject, you may contact Susan Loess-Perez, DePaul University's Director of Research Compliance, in the Office of Research Services at 312-362-7593 or by email at sloesspe@depaul.edu. You may also contact DePaul's Office of Research Services if:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.

You may keep this information for your records.

By completing the Interview/survey you are indicating your agreement to be in the research.

I have explained the study to you. You are providing your affirmative agreement verbally to be in the research.

APPENDIX E. RESEARCH ADVERTISEMENT

Big Data holds great promise of unlocking unrealized value in stored data. Yet, Big Data adoption has not been as widespread as many would have thought. This research study attempts to learn more about factors that affect the adoption of Big Data technology in organizations with non-cloud data storage systems.
We are asking for participants in this research who worked or have academic experience in Big Data technology and data storage systems.
Share your thoughts in this survey and get a
1- Chance to win one of 3 Apple iPad Air. This is part of a drawing for participants in this survey.
2- Be the first to get the results of the survey on what are the key factors that are enabling or impeding Big Data adoption on data storage systems.
This is part of a PhD research study in the information system at DePaul University. This work explores the key factors for organization adoption of big data.

You must be age 18 or older to be in this study. This study is not approved for the enrollment of people under the age of 18.

For more information, please follow the following link (Link to either Interview Recruitment OR survey online link)

Research Title: **Empirical Assessment of Big Data Technology Adoption Factors for Organizations with Data Storage Systems**

Principal Investigator: Ahmad B. Alnafoosi, College of Computing/CDM/Ph.D. Candidate, aalnafoo@mail.depaul.edu

Institution: DePaul University, USA

Faculty Advisor: Theresa Steinbach, Ph.D. College of Computing/CDM, tsteinbach@cdm.depaul.edu

APPENDIX F. RESEARCH VERBAL SCRIPTS

Hi, my name is Ahmad Alnafoosi. I am a PhD candidate student in DePaul University's College of Computing and Digital Media. I am conducting research on factors affecting Big Data adoption for organizations with non-cloud data storage systems. I am reaching out to you since I believe that you have work/academic experience in Big Data technology and data storage systems. I am looking for participants in this research (Interview or Survey).

As you know, Big Data holds great promise of unlocking unrealized value in stored data. Yet, Big Data adoption has not been as widespread as many would have thought. This research study attempts to learn more about factors that affect the adoption of Big Data technology in organizations with non-cloud data storage systems.

Share your thoughts in this survey and get a

1- Chance to win one of 3 Apple iPad Air. This is part of a drawing for participants in this survey.

2- Be the first to get the results of the survey on what are the key factors that are enabling or impeding Big Data adoption on data storage systems.

You must be age 18 or older to be in this study.

Here is the information to contact me aalnafoo@mail.depaul.edu , by phone at (847) 920-4987 or my faculty sponsor Theresa Steinbach tsteinbach@cdm.depaul.edu.

APPENDIX G. RESEARCH ONLINE POSTINGS

Big Data holds great promise of unlocking unrealized value in stored data. Yet, Big Data adoption has not been as widespread as many would have thought. This research study attempts to learn more about factors that affect the adoption of Big Data technology in organizations with non-cloud data storage systems.
We are asking for participants in this research who worked or have academic experience in Big Data technology and data storage systems.
Share your thoughts in this survey and get a
1- Chance to win one of 3 Apple iPad Air. This is part of a drawing for participants in this survey.
2- Be the first to get the results of the survey on what are the key factors that are enabling or impeding Big Data adoption on data storage systems.

You must be age 18 or older to be in this study. This study is not approved for the enrollment of people under the age of 18.

For more information, please follow the following link (Link to either Interview Recruitment OR survey online link)
Research Title: **Empirical Assessment of Big Data Technology Adoption Factors for Organizations with Data Storage Systems**

Principal Investigator: Ahmad B. Alnafoosi, College of Computing/CDM/Ph.D. Candidate, aalnafoo@mail.depaul.edu

Institution: DePaul University, USA

Faculty Advisor: Theresa Steinbach, Ph.D. College of Computing/CDM, tsteinbach@cdm.depaul.edu

APPENDIX H. PLS-SEM INTERMEDIATE STEPS

Initial run of PLS algorithm at outer loadings for indicator reliability running factor PLS-SEM analysis using Smart PLS (V 3.3.3) suggested there is a reliability issue for Cost and Legislative Barrier indicators (observable variables). PLS algorithm was run with factor weighting scheme, 500 maximum iterations and stop criterion value of 7.

As discussed in the previous chapter, internal consistency reliability is the measure of multiple indicators to agree in measuring latent variable (Sarstedt & Mooi, 2014). It is assessed by multiple measures. First measure and on the lower end is Cronbach's alpha. Cronbach's alpha acceptable range between 0.6 – 0.95. Value of 0.6 is acceptable for exploratory research like this one. Non exploratory research requires higher threshold of 0.7 for Cronbach's alpha (J. F. Hair et al., 2014). The second reliability measurement is composite reliability. Composite reliability with acceptable values between 0.7 and 0.95 (J. F. Hair et al., 2014). Composite reliability shows low values for these two constructs of Cost and Legislation Barrier. Values outside the acceptable range are highlighted by red. See Table 33 for construct reliability and validity.

The results of that run also show that, there are some indicators that have some convergent validity issues. Convergent validity consists of the outer loading (with acceptable range of > 0.708) and AVE (with acceptable range of > 0.5) (J. F.

Hair et al., 2014). Outer loading values of 0.708 and higher is recommended and values of less than 0.4 can be considered for removal and the impact on validity needs to be examined. Cst1 (Cost) indicator has outer loading value of -0.042 and Lgs3 (Legislation Barrier) indicator has outer loading of 0.241. Values outside the acceptable range are highlighted by red. For AVE see Table 33 and outer loadings see Table 34.

Since the above findings require measurement model modifications, discriminant validity will not be conducted on this mode. It will be conducted on the modified mode. PLS-SEM process is dynamic where modification of model is permitted in these cases.

Construct	Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
BD Adoption	0.799	0.882	0.714
Compute Compatibility	0.891	0.932	0.821
Cost	0.684	0.508	0.366
Enterprise Source SW	0.893	0.933	0.824
Industry Pressure	0.821	0.881	0.65
Latency Compatibility	0.848	0.906	0.763
Legislation Barrier	0.783	0.666	0.445
Market Turbulence	0.707	0.837	0.631
Open Source SW	0.884	0.928	0.812
Storage Interface Compatibility	0.903	0.939	0.838

Table 33 Phase III – Initial Construct Reliability and Validity

Indicator	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
Adp1	0.897									
Adp2	0.797									
Adp3	0.836									
Cmp1		0.924								
Cmp2		0.903								
Cmp3		0.891								
Cst1			-0.042							
Cst2			0.558							
Cst3			0.885							
ESS1				0.895						
ESS2				0.921						
ESS3				0.907						
IPr1					0.781					
IPr2					0.802					
IPr3					0.811					
IPr4					0.831					
Inf1										0.901
Inf2										0.925
Inf3										0.92
Lat1						0.892				
Lat2						0.896				
Lat3						0.83				
Lgs1							0.909			
Lgs2							0.671			
Lgs3							0.241			
Mkt1								0.818		
Mkt2								0.805		
Mkt3								0.759		
OSS1									0.899	
OSS2									0.889	
OSS3									0.915	

Table 34 Phase III – Initial Indicator Reliability

The researcher examined more closely Cst1. Looking at Cost construct's three indicators, it looked like Cst1 is asking more about affordability than cost. Thus, the researcher decided to remove Cst1 from the mod. See the indicators below for Cost construct.

~~Cst1- My organization can not afford the cost of adopting BD.~~

Cst2- Adopting BD is expensive.

Cst3- BD adoption can result in a high level of total cost of ownership in my organization

The researcher examined more closely Lgs3. Looking at legislation barrier construct's three indicators, it looked like Lgs3 is asking more about legislation tightness, inconsistency and change than supporting BD. Thus, the researcher decided to remove Lgs3 from the model. See the indicators below for Legislation barrier construct.

Lgs1- Business laws do not support BD.

Lgs2- There is inadequate legal protection for BD.

~~Lgs3- Business laws related to BD are tight, inconsistent, or changing.~~

After removing Cst1 and Lgs3 from the measurement model, PLS algorithm was run again with the same parameters as before. The modified measurement model construct internal reliability has improved for all constructs in Cronbach's alpha and composite reliability except for Cost construct. Cost construct has

dropped slightly, but still above 0.60 threshold for exploratory research. Thus, cost reliability is acceptable for this exploratory study. In addition, the internal reliability consists of both measures (Cronbach's alpha in the lower end and composite reliability on the higher end), cost construct is within the limit for internal reliability. See Table 35.

Construct	Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
BD Adoption	0.799	0.882	0.714
Compute Compatibility	0.891	0.932	0.821
Cost	0.652	0.827	0.71
Enterprise Source SW	0.893	0.933	0.824
Industry Pressure	0.821	0.881	0.65
Latency Compatibility	0.848	0.906	0.763
Legislation Barrier	0.717	0.862	0.76
Market Turbulence	0.707	0.837	0.631
Open-Source SW	0.884	0.928	0.812
Storage Interface Compatibility	0.903	0.939	0.838

Table 35 Phase III – Modified Model Construct Reliability and Validity

Convergent validity consisting of indicator reliability/outer loadings and AVE is now within the acceptable range (over 0.5) for all indicators for outer loadings and across all constructs for AVE. For AVE see Table 35 and outer loadings see Table 36.

Indicator	BD Adoption	Compute Comp.	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
Adp1	0.897									
Adp2	0.799									
Adp3	0.835									
Cmp1		0.924								
Cmp2		0.903								
Cmp3		0.891								
Cst2			0.702							
Cst3			0.963							
ESS1				0.895						
ESS2				0.921						
ESS3				0.907						
IPr1					0.781					
IPr2					0.803					
IPr3					0.811					
IPr4					0.831					
Inf1										0.901
Inf2										0.925
Inf3										0.92
Lat1						0.892				
Lat2						0.897				
Lat3						0.83				
Lgs1							0.955			
Lgs2							0.78			
Mkt1								0.818		
Mkt2								0.804		
Mkt3								0.759		
OSS1									0.899	
OSS2									0.889	
OSS3									0.915	

Table 36 Phase III – Modified Model Indicator Reliability

Discriminant validity is the extent which a construct is different from other constructs by empirical standards (J. F. Hair et al., 2014). This can be measured using multiple measurements of cross-loadings, Fornell Larcker criterion, and/or

Heterotrait Monotrait Ratio (HTMT) (J. F. Hair et al., 2014). The preference according to Hair et al., 2014 to have more sensitivity discriminant validity and HTMT being the current preferred one. This paper will examine all three.

Cross loading is usually the first approach used for discriminant validity. It lists all the indicators outer loadings in a table and expect the indicator outer loading to be highest value for the latent variable it is measuring (Joseph F Hair Jr et al., 2016). The highest values in each column are highlighted and has **bold font and highlighted.** The indicators loadings values are highest for each latent variable they are measuring. Thus, there is no discriminant validity issue from cross-loadings perspective. See Table 37.

Indicator	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
Adp1	**0.897**	0.239	0.225	0.358	0.595	0.222	-0.125	0.424	0.29	0.25
Adp2	**0.799**	0.177	0.147	0.324	0.491	0.141	0.022	0.365	0.31	0.185
Adp3	**0.835**	0.231	0.199	0.322	0.557	0.189	-0.172	0.346	0.187	0.223
Cmp1	0.253	**0.924**	-0.066	0.026	0.193	0.668	-0.38	0.095	0.049	0.748
Cmp2	0.221	**0.903**	-0.094	-0.012	0.198	0.667	-0.359	0.085	0.037	0.734
Cmp3	0.221	**0.891**	-0.164	0.004	0.175	0.658	-0.447	0.102	-0.059	0.732
Cst2	0.093	-0.16	**0.702**	0.023	0.087	-0.104	0.3	0.123	0.111	-0.176
Cst3	0.245	-0.083	**0.963**	0.234	0.183	-0.067	0.231	0.199	0.186	-0.119
ESS1	0.347	-0.018	0.152	**0.895**	0.315	-0.004	0.092	0.364	0.16	-0.016
ESS2	0.351	0.018	0.164	**0.921**	0.272	0.001	0.033	0.305	0.087	0.024
ESS3	0.381	0.02	0.217	**0.907**	0.261	-0.013	0.033	0.3	0.14	0.003
IPr1	0.505	0.163	0.138	0.286	**0.781**	0.115	-0.043	0.243	0.19	0.142
IPr2	0.505	0.135	0.12	0.28	**0.803**	0.128	-0.039	0.373	0.252	0.108
IPr3	0.538	0.194	0.156	0.214	**0.811**	0.224	-0.147	0.389	0.228	0.187
IPr4	0.549	0.178	0.151	0.226	**0.831**	0.174	-0.1	0.423	0.261	0.196
Inf1	0.224	0.743	-0.156	0.002	0.205	0.618	-0.416	0.071	0.005	**0.901**
Inf2	0.242	0.748	-0.145	0.004	0.162	0.637	-0.4	0.077	-0.026	**0.925**
Inf3	0.25	0.746	-0.116	0.005	0.178	0.662	-0.417	0.097	-0.05	**0.92**
Lat1	0.195	0.63	-0.04	0.024	0.163	**0.892**	-0.439	0.072	-0.095	0.587
Lat2	0.227	0.645	-0.071	0.008	0.193	**0.897**	-0.3	0.093	-0.017	0.626
Lat3	0.138	0.662	-0.138	-0.069	0.167	**0.83**	-0.442	0.058	-0.085	0.632
Lgs1	-0.122	-0.421	0.232	0.044	-0.123	-0.445	**0.955**	-0.003	0.177	-0.43
Lgs2	-0.058	-0.329	0.293	0.067	-0.03	-0.281	**0.78**	0.018	0.148	-0.347
Mkt1	0.371	0.086	0.174	0.342	0.382	0.047	0.032	**0.818**	0.217	0.103
Mkt2	0.37	0.071	0.134	0.277	0.393	0.095	-0.064	**0.804**	0.148	0.057
Mkt3	0.327	0.092	0.17	0.22	0.277	0.067	0.048	**0.759**	0.209	0.052
OSS1	0.298	0.013	0.16	0.164	0.304	-0.067	0.198	0.266	**0.899**	-0.025
OSS2	0.249	-0.013	0.173	0.091	0.213	-0.066	0.132	0.193	**0.889**	-0.043
OSS3	0.287	0.029	0.171	0.124	0.257	-0.053	0.168	0.184	**0.915**	-0.007

Table 37 Phase III – Modified Model Cross Loadings Criterion

Fornell Larcker criterion was used by comparing the square root of AVE with its correlation to other constructs. The expected values for the construct square

root of AVE to itself should be higher than all other correlation with other constructs. That is the case with this analysis the highest values are highlighted and has **bold font and highlighted**. Thus, there is no discriminant validity issue from Fornell Larcker criterion perspective. See Table 38.

Construct	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open-Source SW	Storage Interface Compatibility
BD Adoption	**0.845**									
Compute Compatibility	0.257	**0.906**								
Cost	0.228	-0.117	**0.843**							
Enterprise Source SW	0.397	0.008	0.197	**0.908**						
Industry Pressure	0.651	0.208	0.176	0.31	**0.806**					
Latency Compatibility	0.221	0.733	-0.087	-0.006	0.2	**0.873**				
Legislation Barrier	-0.113	-0.435	0.28	0.057	-0.104	-0.436	**0.872**			
Market Turbulence	0.449	0.104	0.2	0.355	0.445	0.088	0.004	**0.794**		
Open-Source SW	0.31	0.012	0.186	0.142	0.289	-0.069	0.186	0.24	**0.901**	
Storage Interface Compatibility	0.261	0.814	-0.151	0.004	0.198	0.699	-0.449	0.09	-0.027	**0.915**

Table 38 Phase III – Modified Model Fornell Larcker Criterion

According to Hair, et al., 2014, the current best approach to discriminant validity is Heterotrait-Monotrait Ratio (HTMT). HTMT measures the ratio of the between-trait correlations to the within-trait correlation. In other words, it is the measure of similarities between latent variables. Higher HTMT value indicate higher similarity and lack of discriminant validity. Value of 0.85 is suggested as threshold for conceptually distinct model and value of 0.90 is suggested for conceptually similar models (as in the case of this research for technical compatibility) (J. F. Hair et al., 2014). All the HTMT values of this model are

below 0.9 value. Except for, compute compatibility vs latency compatibility HTMT value is 0.908. See Table 39.

Construct	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW
BD Adoption									
Compute Compatibility	0.302								
Cost	0.269	0.187							
Enterprise Source SW	0.469	0.026	0.196						
Industry Pressure	0.801	0.243	0.214	0.365					
Latency Compatibility	0.256	0.848	0.141	0.049	0.237				
Legislation Barrier	0.162	0.533	0.474	0.079	0.113	0.542			
Market Turbulence	0.595	0.132	0.278	0.445	0.577	0.11	0.076		
Open-Source SW	0.368	0.059	0.228	0.157	0.335	0.089	0.229	0.302	
Storage Interface Compatibility	0.305	**0.908**	0.224	0.019	0.229	0.802	0.547	0.112	0.04

Table 39 Phase III – Modified Model HTMT

Per Henseler et al. (2015), the first step in addressing HTMT discriminant issue is to preserve the construct and to attempt to increase the average monotrait-heteromethod correlation and/or decreasing the average heteromethod-heterotrait correlation of the construct measures. This is done by indicator that has the lowest correlation with the other indicators measuring the same item (Henseler et al., 2015). In this model, the lowest correlation indicator is cmp3 for compute compatibility construct see Table 37 above. Thus, cmp3 was removed from the model and HTMT was recalculated and that lowered all the HTMT values below 0.9 threshold see Table 40. Since an indicator was removed from the model all the above measurement model evaluations were re-done and all the values are similar

with the same conclusions. Thus, this model will be called the final measurement mode.

Construct	BD Adoption	Compute Compatibility	Cost	Enterprise Source SW	Industry Pressure	Latency Compatibility	Legislation Barrier	Market Turbulence	Open Source SW	Storage Interface Compatibility
BD Adoption										
Compute Compatibility	0.303									
Cost	0.269	0.152								
Enterprise Source SW	0.469	0.025	0.196							
Industry Pressure	0.801	0.247	0.214	0.365						
Latency Compatibility	0.256	0.831	0.141	0.049	0.237					
Legislation Barrier	0.162	0.491	0.474	0.079	0.113	0.542				
Market Turbulence	0.595	0.124	0.278	0.445	0.577	0.11	0.076			
Open Source SW	0.368	0.051	0.228	0.157	0.335	0.089	0.229	0.302		
Storage Interface Compatibility	0.305	0.894	0.224	0.019	0.229	0.802	0.547	0.112	0.04	

Table 40 Phase III – Final Model HTMT

APPENDIX I. HOLDOUT SAMPLE ANALYSIS

Step 3- Standardize the Holdout Sample

Lat1Std	Lat2Std	Lat3Std	OSS1Std	OSS2Std	OSS3Std	IPr1Std	IPr2Std	IPr3Std	IPr4	Adp1Std	Adp2Std	Adp3Std	Cmp1Std	Cmp2Std	Cst2Std	Cst3Std	Lgs1Std	Lgs2Std	Inf1Std	Inf2Std	Inf3Std	Ess1Std	Ess2Std	Ess3Std	Mkt1Std	Mkt2Std	Mkt3Std
-0.1	-0.7	1.3	0.2	0.2	0.18	0.8	1	-0.3	-0	0.61	0.88	-1.5	-0.1	0.69	1.29	0.2	-0	-1.2	-1.16	0.7	-0	0.2	-1	0.2	-0.1	-0.2	-0.1
-0.1	0.7	0.6	-1.4	-1.3	-1.3	-0	0	0.7	1	-0.4	0.88	-0.4	-0.1	0.69	0.35	0.2	-0.9	-0.3	0.8	0.7	0.7	0.2	0.2	0.2	0.9	-0.2	-0.1
-1.8	-1.4	-1.6	0.2	0.96	0.92	-0	0	-0.3	-0	-0.4	-0.2	0.67	-1.5	-1.5	0.35	0.2	0.8	0.6	-1.6	-1	-2	0.2	0.2	1.1	0.9	-0.2	0.96
1.5	-0.7	1.3	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	1.38	1.42	-0.6	1.2	0.8	-0.3	0	0.7	0.7	0.2	0.2	-1.8	-0.1	-0.2	-0.1
-1.8	0.7	0.6	0.99	0.96	0.92	-2	0	-0.3	-0	0.61	-0.2	0.67	1.38	1.42	0.35	1.2	-0	0.6	0.8	0.7	0.7	-2	0.2	-0.8	0.9	-0.2	0.96
0.7	0.7	-0.1	-0.6	0.2	0.18	-0	0	-0.3	1	0.61	0.88	0.67	0.65	0.69	-1.5	-1.8	-0.9	-0.3	0.8	0.7	1.5	1.2	1.1	1.1	0.9	-0.2	-0.1
0.7	0.7	0.6	-1.4	-1.3	0.18	-0	-1	-2.5	-2	-1.5	-2.5	-2.6	0.65	-0.8	0.35	1.2	-0	-0.3	0.8	0.7	0.7	-1	-1	-0.8	-0.1	-0.2	-1.1
-1	1.4	-1.6	-0.6	-1.3	-2.1	-3	-1	-0.3	1	0.61	0.88	-3.7	1.38	-0	-0.6	-1.8	-0	1.5	-0.8	-1	-2	0.2	-1	1.1	-2.1	-1.3	-0.1
-0.1	-0.7	-0.1	0.99	0.96	0.92	-0	-2	0.7	-0	0.61	0.88	0.67	1.38	1.42	0.35	1.2	-0.9	0.6	-0.8	-1	0.7	0.2	-2	1.1	-1.1	-2.5	0.96
-1.8	-1.4	-1.6	0.99	0.96	0.18	0.8	0	-0.3	-0	0.61	-0.2	-0.4	-1.5	-0.8	1.29	1.2	0.8	0.6	-0.8	-1	-2	-1	1.1	0.2	-0.1	-0.2	-0.1
1.5	1.4	1.3	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	1.38	1.42	-1.5	1.8	1.8	-2.2	1.6	1.5	1.5	1.2	1.1	1.1	0.9	0.89	0.96
-1	-0.7	-0.8	0.2	0.96	0.92	-0	1	-0.3	1	-0.4	0.88	-0.4	-0.8	-0.8	0.35	1.2	0.8	1.5	-1.6	-1	-1	0.2	1.1	0.2	-0.1	0.89	-0.1
0.7	-0.7	0.6	-1.4	-2.1	-1.3	0.8	1	0.7	1	0.61	-1.3	0.67	-0.1	-0.8	0.35	0.2	-0	0.6	0.8	-1	0.7	0.2	-2	0.2	-0.1	-2.5	-0.1
-1	-0.7	-0.8	0.2	0.2	0.18	0.8	1	0.7	1	0.61	0.88	0.67	-0.8	-0.8	0.35	0.2	1.7	1.5	-0.8	-1	-1	1.2	1.1	1.1	-0.1	-0.2	-0.1
0.7	0	0.6	-1.4	-1.3	-1.3	-0	1	0.7	-2	0.61	-1.3	-1.5	-0.1	1.42	-0.6	-0.8	-0	-1.2	0	0.7	0.7	0.2	1.1	0.2	0.9	-0.2	0.96
-0.1	0	-0.1	-2.2	-2.1	-2.1	-2	-1	-1.4	-2	-1.5	-1.3	-1.5	-0.1	-0	-0.6	-0.8	-0	-0.3	0	0	-0	-1	-1	-0.8	-1.1	-1.3	-1.1
-1	-1.4	-1.6	0.99	0.2	0.18	-0	0	0.7	-0	-0.4	0.88	-0.4	-1.5	-1.5	0.35	0.2	0.8	0.6	-1.6	-1	-2	0.2	0.2	0.2	-0.1	-0.2	-0.1
0.7	0	1.3	0.99	0.96	0.92	0.8	0	0.7	1	-0.4	-0.2	0.67	1.38	0.69	0.35	1.2	-0	0.6	0	0.7	0.7	-2	-1	-0.8	-0.1	0.89	-0.1
-1	-1.4	0.6	0.99	0.96	0.92	0.8	0	0.7	1	0.61	0.88	0.67	-0.8	-0.8	0.35	0.2	-0.9	1.5	-0.8	0.7	0.7	0.2	0.2	1.1	0.9	0.89	0.96
-1.8	-1.4	-1.6	-0.6	0.2	-2.1	-3	-2.9	-2.5	1	-0.4	-3.6	-1.5	-1.5	-1.5	-2.5	-2.7	-1.8	-2.2	-1.6	-1	-2	-3	-3	-2.7	-3.1	-3.6	-3.1

0.7	0.7	0.6	0.99	0.96	0.92	0.8	0	0.7	1	0.61	0.88	0.67	1.38	1.42	0.35	1.2	0.8	-0.3	0.8	0.7	1.5	1.2	1.1	0.2	-0.1	0.89	-0.1
-0.1	-0.7	1.3	0.2	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	-0.8	-0.8	0.35	0.2	-1.8	-0.3	-0.8	-1	-1	1.2	1.1	1.1	0.9	0.89	-0.1
-1	-0.7	-1.6	0.99	0.96	0.92	-0	0	-0.3	-0	0.61	0.88	0.67	-1.5	-1.5	-1.5	0.2	-1.8	-2.2	-1.6	-1	-2	0.2	0.2	0.2	-0.1	-0.2	-1.1
-1	-0.7	-0.8	0.2	0.2	0.18	-0	0	-0.3	-0	-0.4	-0.2	-0.4	-0.8	-0.8	0.35	0.2	0.8	0.6	-0.8	-1	-1	0.2	1.1	0.2	-0.1	-0.2	-0.1
1.5	1.4	1.3	-2.2	-2.1	-2.1	0.8	1	-0.3	-2	0.61	0.88	0.67	1.38	1.42	1.29	-0.8	-0	-0.3	1.6	1.5	1.5	-1	-1	-0.8	-0.1	-1.3	-1.1
1.5	1.4	1.3	-0.6	-0.6	-0.6	-2	-1	-0.3	-0	0.61	-0.2	0.67	-0.1	-0	0.35	0.2	-0	-0.3	0.8	0.7	0.7	0.2	0.2	0.2	-1.1	-0.2	-2.1
0.7	-0.7	0.6	-1.4	-1.3	-1.3	-0	-2	-2.5	-3	0.61	-0.2	-0.4	-0.8	-0.8	1.29	0.2	0.8	1.5	0.8	-1	-1	0.2	0.2	0.2	0.9	-2.5	-2.1
0.7	0	0.6	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	0.65	0.69	-1.5	0.2	-0	0.6	0	0.7	0.7	0.2	0.2	0.2	0.9	0.89	0.96
0.7	0.7	0.6	0.2	0.2	0.18	-2	-1	0.7	1	-1.5	-1.3	-0.4	-0.1	-0	0.35	0.2	-0	0.3	0.8	0.7	0.7	0.2	0.2	0.2	0.9	0.89	0.96
-1.8	-1.4	-0.8	0.2	0.2	0.92	0.8	0	0.3	1	-0.4	0.88	-0.4	-1.5	-0.8	0.35	1.2	1.7	0.6	-1.6	-1	-2	1.2	0.2	0.2	0.9	0.2	-0.1
-1.8	-1.4	-1.6	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	-1.5	-1.5	1.29	1.2	1.7	1.5	-1.6	-1	-2	1.2	1.1	1.1	0.9	0.89	0.96
0.7	0.7	-0.1	0.2	0.96	0.92	-0	0	-0.3	1	0.61	-1.3	0.67	1.38	0.69	0.35	-0.8	-0.9	-1.2	1.6	0.7	0.7	0.2	1.1	1.1	0.9	0.89	0.96
-0.1	0	-0.1	-0.6	-1.3	-1.3	0.8	1	0.7	1	0.61	0.88	0.67	0.65	0.69	-0.6	0.2	-0	-0.3	0	0	-0	0.2	0.2	0.2	-0.1	0.89	-0.1
1.5	0.7	0.6	0.99	-0.6	-0.6	-0	-2	-0.3	1	-0.4	-0.2	0.67	-0.1	0.69	0.35	0.2	0.8	-0.3	0.8	0.7	0.7	0.2	1.1	1.1	0.9	-0.2	0.96
0.7	0.7	1.3	0.2	0.2	-0.6	0.8	0	0.7	-2	0.61	-1.3	0.67	1.38	0.69	0.35	-0.8	-0.9	-0.3	0.8	0.7	1.5	0.7	-1	-1	-0.8	-1.1	-1.3	0.96
-0.1	-0.7	-0.1	0.2	0.2	-0.6	0.8	0	0.7	1	0.61	-0.2	0.67	-0.1	-0	-0.6	0.2	-1.8	-0.3	-0.8	-1	-1	0.2	1.1	1.1	0.9	0.89	0.96
-1	-1.4	-1.6	0.2	0.2	0.92	0.8	0	0.7	-0	0.61	-0.2	-0.4	-1.5	-1.5	1.29	0.2	1.7	0.6	-0.8	-1	-1	0.2	0.2	0.2	0.9	0.89	-0.1
1.5	1.4	1.3	0.99	0.96	0.92	-0	-1	-0.3	-2	0.61	0.88	0.67	0.65	-0.8	0.35	0.2	0.8	0.6	0	0	0	0.2	0.2	0.2	-0.1	-0.2	-0.1
1.5	1.4	1.3	0.99	-0.6	-0.6	0.8	1	0.7	1	0.61	0.88	0.67	1.38	1.42	0.35	0.2	-0	0.3	1.6	1.5	0.7	1.2	0.2	0.2	-0.1	0.89	0.96
-0.1	-0.7	-0.8	0.2	0.2	0.18	0.8	1	0.7	1	0.61	0.88	0.67	-0.8	-0.8	0.35	-0.8	-0	-0.3	-0.8	-1	-0	-1	-2	-0.8	-0.1	0.89	0.96
-0.1	0.7	-0.1	-1.4	-0.6	0.18	-3	-2	-2.5	-3	-3.5	-2.5	-2.6	-1.5	-0.8	1.29	0.2	-0.9	-0.3	0	-1	-1	-3	-2	-0.8	-3.1	-1.3	-1.1
1.5	1.4	1.3	-2.2	-2.1	-2.1	0.8	1	0.7	1	0.61	0.88	0.67	1.38	1.42	0.35	-1.8	-0.9	0.6	1.6	1.5	1.5	1.2	1.1	1.1	-0.1	-0.2	-2.1
0.7	0.7	0.6	0.2	0.2	0.18	0.8	0	-1.4	-0	0.61	0.88	0.67	-0.1	-0	-1.5	-1.8	-0	-0.3	0	0	-0	0.2	-1	0.2	-0.1	-0.2	-0.1
1.5	0.7	0.6	-2.2	-2.1	-2.1	-0	0	-0.3	1	0.61	0.88	0.67	1.38	1.42	-1.5	-1.8	-0.9	-1.2	-0.8	-1	-1	1.2	1.1	1.1	0.9	0.89	0.96

-0.1	0.7	-0.8	-0.6	0.2	0.18	-2	0	-2.5	-0	-2.5	-0.2	-1.5	0.65	-0	-0.6	-0.8	0.8	-0.3	0.8	1.5	-0	-2	0.2	-0.8	-0.1	-2.5	-1.1
0.7	-1.4	1.3	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	-1.5	0.69	-1.5	-1.8	-1.8	-2.2	0.8	0.7	0.7	1.2	1.1	1.1	-1.1	-0.2	-1.1
0.7	0.7	1.3	-1.4	-1.3	-1.3	-0	-1	-0.3	1	-1.5	-0.2	0.67	0.65	-0	-0.6	-0.8	-0	-1.2	1.6	1.5	1.5	0.2	0.2	0.2	0.9	0.89	-0.1
-1	0	-0.8	-1.4	-1.3	-0.6	0.8	1	0.7	-0	-1.5	-1.3	0.67	-1.5	-1.5	-0.6	-0.8	-0	-0.3	-0.8	0	-0	-1	-1	-0.8	-0.1	-0.2	-0.1
0.7	0.7	-0.1	-1.4	-1.3	-0.6	-2	-1	-1.4	-0	-1.5	-0.2	-1.5	0.65	1.42	1.29	-0.8	-0	-0.3	0	0.7	-0	0.2	0.2	0.2	0.9	-1.3	0.96
0.7	0	0.6	-1.4	-1.3	-1.3	-0	-2	-0.3	-2	-0.4	-0.2	-0.4	-0.1	0.69	-0.6	-0.8	1.7	1.5	0	0.7	0.7	-1	-1	-0.8	0.9	0.89	0.96
-0.1	0.7	0.6	0.2	0.2	0.92	-0	-1	0.7	1	0.61	-0.2	0.67	0.65	0.69	-0.6	-0.8	-0	-0.3	-0.8	0.7	-1	0.2	0.2	0.2	-2.1	0.89	-1.1
-1	-0.7	-1.6	0.99	0.96	0.92	-0	0	0.7	-0	0.61	0.88	-0.4	-1.5	-1.5	0.35	1.2	1.7	1.5	-0.8	-1	-1	1.2	0.2	1.1	-1.1	-1.3	-1.1
-1.8	-1.4	-0.8	0.2	0.96	0.18	-0	0	0.7	1	-0.4	-0.2	-0.4	-1.5	-1.5	0.35	1.2	0.8	0.6	-1.6	-1	-1	1.2	1.1	0.2	-0.1	0.89	-0.1
0.7	0.7	-0.8	0.99	0.2	0.18	0.8	1	0.7	1	0.61	0.88	0.67	-0.8	-0.8	0.35	0.2	0.8	1.5	0	-1	0.7	0.2	0.2	0.2	0.9	0.89	-0.1
1.5	1.4	1.3	-0.6	-0.6	-0.6	-0	-1	0.7	-0	-1.5	-0.2	-1.5	-0.8	-0.8	-0.6	-0.8	0.8	0.6	-0.8	-1	0.7	-1	0.2	0.2	-3.1	-1.3	-2.1
-1	-0.7	-0.8	0.99	0.96	0.92	0.8	1	0.7	1	-0.4	-0.2	-0.4	-0.8	-0.8	0.35	0.2	1.7	1.5	0	0	-0	0.2	0.2	0.2	-0.1	-0.2	-0.1
-1	-1.4	-1.6	0.2	0.2	0.92	0.8	0	0.7	-0	0.67	-1.5	-0.8	1.29	0.2	0.8	0.6	-1.6	-1	-1	0.2	1.1	0.2	0.9	-0.2	-1.3	-0.1
-0.1	0.7	-0.1	0.2	0.96	0.92	0.8	1	0.7	1	0.61	-0.2	0.67	0.65	1.42	1.29	1.2	-0.9	-0.3	0	0	0.7	0.2	0.2	0.2	0.9	0.89	-0.1
-0.1	0	-0.1	0.99	0.2	0.92	-0	0	-1.4	-2	-0.4	-1.3	-0.4	0.65	-0	1.29	-0.8	0.8	0.6	0	0	-0	0.2	-1	0.2	-1.1	-1.3	-0.1
-0.1	0.7	-0.8	-2.2	-2.1	-2.1	-0	0	-0.3	-0	-3.5	0.88	-0.4	0.65	0.69	-0.6	1.2	-1.8	-2.2	0.8	0.7	0.7	0.2	0.2	0.2	-2.1	-0.2	0.96
-0.1	1.4	1.3	-2.2	-2.1	-2.1	0.8	1	0.7	1	0.61	0.88	0.67	-0.1	1.42	-2.5	1.2	0.8	0.6	1.6	1.5	1.5	1.2	1.1	1.1	0.9	0.89	0.96
-1	-1.4	-1.6	0.2	0.96	0.92	0.8	1	-0.3	1	0.61	-0.2	0.67	-1.5	-0.8	0.35	1.2	1.7	0.6	-0.8	-1	-1	1.2	0.2	0.2	0.9	0.89	0.96
0.7	0.7	0.6	-0.6	0.2	0.18	-0	0	0.7	1	0.61	-0.2	0.67	0.65	0.69	-1.5	-0.8	0.7	0.7	0.2	0.7	0.7	0.2	1.1	1.1	-0.1	0.89	-1.1
-0.1	-1.4	-1.6	0.2	0.2	0.92	-0	-1	-0.4	-0	-0.4	-0.2	-0.4	1.38	1.42	-0.6	-0.8	1.7	1.5	-1.6	-1	-1	0.2	0.2	0.2	0.9	0.89	0.96
-0.1	1.4	0.6	0.99	0.96	0.92	0.8	1	0.7	1	-0.4	-0.2	0.67	0.65	-0.8	0.35	0.2	-0.9	-1.2	0.8	0.7	0.7	-1	-1	0.2	0.9	0.89	0.96
1.5	-1.4	1.3	0.2	0.2	0.18	-0	-2	-0.3	-0	0.61	-2.5	0.67	0.65	-0	1.29	1.2	-0	-0.3	0.8	0.7	0.7	-1	-1	0.2	0.9	-0.2	-2.1
-1	-1.4	-0.8	0.2	0.2	0.18	-2	0	-0.3	-2	-0.4	-0.2	-1.5	-0.8	-1.5	0.35	0.2	-0	0.6	-1.6	-1	-2	-1	0.2	0.2	-0.1	0.89	-0.1
0.7	-1.4	1.3	0.2	0.96	0.18	0.8	1	0.7	1	0.61	0.88	0.67	-0.8	0.69	0.35	0.2	0.8	0.6	0.8	0.7	0.7	-1	-1	0.2	0.9	0.89	-0.1

0.7	0.7	0.6	0.2	-0.6	-0.6	-0	0	-2.5	-3	-1.5	-0.2	-1.5	0.65	0.69	0.35	0.2	-0	-0.3	0.8	0.7	-1	-2	-2	-1.8	-1.1	-1.3	-1.1
0.7	-0.7	-0.8	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	1.38	1.42	-1.5	0.2	-1.8	-1.2	0	-1	-1	-1	0.2	-0.8	0.9	0.89	-0.1
0.7	0.7	-0.8	0.2	-0.6	-1.3	0.8	0	0.7	1	0.61	0.88	0.67	0.65	0.69	-2.5	-1.8	0.8	-0.3	0.8	0.7	0.7	-1	1.1	1.1	-0.1	-0.2	-0.1
0.7	-0.7	-0.8	0.2	0.2	-0.6	0.8	1	0.7	1	0.61	-0.2	-0.4	0.65	0.69	0.35	1.2	-0	0.6	0	0	-0	0.2	0.2	0.2	0.9	0.89	0.96
-0.1	0.7	0.6	0.99	0.96	0.92	-0	-2	-1.4	-2	-2.5	-1.3	-2.6	-0.1	-0	-0.6	1.2	-0	0.6	0.8	0	0	-0	-3	-3	-2.7	-1.1	-0.2	-1.1
-1	-0.7	-0.8	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	-0.8	-0.8	0.35	0.2	0.8	0.6	-0.8	-1	-1	1.2	1.1	1.1	0.9	0.89	0.96
0.7	0.7	0.6	0.2	0.2	0.18	-0	0	-0.3	-0	-0.4	-0.2	-0.4	0.65	0.69	0.35	-0.8	-0	-1.2	0	0	-1	-1	-1	-0.8	-0.1	-0.2	0.96
-0.1	0.7	-0.1	0.99	0.96	0.92	-0	0	-0.3	-0	0.61	-0.2	0.67	0.65	-0.8	0.35	0.2	-0	-0.3	0	0	-0	-2	0.2	-1.8	-0.1	0.89	0.96
1.5	-1.4	-1.6	0.2	0.2	-0.6	0.8	-1	0.7	1	0.61	-1.3	0.67	-0.1	-1.5	0.35	0.2	-1.8	0.6	1.6	-1	1.5	-1	1.1	1.1	0.9	0.89	0.96
-1.8	-1.4	-1.6	0.2	0.2	0.92	-0	0	0.7	1	-0.4	0.2	-0.4	-1.5	-1.5	0.35	1.2	0.8	0.6	-0.8	-1	-2	0.2	0.2	1.1	-0.1	-0.2	0.96
0.7	0.7	-0.1	0.99	0.96	0.92	0.8	1	0.7	1	0.61	-0.2	0.67	0.65	0.69	-0.6	0.2	-0.9	-0.3	0.8	0.7	-0	-1	-1	-0.8	-0.1	-0.2	0.96
-0.1	1.4	0.6	0.2	0.2	-1.3	0.8	1	0.7	1	0.61	0.88	0.67	0.65	0.69	-1.5	0.2	-0	0.6	0.8	0.7	0.7	-3	-3	-2.7	0.9	0.89	0.96
0.7	1.4	1.3	0.2	-2.1	-0.6	-2	-2	0.7	-0	0.61	-0.2	-0.4	0.65	1.42	1.29	1.2	0.8	1.5	0.8	1.5	1.5	1.2	1.1	1.1	-0.1	0.89	-1.1
1.5	1.4	1.3	0.99	0.96	0.92	0.8	1	-0.3	1	0.61	0.88	0.67	1.38	1.42	1.29	1.2	-1.8	-2.2	1.6	1.5	1.5	1.2	1.1	1.1	0.9	0.89	0.96
0.7	0	-0.1	-2.2	-2.1	-2.1	-2	-2	-1.4	-0	-0.4	0.88	-1.5	-0.1	-0.8	1.29	-0.8	-0.9	-0.3	-0.8	0.7	0.7	1.2	-3	-2.7	-0.1	-1.3	-2.1
-0.1	0	-0.1	-2.2	-2.1	-2.1	-0	0	-0.3	-0	0.61	0.88	-0.4	0.65	0.69	0.35	0.2	-0	0.6	0.8	0.7	-0	0.2	0.2	-2.7	-0.1	-2.5	-0.1
0.7	1.4	1.3	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	1.38	-0	1.29	-0.8	-1.8	-1.2	1.6	1.5	1.5	1.2	1.1	1.1	-1.1	0.89	0.96
-1	0	-0.8	0.99	0.96	0.92	0.8	0	-0.3	1	-0.4	-0.2	-1.5	-1.5	-0.8	1.29	1.2	-0	0.6	-1.6	-1	-1	1.2	1.1	0.2	-0.1	-0.2	-0.1
-1	-0.7	-0.8	0.2	0.2	0.18	0.8	1	0.7	1	0.61	0.88	0.67	-0.8	-0.8	0.35	0.2	0.8	0.6	0.8	-1	-1	0.2	0.2	0.2	-0.1	-0.2	-0.1
0.7	-1.4	0.6	0.99	0.96	0.92	0.8	1	0.7	1	0.61	0.88	0.67	0.65	-1.5	-2.5	0.2	-0.9	-1.2	0.8	0.7	0.7	0.2	0.2	0.2	-0.1	0.89	-2.1
-1	-1.4	-1.6	0.2	0.2	0.92	-0	0	-0.3	1	0.61	-0.2	0.67	-1.5	-1.5	0.35	0.2	0.8	0.6	-1.6	-1	-1	1.2	1.1	0.2	-0.1	0.89	0.96
0.7	0.7	0.6	-0.6	-0.6	-0.6	0.8	1	0.7	1	0.61	-0.2	0.67	0.65	0.69	-1.5	-0.8	-0	-0.3	0.8	0.7	0.7	0.2	0.2	0.2	0.9	0.89	-0.1
-1	0.7	-0.8	-0.6	-0.6	-1.3	-0	-1	-2.5	-3	-2.5	-2.5	-2.6	-0.8	-0.8	0.35	0.2	-0.9	0.6	-0.8	-1	-1	-1	-2	0.2	-0.1	-0.2	-0.1
-1.8	-1.4	-0.8	0.2	0.2	0.18	-0	0	0.7	-0	0.61	0.88	-0.4	-0.8	-0.8	0.35	1.2	0.8	1.5	-1.6	-1	-1	1.2	0.2	1.1	0.9	-0.2	-0.1

-0.1	0.7	0.6	0.2	0.2	0.18	-2	-1	0.7	1	0.61	-0.2	0.67	-0.1	1.42	-1.5	0.2	-0	-1.2	0	0.7	0.7	0.2	0.2	0.2	-1.1	-0.2	-0.1
0.7	0.7	1.3	0.99	0.96	0.92	0.8	1	-0.3	1	0.61	0.88	0.67	0.65	0.69	0.35	1.2	-0	-0.3	0.8	0.7	0.7	1.2	1.1	1.1	0.9	0.89	0.96
-1	-0.7	-0.8	0.2	0.2	0.18	-0	0	-0.3	-0	0.61	0.88	0.67	-0.8	-0.8	0.35	0.2	1.7	1.5	-0.8	-1	-1	0.2	0.2	0.2	-0.1	-0.2	-0.1
-1.8	0.7	0.6	-2.2	-0.6	0.18	-0	0	0.7	1	-0.4	-0.2	0.67	-0.1	-0.8	0.35	-1.8	-0	-0.3	-0.8	-1	-1	0.2	0.2	0.2	-0.1	0.89	0.96
0.7	1.4	0.6	0.99	0.96	0.92	0.8	1	-0.3	-2	0.61	0.88	0.67	1.38	1.42	-2.5	-2.7	-1.8	-2.2	1.6	1.5	1.5	-1	-1	-0.8	-3.1	-0.2	0.96
-0.1	0	-0.1	-0.6	-0.6	-0.6	-4	-2.9	-3.5	-4	-3.5	-2.5	-0.4	-0.1	-0	1.29	0.2	-0	-0.3	0	0	-0	1.2	-3	-2.7	-0.1	-0.2	-3.1
0.7	0.7	0.6	0.99	0.96	0.92	0.8	1	0.7	1	0.61	-0.2	0.67	0.65	0.69	0.35	0.2	-0.9	-0.3	-0.8	-1	-1	-1	-1	-0.8	-0.1	-0.2	-0.1
-1	-1.4	-1.6	0.2	-0.6	-0.6	0.8	0	0.7	1	0.61	-1.3	0.67	-0.8	-1.5	0.35	-0.8	-0.9	-1.2	1.6	1.5	1.5	-1	-1	-0.8	0.9	0.89	0.96

Step 4- Create Construct Scores for The Holdout Sample

Lat	OSS	Ipr	Adp	Cmp	Cst	Lgs	Inf	Ess	Mkt
-0.01	0.21	0.31	0.00	0.30	0.57	-0.47	-0.30	-0.14	-0.17
0.44	-1.49	0.36	-0.03	0.30	0.28	-0.80	0.82	0.19	0.27
-1.80	0.75	-0.34	-0.01	-1.61	0.28	0.85	-1.74	0.57	0.66
0.65	1.07	1.01	0.84	1.49	0.79	0.52	0.55	-0.56	-0.17
-0.22	1.07	-0.69	0.44	1.49	1.09	0.19	0.82	-0.89	0.66
0.57	-0.10	0.02	0.84	0.72	-1.90	-0.80	1.14	1.26	0.27
0.77	-0.92	-1.67	-2.54	-0.01	1.09	-0.14	0.82	-0.87	-0.56
-0.16	-1.46	-0.98	-0.85	0.77	-1.61	0.52	-1.46	0.23	-1.54
-0.41	1.07	-0.58	0.84	1.49	1.09	-0.47	-0.26	-0.10	-1.20
-1.80	0.78	0.02	0.02	-1.25	1.38	0.85	-1.46	0.18	0.22
1.63	1.07	1.01	0.84	1.49	-1.90	-2.13	1.68	1.26	1.15
-0.94	0.75	0.32	-0.03	-0.83	1.09	1.19	-1.42	0.53	0.32
0.12	-1.75	1.01	0.04	-0.42	0.28	0.19	0.28	-0.48	-1.14
-0.94	0.21	1.01	0.84	-0.83	0.28	1.85	-0.88	1.26	-0.17
0.45	-1.49	-0.06	-0.81	0.66	-0.81	-0.47	0.55	0.53	0.66
-0.09	-2.34	-1.69	-1.71	-0.06	-0.81	-0.14	-0.03	-0.87	-1.48
-1.47	0.53	0.00	-0.03	-1.61	0.28	0.85	-1.74	0.19	-0.17
0.65	1.07	0.72	-0.01	1.13	1.09	0.19	0.55	-1.23	0.32
-0.87	1.07	0.72	0.84	-0.83	0.28	-0.14	0.28	0.57	1.15

-1.80	-0.96	-2.24	-2.07	-1.61	-2.99	-2.13	-1.74	-3.01	-4.11
0.77	1.07	0.72	0.84	1.49	1.09	0.52	1.14	0.89	0.32
-0.01	0.75	1.01	0.84	-0.83	0.28	-1.46	-0.88	1.26	0.76
-1.15	1.07	-0.34	0.84	-1.61	-0.30	-2.13	-1.74	0.19	-0.56
-0.94	0.21	-0.34	-0.43	-0.83	0.28	0.85	-0.88	0.53	-0.17
1.63	-2.34	-0.05	0.84	1.49	-0.23	-0.14	1.68	-0.87	-1.04
1.63	-0.64	-0.98	0.44	-0.06	0.28	-0.14	0.82	0.19	-1.39
0.12	-1.49	-2.33	0.02	-0.83	0.57	1.19	-0.34	0.19	-1.48
0.45	1.07	1.01	0.84	0.72	-0.30	0.19	0.55	0.19	1.15
0.77	0.21	-0.28	-1.29	-0.06	0.28	-0.14	0.82	0.19	1.15
-1.60	0.50	0.38	-0.03	-1.25	1.09	1.52	-1.47	0.55	0.27
-1.80	1.07	1.01	0.84	-1.61	1.38	1.85	-1.74	1.26	1.15
0.57	0.75	0.02	0.04	1.13	-0.52	-1.13	1.10	0.91	1.15
-0.09	-1.18	1.01	0.84	0.72	-0.01	-0.14	-0.03	0.19	0.32
1.10	-0.01	-0.56	-0.01	0.30	0.28	0.52	0.82	-0.14	0.66
0.97	-0.07	0.00	0.04	1.13	-0.52	-0.80	0.82	-0.87	-0.71
-0.41	-0.07	0.72	0.44	-0.06	-0.01	-1.46	-0.88	0.91	1.15
-1.47	0.50	0.36	0.02	-1.61	0.57	1.52	-1.15	0.19	0.76
1.63	1.07	-0.99	0.84	-0.01	0.28	0.85	-0.03	0.19	-0.17
1.63	-0.01	1.01	0.84	1.49	0.28	-0.14	1.37	-0.16	0.70
-0.61	0.21	1.01	0.84	-0.83	-0.52	-0.14	-0.57	-1.21	0.70
0.24	-0.67	-3.03	-3.44	-1.25	0.57	-0.80	-0.61	-1.92	-2.37
1.63	-2.34	1.01	0.84	1.49	-1.32	-0.47	1.68	1.26	-0.94
0.77	0.21	-0.32	0.84	-0.06	-1.90	-0.14	-0.03	-0.14	-0.17
1.10	-2.34	0.02	0.84	1.49	-1.90	-1.13	-0.88	1.26	1.15
0.04	-0.10	-1.37	-1.76	0.35	-0.81	0.52	0.78	-0.89	-1.53
-0.01	1.07	1.01	0.84	-0.53	-1.90	-2.13	0.82	1.26	-1.00
0.97	-1.49	-0.27	-0.46	0.35	-0.81	-0.47	1.68	0.19	0.76
-0.62	-1.21	0.65	-0.87	-1.61	-0.81	-0.14	-0.30	-0.87	-0.17
0.57	-1.21	-1.32	-1.31	1.08	-0.23	-0.14	0.24	0.19	0.17
0.45	-1.49	-1.28	-0.43	0.30	-0.81	1.85	0.55	-0.87	1.15
0.44	0.50	0.07	0.44	0.72	-0.81	-0.14	-0.34	0.19	-0.95
-1.15	1.07	0.00	0.42	-1.61	1.09	1.85	-1.15	0.93	-1.48
-1.60	0.47	0.36	-0.43	-1.61	1.09	0.85	-1.42	0.89	0.32
0.37	0.53	1.01	0.84	-0.83	0.28	1.19	0.01	0.19	0.76
1.63	-0.64	-0.29	-1.31	-0.83	-0.81	0.85	-0.26	-0.16	-2.75
-0.94	1.07	1.01	-0.43	-0.83	0.28	1.85	-0.03	0.19	-0.17

-1.47	0.50	0.36	-0.01	-1.25	0.57	0.85	-1.42	0.53	0.27
0.24	0.75	1.01	0.44	1.08	1.38	-0.80	0.28	0.19	0.37
-0.09	0.81	-1.04	-0.84	0.35	-0.23	0.85	-0.03	-0.14	-1.10
0.04	-2.34	-0.34	-1.38	0.72	0.79	-2.13	0.82	0.19	-0.66
0.97	-2.34	1.01	0.84	0.66	0.21	0.85	1.68	1.26	1.15
-1.47	0.75	0.67	0.44	-1.25	1.09	1.52	-1.15	0.55	1.15
0.77	-0.10	0.36	0.44	0.72	-2.70	-0.14	0.82	0.91	0.37
-1.47	0.50	-0.34	-0.43	1.49	-0.81	1.85	-1.42	0.19	1.15
0.77	1.07	1.01	-0.01	-0.01	0.28	-1.13	0.82	-0.50	1.15
0.33	0.21	-0.92	-0.37	0.35	1.38	-0.14	0.82	-0.50	-0.50
-1.27	0.21	-1.05	-0.86	-1.20	0.28	0.19	-1.74	-0.16	0.32
-0.01	0.47	1.01	0.84	-0.11	0.28	0.85	0.82	-0.50	0.76
0.77	-0.32	-1.74	-1.31	0.72	0.28	-0.14	0.20	-1.94	-1.48
-0.28	1.07	1.01	0.84	1.49	-0.30	-1.80	-0.61	-0.54	0.76
0.37	-0.61	0.72	0.84	0.72	-2.19	0.52	0.82	0.55	-0.17
-0.28	-0.07	1.01	0.02	0.72	1.09	0.19	-0.03	0.19	1.15
0.44	1.07	-1.62	-2.58	-0.06	0.79	0.19	0.24	-3.01	-1.00
-0.94	1.07	1.01	0.84	-0.83	0.28	0.85	-0.88	1.26	1.15
0.77	0.21	-0.34	-0.43	0.72	-0.52	-0.47	-0.34	-0.87	0.22
0.24	1.07	-0.34	0.44	-0.01	0.28	-0.14	-0.03	-1.27	0.70
-0.48	-0.07	0.42	0.04	-0.78	0.28	-1.13	0.59	0.55	1.15
-1.80	0.50	0.36	-0.43	-1.61	1.09	0.85	-1.46	0.57	0.22
0.57	1.07	1.01	0.44	0.72	-0.01	-0.80	0.51	-0.87	0.22
0.77	-0.36	1.01	0.84	0.72	-0.30	0.19	0.82	-3.01	1.15
1.30	-0.83	-0.94	0.02	1.08	1.38	1.19	1.41	1.26	-0.07
1.63	1.07	0.67	0.84	1.49	1.38	-2.13	1.68	1.26	1.15
0.25	-2.34	-1.62	-0.45	-0.42	-0.23	-0.80	0.28	-1.58	-1.43
-0.09	-2.34	-0.34	0.42	0.72	0.28	0.19	0.51	-0.93	-1.14
1.30	1.07	1.01	0.84	0.77	-0.23	-1.80	1.68	1.26	0.26
-0.62	1.07	0.38	-0.86	-1.25	1.38	0.19	-1.42	0.89	-0.17
-0.94	0.21	1.01	0.84	-0.83	0.28	0.85	-0.88	0.19	-0.17
-0.21	1.07	1.01	0.84	-0.37	-0.59	-1.13	0.82	0.19	-0.46
-1.47	0.50	0.02	0.44	-1.61	0.28	0.85	-1.42	0.89	0.70
0.77	-0.64	1.01	0.44	0.72	-1.10	-0.14	0.82	0.19	0.76
-0.29	-0.92	-2.03	-2.99	-0.83	0.28	-0.47	-0.88	-0.83	-0.17
-1.60	0.21	0.00	0.42	-0.83	1.09	1.19	-1.42	0.93	0.27
0.44	0.21	-0.28	0.44	0.66	-0.30	-0.47	0.55	0.19	-0.61

0.97	1.07	0.67	0.84	0.72	1.09	-0.14	0.82	1.26	1.15
-0.94	0.21	-0.34	0.84	-0.83	0.28	1.85	-0.88	0.19	-0.17
-0.22	-0.99	0.36	-0.01	-0.42	-1.32	-0.14	-0.88	0.19	0.70
1.10	1.07	-0.05	0.84	1.49	-2.99	-2.13	1.68	-0.87	-1.11
-0.09	-0.64	-4.38	-2.59	-0.06	0.57	-0.14	-0.03	-1.58	-1.33
0.77	1.07	1.01	0.44	0.72	0.28	-0.80	-0.88	-0.87	-0.17
-1.47	-0.32	0.72	0.04	-1.20	-0.52	-1.13	1.68	-0.87	1.15

Step 5 - Standardize the Construct Scores for The Holdout Sample

LatStd	OSSStd	IprStd	AdpStd	CmpStd	CstStd	LgsStd	InfStd	EssStd	MktStd
-0.010	0.203	0.299	-0.003	0.301	0.576	-0.452	-0.295	-0.148	-0.170
0.453	-1.421	0.352	-0.032	0.301	0.284	-0.771	0.808	0.204	0.272
-1.846	0.714	-0.327	-0.012	-1.606	0.284	0.822	-1.701	0.597	0.661
0.668	1.014	0.978	0.874	1.487	0.797	0.503	0.542	-0.582	-0.170
-0.224	1.014	-0.670	0.455	1.487	1.088	0.184	0.808	-0.932	0.661
0.586	-0.097	0.022	0.874	0.714	-1.905	-0.771	1.113	1.320	0.272
0.792	-0.879	-1.619	-2.632	-0.007	1.088	-0.134	0.808	-0.912	-0.558
-0.169	-1.392	-0.947	-0.880	0.767	-1.614	0.503	-1.435	0.245	-1.542
-0.422	1.014	-0.566	0.874	1.487	1.088	-0.453	-0.255	-0.107	-1.199
-1.846	0.743	0.015	0.017	-1.245	1.379	0.822	-1.435	0.184	0.218
1.671	1.014	0.978	0.874	1.487	-1.905	-2.044	1.644	1.320	1.148
-0.966	0.714	0.306	-0.032	-0.832	1.088	1.140	-1.396	0.556	0.317
0.123	-1.661	0.978	0.037	-0.420	0.284	0.184	0.277	-0.500	-1.145
-0.966	0.203	0.978	0.874	-0.832	0.284	1.777	-0.865	1.320	-0.170
0.457	-1.421	-0.063	-0.840	0.661	-0.810	-0.452	0.542	0.556	0.661
-0.087	-2.232	-1.633	-1.775	-0.059	-0.810	-0.134	-0.029	-0.912	-1.488
-1.507	0.503	0.002	-0.032	-1.606	0.284	0.822	-1.701	0.204	-0.170
0.664	1.014	0.694	-0.012	1.127	1.088	0.184	0.542	-1.284	0.317
-0.889	1.014	0.694	0.874	-0.832	0.284	-0.135	0.276	0.597	1.148
-1.846	-0.910	-2.174	-2.145	-1.606	-2.999	-2.044	-1.701	-3.145	-4.124
0.792	1.014	0.694	0.874	1.487	1.088	0.503	1.113	0.927	0.317
-0.010	0.714	0.978	0.874	-0.832	0.284	-1.408	-0.865	1.320	0.760
-1.173	1.014	-0.327	0.874	-1.606	-0.298	-2.044	-1.701	0.204	-0.558
-0.966	0.203	-0.327	-0.451	-0.832	0.284	0.822	-0.865	0.556	-0.170

1.671	-2.232	-0.050	0.874	1.487	-0.228	-0.134	1.644	-0.912	-1.046
1.671	-0.609	-0.954	0.455	-0.059	0.284	-0.134	0.808	0.204	-1.389
0.123	-1.421	-2.253	0.017	-0.832	0.576	1.140	-0.333	0.204	-1.479
0.457	1.014	0.978	0.874	0.714	-0.298	0.184	0.542	0.204	1.148
0.792	0.203	-0.275	-1.337	-0.059	0.284	-0.134	0.808	0.204	1.148
-1.640	0.473	0.365	-0.032	-1.245	1.088	1.459	-1.436	0.575	0.272
-1.846	1.014	0.978	0.874	-1.606	1.379	1.777	-1.701	1.320	1.148
0.586	0.714	0.022	0.037	1.127	-0.519	-1.089	1.073	0.948	1.148
-0.087	-1.121	0.978	0.874	0.714	-0.007	-0.134	-0.029	0.204	0.317
1.131	-0.009	-0.546	-0.012	0.301	0.284	0.503	0.808	-0.148	0.661
0.998	-0.068	-0.004	0.037	1.127	-0.519	-0.771	0.807	-0.912	-0.711
-0.422	-0.068	0.694	0.455	-0.059	-0.007	-1.408	-0.865	0.948	1.148
-1.507	0.473	0.345	0.017	-1.606	0.576	1.459	-1.130	0.204	0.760
1.671	1.014	-0.960	0.874	-0.007	0.284	0.822	-0.029	0.204	-0.170
1.671	-0.009	0.978	0.874	1.487	0.284	-0.134	1.339	-0.169	0.706
-0.628	0.203	0.978	0.874	-0.832	-0.519	-0.134	-0.560	-1.264	0.706
0.247	-0.638	-2.938	-3.567	-1.245	0.576	-0.771	-0.599	-2.008	-2.372
1.671	-2.232	0.978	0.874	1.487	-1.323	-0.453	1.644	1.320	-0.947
0.792	0.203	-0.314	0.874	-0.059	-1.905	-0.134	-0.029	-0.148	-0.170
1.131	-2.232	0.022	0.874	1.487	-1.905	-1.089	-0.865	1.320	1.148
0.041	-0.097	-1.329	-1.824	0.354	-0.810	0.503	0.768	-0.932	-1.533
-0.006	1.014	0.978	0.874	-0.525	-1.905	-2.044	0.808	1.320	-1.000
0.998	-1.421	-0.262	-0.479	0.354	-0.810	-0.452	1.644	0.204	0.760
-0.632	-1.150	0.629	-0.898	-1.606	-0.810	-0.134	-0.295	-0.912	-0.170
0.586	-1.150	-1.283	-1.356	1.074	-0.228	-0.134	0.237	0.204	0.173
0.457	-1.421	-1.244	-0.451	0.301	-0.810	1.777	0.542	-0.912	1.148
0.453	0.473	0.068	0.455	0.714	-0.810	-0.134	-0.334	0.204	-0.955
-1.173	1.014	0.002	0.435	-1.606	1.088	1.777	-1.130	0.968	-1.488
-1.640	0.443	0.352	-0.451	-1.606	1.088	0.822	-1.396	0.927	0.317
0.380	0.503	0.978	0.874	-0.832	0.284	1.140	0.011	0.204	0.760
1.671	-0.609	-0.282	-1.356	-0.832	-0.810	0.822	-0.255	-0.168	-2.761
-0.966	1.014	0.978	-0.451	-0.832	0.284	1.777	-0.029	0.204	-0.170
-1.507	0.473	0.345	-0.012	-1.245	0.576	0.822	-1.396	0.556	0.272
0.247	0.714	0.978	0.455	1.074	1.379	-0.771	0.276	0.204	0.371
-0.087	0.773	-1.006	-0.869	0.354	-0.228	0.822	-0.029	-0.148	-1.100
0.041	-2.232	-0.327	-1.434	0.714	0.797	-2.044	0.808	0.204	-0.666
0.994	-2.232	0.978	0.874	0.661	0.215	0.822	1.644	1.320	1.148

-1.507	0.714	0.649	0.455	-1.245	1.088	1.459	-1.130	0.575	1.148
0.792	-0.097	0.352	0.455	0.714	-2.708	-0.134	0.808	0.948	0.371
-1.507	0.473	-0.327	-0.451	1.487	-0.810	1.777	-1.396	0.204	1.148
0.788	1.014	0.978	-0.012	-0.007	0.284	-1.089	0.808	-0.520	1.148
0.333	0.203	-0.895	-0.382	0.354	1.379	-0.134	0.808	-0.520	-0.504
-1.301	0.203	-1.019	-0.889	-1.193	0.284	0.184	-1.701	-0.168	0.317
-0.006	0.443	0.978	0.874	-0.112	0.284	0.822	0.808	-0.520	0.760
0.792	-0.309	-1.685	-1.356	0.714	0.284	-0.134	0.197	-2.029	-1.488
-0.289	1.014	0.978	0.874	1.487	-0.298	-1.726	-0.599	-0.561	0.760
0.380	-0.580	0.694	0.874	0.714	-2.196	0.503	0.808	0.577	-0.170
-0.289	-0.068	0.978	0.017	0.714	1.088	0.184	-0.029	0.204	1.148
0.453	1.014	-1.574	-2.681	-0.059	0.797	0.184	0.237	-3.145	-1.000
-0.966	1.014	0.978	0.874	-0.832	0.284	0.822	-0.865	1.320	1.148
0.792	0.203	-0.327	-0.451	0.714	-0.519	-0.452	-0.334	-0.912	0.218
0.247	1.014	-0.327	0.455	-0.007	0.284	-0.134	-0.029	-1.325	0.706
-0.491	-0.068	0.410	0.037	-0.780	0.284	-1.090	0.583	0.577	1.148
-1.846	0.473	0.352	-0.451	-1.606	1.088	0.822	-1.435	0.597	0.218
0.586	1.014	0.978	0.455	0.714	-0.007	-0.771	0.502	-0.912	0.218
0.788	-0.339	0.978	0.874	0.714	-0.298	0.184	0.808	-3.145	1.148
1.333	-0.791	-0.908	0.017	1.074	1.379	1.140	1.378	1.320	-0.071
1.671	1.014	0.649	0.874	1.487	1.379	-2.044	1.644	1.320	1.148
0.251	-2.232	-1.567	-0.470	-0.420	-0.228	-0.771	0.276	-1.659	-1.434
-0.087	-2.232	-0.327	0.435	0.714	0.284	0.184	0.502	-0.974	-1.145
1.333	1.014	0.978	0.874	0.767	-0.228	-1.726	1.644	1.320	0.264
-0.632	1.014	0.365	-0.889	-1.245	1.379	0.184	-1.396	0.927	-0.170
-0.966	0.203	0.978	0.874	-0.832	0.284	0.822	-0.865	0.204	-0.170
-0.212	1.014	0.978	0.874	-0.367	-0.589	-1.089	0.808	0.204	-0.459
-1.507	0.473	0.022	0.455	-1.606	0.284	0.822	-1.396	0.927	0.706
0.792	-0.609	0.978	0.455	0.714	-1.101	-0.134	0.808	0.204	0.760
-0.297	-0.880	-1.969	-3.099	-0.832	0.284	-0.453	-0.865	-0.872	-0.170
-1.640	0.203	0.002	0.435	-0.832	1.088	1.140	-1.396	0.968	0.272
0.453	0.203	-0.275	0.455	0.661	-0.298	-0.452	0.542	0.204	-0.612
0.998	1.014	0.649	0.874	0.714	1.088	-0.134	0.808	1.320	1.148
-0.966	0.203	-0.327	0.874	-0.832	0.284	1.777	-0.865	0.204	-0.170
-0.224	-0.938	0.352	-0.012	-0.420	-1.323	-0.134	-0.865	0.204	0.706
1.127	1.014	-0.050	0.874	1.487	-2.999	-2.044	1.644	-0.912	-1.108
-0.087	-0.609	-4.243	-2.690	-0.059	0.576	-0.134	-0.029	-1.659	-1.335

0.792	1.014	0.978	0.455	0.714	0.284	-0.771	-0.865	-0.912	-0.170
-1.507	-0.309	0.694	0.037	-1.193	-0.519	-1.089	1.644	-0.912	1.148

Step 6 - Create Prediction Scores

LatStd	OSSStd	IprStd	AdpStd	CmpStd	CstStd	LgsStd	InfStd	EssStd	MktStd	AdpPred ALL
-0.010	0.203	0.299	-0.003	0.301	0.576	-0.452	-0.295	-0.148	-0.170	0.089
0.453	-1.421	0.352	-0.032	0.301	0.284	-0.771	0.808	0.204	0.272	0.455
-1.846	0.714	-0.327	-0.012	-1.606	0.284	0.822	-1.701	0.597	0.661	-0.591
0.668	1.014	0.978	0.874	1.487	0.797	0.503	0.542	-0.582	-0.170	0.728
-0.224	1.014	-0.670	0.455	1.487	1.088	0.184	0.808	-0.932	0.661	0.170
0.586	-0.097	0.022	0.874	0.714	-1.905	-0.771	1.113	1.320	0.272	0.558
0.792	-0.879	-1.619	-2.632	-0.007	1.088	-0.134	0.808	-0.912	-0.558	-0.623
-0.169	-1.392	-0.947	-0.880	0.767	-1.614	0.503	-1.435	0.245	-1.542	-1.488
-0.422	1.014	-0.566	0.874	1.487	1.088	-0.453	-0.255	-0.107	-1.199	-0.239
-1.846	0.743	0.015	0.017	-1.245	1.379	0.822	-1.435	0.184	0.218	-0.336
1.671	1.014	0.978	0.874	1.487	-1.905	-2.044	1.644	1.320	1.148	1.529
-0.966	0.714	0.306	-0.032	-0.832	1.088	1.140	-1.396	0.556	0.317	-0.152
0.123	-1.661	0.978	0.037	-0.420	0.284	0.184	0.277	-0.500	-1.145	0.137
-0.966	0.203	0.978	0.874	-0.832	0.284	1.777	-0.865	1.320	-0.170	0.252
0.457	-1.421	-0.063	-0.840	0.661	-0.810	-0.452	0.542	0.556	0.661	0.144
-0.087	-2.232	-1.633	-1.775	-0.059	-0.810	-0.134	-0.029	-0.912	-1.488	-1.438
-1.507	0.503	0.002	-0.032	-1.606	0.284	0.822	-1.701	0.204	-0.170	-0.636
0.664	1.014	0.694	-0.012	1.127	1.088	0.184	0.542	-1.284	0.317	0.578
-0.889	1.014	0.694	0.874	-0.832	0.284	-0.135	0.276	0.597	1.148	0.809
-1.846	-0.910	-2.174	-2.145	-1.606	-2.999	-2.044	-1.701	-3.145	-4.124	-3.046
0.792	1.014	0.694	0.874	1.487	1.088	0.503	1.113	0.927	0.317	1.184
-0.010	0.714	0.978	0.874	-0.832	0.284	-1.408	-0.865	1.320	0.760	0.620
-1.173	1.014	-0.327	0.874	-1.606	-0.298	-2.044	-1.701	0.204	-0.558	-0.660
-0.966	0.203	-0.327	-0.451	-0.832	0.284	0.822	-0.865	0.556	-0.170	-0.418
1.671	-2.232	-0.050	0.874	1.487	-0.228	-0.134	1.644	-0.912	-1.046	0.076
1.671	-0.609	-0.954	0.455	-0.059	0.284	-0.134	0.808	0.204	-1.389	-0.271
0.123	-1.421	-2.253	0.017	-0.832	0.576	1.140	-0.333	0.204	-1.479	-1.475
0.457	1.014	0.978	0.874	0.714	-0.298	0.184	0.542	0.204	1.148	0.920
0.792	0.203	-0.275	-1.337	-0.059	0.284	-0.134	0.808	0.204	1.148	0.433
-1.640	0.473	0.365	-0.032	-1.245	1.088	1.459	-1.436	0.575	0.272	-0.201

-1.846	1.014	0.978	0.874	-1.606	1.379	1.777	-1.701	1.320	1.148	0.282
0.586	0.714	0.022	0.037	1.127	-0.519	-1.089	1.073	0.948	1.148	0.851
-0.087	-1.121	0.978	0.874	0.714	-0.007	-0.134	-0.029	0.204	0.317	0.386
1.131	-0.009	-0.546	-0.012	0.301	0.284	0.503	0.808	-0.148	0.661	0.133
0.998	-0.068	-0.004	0.037	1.127	-0.519	-0.771	0.807	-0.912	-0.711	0.067
-0.422	-0.068	0.694	0.455	-0.059	-0.007	-1.408	-0.865	0.948	1.148	0.355
-1.507	0.473	0.345	0.017	-1.606	0.576	1.459	-1.130	0.204	0.760	-0.156
1.671	1.014	-0.960	0.874	-0.007	0.284	0.822	-0.029	0.204	-0.170	-0.313
1.671	-0.009	0.978	0.874	1.487	0.284	-0.134	1.339	-0.169	0.706	1.086
-0.628	0.203	0.978	0.874	-0.832	-0.519	-0.134	-0.560	-1.264	0.706	0.044
0.247	-0.638	-2.938	-3.567	-1.245	0.576	-0.771	-0.599	-2.008	-2.372	-2.190
1.671	-2.232	0.978	0.874	1.487	-1.323	-0.453	1.644	1.320	-0.947	0.854
0.792	0.203	-0.314	0.874	-0.059	-1.905	-0.134	-0.029	-0.148	-0.170	-0.365
1.131	-2.232	0.022	0.874	1.487	-1.905	-1.089	-0.865	1.320	1.148	-0.330
0.041	-0.097	-1.329	-1.824	0.354	-0.810	0.503	0.768	-0.932	-1.533	-0.775
-0.006	1.014	0.978	0.874	-0.525	-1.905	-2.044	0.808	1.320	-1.000	0.905
0.998	-1.421	-0.262	-0.479	0.354	-0.810	-0.452	1.644	0.204	0.760	0.433
-0.632	-1.150	0.629	-0.898	-1.606	-0.810	-0.134	-0.295	-0.912	-0.170	-0.256
0.586	-1.150	-1.283	-1.356	1.074	-0.228	-0.134	0.237	0.204	0.173	-0.565
0.457	-1.421	-1.244	-0.451	0.301	-0.810	1.777	0.542	-0.912	1.148	-0.728
0.453	0.473	0.068	0.455	0.714	-0.810	-0.134	-0.334	0.204	-0.955	-0.188
-1.173	1.014	0.002	0.435	-1.606	1.088	1.777	-1.130	0.968	-1.488	-0.343
-1.640	0.443	0.352	-0.451	-1.606	1.088	0.822	-1.396	0.927	0.317	-0.094
0.380	0.503	0.978	0.874	-0.832	0.284	1.140	0.011	0.204	0.760	0.589
1.671	-0.609	-0.282	-1.356	-0.832	-0.810	0.822	-0.255	-0.168	-2.761	-0.797
-0.966	1.014	0.978	-0.451	-0.832	0.284	1.777	-0.029	0.204	-0.170	0.475
-1.507	0.473	0.345	-0.012	-1.245	0.576	0.822	-1.396	0.556	0.272	-0.214
0.247	0.714	0.978	0.455	1.074	1.379	-0.771	0.276	0.204	0.371	0.923
-0.087	0.773	-1.006	-0.869	0.354	-0.228	0.822	-0.029	-0.148	-1.100	-0.600
0.041	-2.232	-0.327	-1.434	0.714	0.797	-2.044	0.808	0.204	-0.666	0.070
0.994	-2.232	0.978	0.874	0.661	0.215	0.822	1.644	1.320	1.148	1.179
-1.507	0.714	0.649	0.455	-1.245	1.088	1.459	-1.130	0.575	1.148	0.182
0.792	-0.097	0.352	0.455	0.714	-2.708	-0.134	0.808	0.948	0.371	0.412
-1.507	0.473	-0.327	-0.451	1.487	-0.810	1.777	-1.396	0.204	1.148	-0.642
0.788	1.014	0.978	-0.012	-0.007	0.284	-1.089	0.808	-0.520	1.148	1.024
0.333	0.203	-0.895	-0.382	0.354	1.379	-0.134	0.808	-0.520	-0.504	-0.058
-1.301	0.203	-1.019	-0.889	-1.193	0.284	0.184	-1.701	-0.168	0.317	-1.096

-0.006	0.443	0.978	0.874	-0.112	0.284	0.822	0.808	-0.520	0.760	0.788
0.792	-0.309	-1.685	-1.356	0.714	0.284	-0.134	0.197	-2.029	-1.488	-1.210
-0.289	1.014	0.978	0.874	1.487	-0.298	-1.726	-0.599	-0.561	0.760	0.408
0.380	-0.580	0.694	0.874	0.714	-2.196	0.503	0.808	0.577	-0.170	0.391
-0.289	-0.068	0.978	0.017	0.714	1.088	0.184	-0.029	0.204	1.148	0.709
0.453	1.014	-1.574	-2.681	-0.059	0.797	0.184	0.237	-3.145	-1.000	-1.104
-0.966	1.014	0.978	0.874	-0.832	0.284	0.822	-0.865	1.320	1.148	0.565
0.792	0.203	-0.327	-0.451	0.714	-0.519	-0.452	-0.334	-0.912	0.218	-0.404
0.247	1.014	-0.327	0.455	-0.007	0.284	-0.134	-0.029	-1.325	0.706	-0.149
-0.491	-0.068	0.410	0.037	-0.780	0.284	-1.090	0.583	0.577	1.148	0.727
-1.846	0.473	0.352	-0.451	-1.606	1.088	0.822	-1.435	0.597	0.218	-0.178
0.586	1.014	0.978	0.455	0.714	-0.007	-0.771	0.502	-0.912	0.218	0.680
0.788	-0.339	0.978	0.874	0.714	-0.298	0.184	0.808	-3.145	1.148	0.264
1.333	-0.791	-0.908	0.017	1.074	1.379	1.140	1.378	1.320	-0.071	0.363
1.671	1.014	0.649	0.874	1.487	1.379	-2.044	1.644	1.320	1.148	1.728
0.251	-2.232	-1.567	-0.470	-0.420	-0.228	-0.771	0.276	-1.659	-1.434	-1.320
-0.087	-2.232	-0.327	0.435	0.714	0.284	0.184	0.502	-0.974	-1.145	-0.504
1.333	1.014	0.978	0.874	0.767	-0.228	-1.726	1.644	1.320	0.264	1.569
-0.632	1.014	0.365	-0.889	-1.245	1.379	0.184	-1.396	0.927	-0.170	0.002
-0.966	0.203	0.978	0.874	-0.832	0.284	0.822	-0.865	0.204	-0.170	0.108
-0.212	1.014	0.978	0.874	-0.367	-0.589	-1.089	0.808	0.204	-0.459	0.854
-1.507	0.473	0.022	0.455	-1.606	0.284	0.822	-1.396	0.927	0.706	-0.276
0.792	-0.609	0.978	0.455	0.714	-1.101	-0.134	0.808	0.204	0.760	0.717
-0.297	-0.880	-1.969	-3.099	-0.832	0.284	-0.453	-0.865	-0.872	-0.170	-1.463
-1.640	0.203	0.002	0.435	-0.832	1.088	1.140	-1.396	0.968	0.272	-0.287
0.453	0.203	-0.275	0.455	0.661	-0.298	-0.452	0.542	0.204	-0.612	0.081
0.998	1.014	0.649	0.874	0.714	1.088	-0.134	0.808	1.320	1.148	1.244
-0.966	0.203	-0.327	0.874	-0.832	0.284	1.777	-0.865	0.204	-0.170	-0.537
-0.224	-0.938	0.352	-0.012	-0.420	-1.323	-0.134	-0.865	0.204	0.706	-0.309
1.127	1.014	-0.050	0.874	1.487	-2.999	-2.044	1.644	-0.912	-1.108	0.273
-0.087	-0.609	-4.243	-2.690	-0.059	0.576	-0.134	-0.029	-1.659	-1.335	-2.389
0.792	1.014	0.978	0.455	0.714	0.284	-0.771	-0.865	-0.912	-0.170	0.132
-1.507	-0.309	0.694	0.037	-1.193	-0.519	-1.089	1.644	-0.912	1.148	0.876

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *J. Assoc. Inf. Syst. Journal of the Association of Information Systems, 17*(2), 1-32.

Abrahamson, E. (1991). Managerial Fads and Fashions: The Diffusion and Rejection of Innovations. *AMR Academy of Management Review, 16*(3), 586-612.

Abutu, O. P. (2014). Consequences of the January 2012 oil subsidy removal in Nigeria. *Journal of Business and Retail Management Research, 8*(2).

Afthanorhan, W. (2013). A comparison of partial least square structural equation modeling (PLS-SEM) and covariance based structural equation modeling (CB-SEM) for confirmatory factor analysis. *International Journal of Engineering Science and Innovative Technology, 2*(5), 198-205.

Agrawal, K. (2015). Investigating the determinants of Big Data Analytics (BDA) adoption in Asian emerging economies.

Agrawal, K. P. (2013). The assimilation of Big Data Analytics (BDA) by Indian firms: a technology diffusion perspective.

Ainur, A., Sayang, M., Jannoo, Z., & Yap, B. (2017). Sample Size and Non-Normality Effects on Goodness of Fit Measures in Structural Equation Models. *Pertanika Journal of Science & Technology, 25*(2).

Ajila, S. A., & Wu, D. (2007). Empirical study of the effects of open source adoption on software development economics. *JOURNAL OF SYSTEMS AND SOFTWARE, 80*(9), 1517-1529.

Ajimoko, O. J. (2017). *Exploring the Cloud-Based Big Data Analytics Adoption Criteria for Small Business Enterprises.* Colorado Technical University,

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes, 50*(2), 179-211.

Akter, S., D'ambra, J., & Ray, P. (2011). An evaluation of PLS based complex models: the roles of power analysis, predictive relevance and GoF index.

Akter, S., Fosso Wamba, S., & Dewan, S. (2017). Why PLS-SEM is suitable for complex modelling? An empirical illustration in big data analytics quality. *Production Planning & Control, 28*(11-12), 1011-1021.

Alase, A. (2017). The interpretative phenomenological analysis (IPA): A guide to a good qualitative research approach. *International Journal of Education and Literacy Studies, 5*(2), 9-19.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology, 112*(4), 545.

Alnafoosi, A. B., & Steinbach, T. (2013). An integrated framework for evaluating big-data storage solutions - IDA case study. 947-956.

Alrousan, M. K. (2015). E-commerce adoption by travel agencies in Jordan. Retrieved from http://hdl.handle.net/10369/7539

Alvesson, M. (2003). Beyond neopositivists, romantics, and localists: A reflexive approach to interviews in organizational research. *Academy of management review*, *28*(1), 13-33.

Amabile, T. M., Patterson, C., Mueller, J., Wojcik, T., Odomirok, P. W., Marsh, M., & Kramer, S. J. (2001). Academic-Practitioner Collaboration in Management Research: A Case of Cross-Profession Collaboration. *The Academy of Management Journal*, *44*(2), 418-431.

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2018). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*.

Apache. (2019). Apache Hadoop. Retrieved from http://hadoop.apache.org

Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ*, *1986*, 23-28.

Barham, H. (2017). *Achieving competitive advantage through Big Data: a literature review*. Paper presented at the 2017 Portland international conference on management of engineering and technology (PICMET).

Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *INFORMATION TECHNOLOGY LEARNING AND PERFORMANCE JOURNAL*, *19*(1), 43-50.

Becker, J.-M., Ringle, C. M., & Sarstedt, M. (2018). Estimating moderating effects in PLS-SEM and PLSc-SEM: Interaction term generation* data treatment. *Journal of Applied Structural Equation Modeling*, *2*(2), 1-21.

Bellinger, G., Castro, D., & Mills, A. (2004). Data, information, knowledge, and wisdom.

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

Benitez, J. H. J. r. C. A. S. F. (2020). How to perform and report an impactful analysis using partial least squares : guidelines for confirmatory and explanatory IS research. *Information & Management*, *57*, 1-16.

Bhattacherjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *misquarterly MIS Quarterly*, *25*(3), 351-370.

Boonsiritomachai, W. (2014). *Enablers affecting the adoption of business intelligence: a study of Thai small and medium-sized enterprises.* Victoria University,

Bremser, C. (2018). *Starting Points for Big Data Adoption.* Paper presented at the ECIS.

Brocki, J. M., & Wearden, A. J. (2006). A critical evaluation of the use of interpretative phenomenological analysis (IPA) in health psychology. *Psychology and health, 21*(1), 87-108.

Brown, C., Smith, P., Arduengo, N., & Taylor, M. (2016). Trusting telework in the federal government. *Qualitative Report, 21*(1).

Brynjolfsson, E., Hu, Y., & Simester, D. (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science, 57*(8), 1373-1386.

Bsharah, F., & Less, M. (2000). Requirements and strategies for the retention of automotive product data. *JCAD</cja:jid> Computer-Aided Design, 32*(2), 145-158.

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.

Cepeda Carrion, G. H. J. r. R. C. M. R. n. J. L. (2016). Prediction-oriented modeling in business research by means of PLS path modeling: Introduction to a JBR special section. *JBR Journal of Business Research, 69*(10), 4545-4551.

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod record, 26*(1), 65-74.

Chen, H.-M., Kazman, R., & Matthes, F. (2015). *Demystifying big data adoption: Beyond IT fashion and relative advantage.* Paper presented at the Proceedings of Pre-ICIS (International Conference on Information System) DIGIT workshop.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *misquarterly MIS Quarterly, 36*(4), 1165-1188.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Netw Appl Mobile Networks and Applications : The Journal of SPECIAL ISSUES on Mobility of Systems, Users, Data and Computing, 19*(2), 171-209.

Chen, P. M. (1993). *RAID : high-performance, reliable secondary storage*. Berkeley, Calif.: University of California, Berkeley, Computer Science Division.

Chin, W. W. (1998a). Commentary: Issues and opinion on structural equation modeling. In: JSTOR.

Chin, W. W. (1998b). The partial least squares approach to structural equation modeling. *Modern methods for business research, 295*(2), 295-336.

Cho, J. Y., & Lee, E.-H. (2014). Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *The qualitative report, 19*(32), 1-20.

Columbus, L. (2017). 53% Of Companies Are Adopting Big Data Analytics. Retrieved from https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/#6a1f723039a1

Comrey, A., & Lee, H. (1992). Interpretation and application of factor analytic results. *Comrey AL, Lee HB. A first course in factor analysis, 2*, 1992.

Connelly, L. M. (2008). Pilot studies. *Medsurg Nursing, 17*(6), 411.

Conway, P. (1996). *Preservation in the digital world*. [S.l.].

Coughlin, T. (2016). The Costs Of Storage. Retrieved from https://www.forbes.com/sites/tomcoughlin/2016/07/24/the-costs-of-storage/#2bfda8d53239

Cox, M., & Ellsworth, D. (1997). *Managing big data for scientific visualization.* Paper presented at the ACM Siggraph.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice, 39*(3), 124-130.

Datoo, S. (2014, 14 Jan 2014). Big data: 4 predictions for 2014. Retrieved from https://www.theguardian.com/technology/datablog/2014/jan/14/big-data-4-predictions-for-2014

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *misquarterly MIS Quarterly, 13*(3), 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace 1. *Journal of applied social psychology, 22*(14), 1111-1132.

de Camargo Fiorini, P., Seles, B. M. R. P., Jabbour, C. J. C., Mariano, E. B., & de Sousa Jabbour, A. B. L. (2018). Management theory and big data literature: From a review to a research agenda. *International Journal of Information Management, 43*, 112-129.

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. Retrieved from https://nls.ldls.org.uk/welcome.html?ark:/81055/vdc_100032418576.0x000064

DePietro, R., Wiarda, E., & Fleischer, M. (1990). The context for Change: Organization, Technology and Environment. In L. G. T. M. F. A. K. Chakrabarti (Ed.), *The processes of technological innovation* (pp. 151-175). Lexington, Mass: Lexington Books.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical education, 40*(4), 314-321.

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS quarterly, 39*(2).

Dijkstra, T. K., & Schermelleh-Engel, K. (2014). Consistent Partial Least Squares for Nonlinear Structural Equation Models. *Psychometrika Psychometrika, 79*(4), 585-604.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*: John Wiley & Sons.

Dremel, C., Herterich, M. M., Wulf, J., Brenner, W., Herterich, M. M., & Waizmann, J. C. (2017). How AUDI AG established big data analytics in its digital transformation. *MIS Q. Exec. MIS Quarterly Executive, 16*(2), 81-100.

Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2016). The impact of big data on world-class sustainable manufacturing. *The International Journal of Advanced Manufacturing Technology, 84*(1-4), 631-645.

Durdyev, S., Ihtiyar, A., Banaitis, A., & Thurnell, D. (2018). The construction client satisfaction model: a PLS-SEM approach. *Journal of Civil Engineering and Management, 24*(1), 31-42.

Eggers, J., & Hein, A. (2020). *Turning Big Data into Value: A Literature Review on Business Value Realization from Process Mining.* Paper presented at the ECIS.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling, 8*(3), 430-457.

Esteves, J., & Curto, J. (2013). A risk and benefits behavioral model to assess intentions to adopt big data. *J. Intell. Stud. Bus. Journal of Intelligence Studies in Business, 3*(3), 37-46.

Fan, M., Stallaert, J., & Whinston, A. B. (2000). The adoption and design methodologies of component-based enterprise systems. *EUROPEAN JOURNAL OF INFORMATION SYSTEMS, 9*, 25-35.

Felin, T., & Zenger, T. R. (2014). Closed or open innovation? Problem solving and the governance choice. *Research policy, 43*(5), 914-925.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior : an introduction to theory and research*. Reading, Mass.; London: Addison-Wesley.

Fontana Jr, R. E., Decad, G., & Hetzler, S. R. (2012). Technology roadmap comparisons for TAPE, HDD, and NAND flash: Implications for data storage applications. *IEEE Trans Magn IEEE Transactions on Magnetics, 48*(5 PART 1), 1692-1696.

Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *Int J Inf Manage International Journal of Information Management, 36*(3), 403-413.

Fulk, J. (1987). A Social Information Processing Model of Media Use in Organizations. *Communication Research: An International Quarterly, 14*(5), 529-552.

Ganger, G. R., Khosla, P. K., Bakkaloglu, M., Bigrigg, M. W., Goodson, G. R., Oguz, S., . . . Wylie, J. J. (2001). *Survivable storage systems.* Paper presented at the Proceedings DARPA Information Survivability Conference and Exposition II. DISCEX'01.

Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iview, 1142*(2011), 1-12.

Gauci, M. G. (2019). WASP (Write a Scientific Paper): Interpretative phenomenological analysis: Its attraction and relevance to the medical field. *Early human development, 133*, 52-56.

Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika, 61*(1), 101-107.

Gerow, J. E., Grover, V., Roberts, N., & Thatcher, J. B. (2010). The diffusion of second-generation statistical techniques in information systems research from 1990-2008. *JITTA: Journal of Information Technology Theory and Application, 11*(4), 5.

Ghosh, B. (2018). Exploratory Study of Organizational Adoption of Cloud based Big Data Analytics. *Journal of Information Systems Applied Research, 11*(3), 4.

Gibbs, J. L., & Kraemer, K. L. (2004). A cross-country investigation of the determinants of scope of e-commerce use: an institutional approach. *Electronic markets, 14*(2), 124-137.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory : strategies for qualitative research*. Oxon, London: Routledge.

Gong, Y., & Janssen, M. (2017). Enterprise architectures for supporting the adoption of big data. *ACM Int. Conf. Proc. Ser. ACM International Conference Proceeding Series, Part F128275*, 505-510.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS quarterly*, 213-236.

Gregor, S., & R. Hevner, A. (2014). The Knowledge Innovation Matrix (KIM): A Clarifying Lens for Innovation. *InformingSciJ Informing Science: The International Journal of an Emerging Transdiscipline, 17*, 217-239.

Grover, V., Fiedler, K., & Teng, J. (1997). Empirical Evidence on Swanson's Tri-Core Model of Information Systems Innovation. *infosysres Information Systems Research, 8*(3), 273-287.

Hair J.F, S. M. R. C. M. M. J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *J. Acad. Mark. Sci. Journal of the Academy of Marketing Science, 40*(3), 414-433.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice, 19*(2), 139-152.
Hair, J. F., Ringle, C. M., & Sarstedt, M. (2014). Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. *Long Range Planning, 46*(1-2), 1-12. doi:10.1016/j.lrp.2013.08.016
Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*.
Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*: Sage publications.
Hair Jr, J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: updated guidelines on which method to use. *IJMDA International Journal of Multivariate Data Analysis, 1*(2), 107.
Hair Jr, J. F., Sarstedt, M., Matthews, L. M., & Ringle, C. M. (2016). Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I–method. *European Business Review*.
Halevi, G., & Moed, H. (2012). The evolution of big data as a research and scientific topic: overview of the literature. *Research trends, 30*(1), 3-6.
Hall, C. (2013). Augmenting Digital Marketing: via SaaS Business Intelligence.
Hameed, M. A., Counsell, S., & Swift, S. (2012). A conceptual model for the process of IT innovation adoption in organizations. *Journal of Engineering and Technology Management, 29*(3), 358-390.
Hanafizadeh, P., & Zare Ravasan, A. (2018). An empirical analysis on outsourcing decision: the case of e-banking services. *Journal of Enterprise Information Management, 31*(1), 146-172.
Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science, 43*(1), 115-135.
Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review, 34*(1), 135-174. doi:10.1111/dpr.12142
Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science (New York, N.Y.), 332*(6025), 60-65.
Hirsch, D., D. (2014). The Glass House Effect: Big Data, The New Oil, and the Power of Analogy. *Maine Law Review, 66*(2).
Hofstadter, D. R. (1979). Godel, Escher, Bach : an eternal golden braid. New York: Basic Books.
Hogan, M., & Shepherd, T. (2015). Information ownership and materiality in an age of big data surveillance. *Journal of Information Policy, 5*, 6-31.

Holst, A. (2021, 06/07/2021). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025 (in zettabytes) Retrieved from https://www.statista.com/statistics/871513/worldwide-data-created/

Hood-Clark, S. F. (2016). *Influences on the use and behavioral intention to use big data.* Capella University,

Hung, S.-Y., Huang, Y.-W., Lin, C.-C., Chen, K., & Tarn, J. M. (2016). *Factors Influencing Business Intelligence Systems Implementation Success in the Enterprises.* Paper presented at the PACIS.

IBM. (2013). Introduction to Structural Equation Modeling using IBM SPSS Statistics and AMOS. In (Second Edition ed.). doi:10.4135/9781526402257

IDC. (2018). 49% of Businesses Suffered Unrecoverable Data Event in Last 3 Years - StorageNewsletter. Retrieved from https://www.storagenewsletter.com/2018/08/23/49-of-businesses-suffered-unrecoverable-data-event-in-last-3-years/

IEEE. (2001). *IEEE 100 : the authoritative dictionary of IEEE standards terms*. New York: Institute of Electrical and Electronics Engineers.

Isaac, S., & Michael, W. B. (1997). *Handbook in research and evaluation : a collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences*. San Diego, Calif.: EdITS.

Jara, A. J., Genoud, D., & Bocchi, Y. (2014). Big Data in Smart Cities: From Poisson to Human Dynamics. 785-790.

Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: antecedents and consequences. *Journal of marketing, 57*(3), 53-70.

Johnson, M. P. (1989). Compatible information systems a key to merger success. *Healthcare financial management : journal of the Healthcare Financial Management Association, 43*(6), 60-61.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher, 33*(7), 14-26.

Joshi, M., & Biswas, P. (2018). *An Empirical Investigation of Impact of Organizational Factors on Big Data Adoption.* Paper presented at the Proceedings of First International Conference on Smart System, Innovations and Computing.

Jourová, V. (2016). *The EU Data Protection Reform and Big Data*. https://ec.europa.eu Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwj9gO-brunjAhUIX80KHYcuBJ4QFjAAegQIABAC&url=http%3A%2F%2Fec.e

uropa.eu%2Fnewsroom%2Fjust%2Fdocument.cfm%3Fdoc_id%3D41523&usg=AOvVaw0ST2KcGgcpQzrQ8t4JTG3F

Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry, 4*(4), 287-291.

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

Kim, M.-K., & Park, J.-H. (2017). Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare. *Information Development Information Development, 33*(3), 257-269.

Kune, R., Konugurthi, P. K., Agarwal, A., Chillarige, R. R., & Buyya, R. (2016). The anatomy of big data computing. *SPE Software: Practice and Experience, 46*(1), 79-105.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *JJIM International Journal of Information Management, 34*(3), 387-394.

Kwon, T. H., & Zmud, R. W. (1987). Unifying the fragmented models of information systems implementation. In *Critical issues in information systems research* (pp. 227-251).

Lamba, H. S., & Dubey, S. K. (2015). *Analysis of requirements for big data adoption to maximize IT business value.* Paper presented at the 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions).

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

Lee, Y., & Kozar, K. A. (2008). An empirical investigation of anti-spyware software adoption: A multitheoretical perspective. *Information & Management, 45*(2), 109-119.

Leedy, P. D., & Ormrod, J. E. (2016). *Practical research : planning and design* (Eleventh edition. ed.). Boston: Pearson.

Li, X., Lillibridge, M., & Uysal, M. (2011). Reliability analysis of deduplicated and erasure-coded storage. *SIGMETRICS Perform. Eval. Rev. ACM SIGMETRICS Performance Evaluation Review, 38*(3), 4.

Li, Y., Tan, C.-H., Teo, H.-H., & Siow, A. (2005). *A Human Capital Perspective of Organizational Intention to Adopt Open Source Software.* Paper presented at the ICIS.

Lin, H.-F. (2008). Empirically testing innovation characteristics and organizational learning capabilities in e-business implementation success. *Internet Research, 18*(1), 60-78.
Liu, Y., & Katramatos, D. (2019). *A Software Defined Network Design for Analyzing Streaming Data in Transit.* Paper presented at the Proceedings of the ACM Workshop on Systems and Network Telemetry and Analytics.
Liu, Z., Min, Q., & Ji, S. (2008). A Comprehensive Review of Research in IT Adoption. 1-5.
Liu, Z. H., & Krishnamurthy, V. (2012). *Towards Business Intelligence over Unified Structured and Unstructured Data Using XML*: INTECH Open Access Publisher.
Lombardo, G. (2018). *Predicting the Adoption of Big Data Security Analytics for Detecting Insider Threats.* Capella University,
Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE transactions on professional communication, 57*(2), 123-146.
Ma, X., Vazhkudai, S. S., & Zhang, Z. (2009). Improving Data Availability for Better Access Performance: A Study on Caching Scientific Data on Distributed Desktop Workstations. *J Grid Computing Journal of Grid Computing, 7*(4), 419-438.
Mach-Król, M. (2017). Big Data analytics in Polish companies–selected research results. *ICT Management for Global Competitiveness and Economic Growth in Emerging Economies (ICTM)*, 85.
Maddox, T. (2015). Research: Local data storage preferred by 69 percent. Retrieved from https://www.zdnet.com/article/research-69-prefer-to-store-data-locally/
Mahesh, D. D., Vijayapala, S., & Dasanayaka, S. W. S. B. (2018). Factors Affecting the Intention to Adopt Big Data Technology : A Study Based on Financial Services Industry of Sri Lanka. 420-425.
Malladi, S., & Krishnan, M. (2013). Determinants of usage variations of business intelligence & analytics in organizations–an empirical analysis.
Mapstone, J., Elbourne, D., & Roberts, I. (2007). Strategies to improve recruitment to research studies. *Cochrane Database Syst Rev, 2*.
Marzullo, K. (2016). Administration Issues Strategic Plan for Big Data Research and Development. Retrieved from https://obamawhitehouse.archives.gov/blog/2016/05/23/administration-issues-strategic-plan-big-data-research-and-development
Mathai, N. (2019). *Factors influencing health care consumer adoption of electronic health records: An empirical investigation.* Murdoch University,

Mayer-Schönberger, V., & Cukier, K. (2017). *Big data : a revolution that will transform how we live, work and think*.
McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review, 90*(10), 60-66.
McCracken, G. (1988). *The long interview* (Vol. 13): Sage.
McGuire, T., Manyika, J., & Chui, M. (2012). WHY BIG DATA IS THE NEW COMPETITIVE ADVANTAGE. *Ivey Business Journal, 76*(4).
Melby, C. (2013). Hortonworks Wants To Own Big Data Without Owning Anything. *Forbers*. Retrieved from https://www.forbes.com/sites/calebmelby/2013/12/19/hortonworks-wants-to-own-big-data-without-owning-anything/#70706ae23132
Melone, N. P. (1990). A Theoretical Assessment of the User-Satisfaction Construct in Information Systems Research. *Management Science Management Science, 36*(1), 76-91.
Mesnier, M. P., & Akers, J. B. (2011). Differentiated storage services. *SIGOPS Oper. Syst. Rev. ACM SIGOPS Operating Systems Review, 45*(1), 45.
Micheni, E. M. (2015). Diffusion of big data and analytics in developing countries.
Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management, 16*(3), 547-578.
Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research, 70*, 1-16.
Miller, N. E., & Dollard, J. (1979). *Social learning and imitation*. Westport, Conn.: Greenwood Press.
Mingers, J. (2001). Embodying information systems: the contribution of phenomenology. *Information and organization, 11*(2), 103-128.
Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information systems research, 2*(3), 192-222.
Morris, R. J. T., & Truskowski, B. (2003). The evolution of storage systems. *IBM SYSTEMS JOURNAL, 42*, 205-217.
Motau, M., & Kalema, B. M. (2016). Big Data Analytics readiness: A South African public sector perspective. 265-271.
Naing, L., Winn, T., & Rusli, B. (2006). Practical issues in calculating the sample size for prevalence studies. *Archives of orofacial Sciences, 1*, 9-14.
Nam, D. W., Kang, D., & Kim, S. H. (2015). Process of big data analysis adoption: Defining big data as a new IS innovation and examining factors affecting the process. *Proc. Annu. Hawaii Int. Conf. Syst. Sci. Proceedings*

of the Annual Hawaii International Conference on System Sciences, 2015-March, 4792-4801.
Netapp. (2019). Data Standards – Data Security and Management Standards. Retrieved from https://www.netapp.com/us/company/leadership/industry-standards/index.aspx
Newman, N. (2017). Journalism, media, and technology trends and predictions 2017.
Nguyen, T., & Petersen, T. E. (2017). *Technology adoption in Norway: organizational assimilation of big data.*
Nielsen, J. (1999). User interface directions for the web. *Communications of the ACM, 42*(1).
Ochieng, G. (2015). *The adoption of big data analytics by supermarkets in Kisumu County.* Doctoral dissertation, University of Nairobi,
Olszak, C. M., & Mach-Król, M. (2018). Conceptual Framework for Assessing Organization's Readiness to Big Data Adoption.
Park, J.-H., Kim, M.-K., & Paik, J.-H. (2015). The Factors of Technology, Organization and Environment Influencing the Adoption and Usage of Big Data in Korean Firms.
Parkins, D. (2017). Regulating the internet giants: The world's most valuable resource is no longer oil, but data. *Economist (United Kingdom), 413*(9035).
Parthasarathy, R. (2017). *Empirical Assessment of the Role of Technology-Related Factors and Organization-Related Factors in Electronic Medical Records Implementation Success*. Retrieved from OAIster database. DePaul University.
Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability maturity model, version 1.1. *IEEE software, 10*(4), 18-27.
Premkumar, G., & Roberts, M. (1999). Adoption of new information technologies in rural small businesses. *OMEGA -OXFORD- PERGAMON PRESS-, 27*(4), 467-484.
Prescott, M. B., & Conger, S. A. (1995). Information technology innovations a classification by IT locus of impact and research approach. *SIGMIS Database ACM SIGMIS Database, 26*(2-3), 20-41.
Puklavec, B., Oliveira, T., & Popovič, A. (2014). Unpacking business intelligence systems adoption determinants: An exploratory study of small and medium enterprises. *Economic & Business Review, 16*(2).
Qingchen, Z., Zhikui, C., Ailing, L., Liang, Z., Fangyi, L., Jian, Z., . . . Social Computing Beijing, C. A. A. (2014). A Universal Storage Architecture for Big Data in Cloud Environment. In *2013 IEEE International Conference on Green Computing and Communications (GreenCom) and IEEE*

Internet of Things(iThings) and IEEE Cyber, Physical and Social Computing(CPSCom) (pp. 476-480).
Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative research in accounting & management, 8*(3), 238-264.
Qualtrics. (2020). Sample Size Calculator. Retrieved from https://www.qualtrics.com/blog/calculating-sample-size/
Ramdani, B., Lorenzo, O., & Kawalek, P. (2009). Information Systems Innovations Adoption and Diffusion Among SMEs Current Status and Future Prospects. *International Journal of E-Adoption, 1*(1), 33-45.
Reinsel, D. (2013). Where in the World Is Storage: A Look at Byte Density Across the Globe. Retrieved from https://www.idc.com/downloads/where_is_storage_infographic_243338.pdf
Reinsel, D., Gantz, J., & Rydning, J. (2017). Data age 2025: The evolution of data to life-critical don't focus on big data. *Focus on the Data That's Big Sponsored by Seagate The Evolution of Data to Life-Critical Don't Focus on Big Data*.
Rivituso, J. (2014). Cyberbullying victimization among college students: An interpretive phenomenological analysis. *Journal of Information Systems Education, 25*(1), 71.
Rogers, E. M. (1962). *Diffusion of innovations*. New York: Free Press of Glencoe.
Rogers, E. M. (2003). Diffusion of innovations. Free Press. *New York, 551*.
Romijn, J. H. (2014). *Using Big Data in the Public Sector. Uncertainties and Readiness in the Dutch Public Executive Sector*. Retrieved from OAIster database.
RUI, G. (2007). *Information systems innovation adoption among organizations-A match-based framework and empirical studies.*
Russom, P. (2011). Big Data Analytics, TDWI best practices report. *Fourth quarter*, 1-35.
Rydning, J. S., Michael. (2021, 03/24/2021). Data Creation and Replication Will Grow at a Faster Rate than Installed Storage Capacity, According to the IDC Global DataSphere and StorageSphere Forecasts. Retrieved from https://www.idc.com/getdoc.jsp?containerId=prUS47560321
Salleh, K. A., & Janczewski, L. J. (2018). *An Implementation of Sec-TOE Framework: Identifying Security Determinants of Big Data Solutions Adoption.* Paper presented at the PACIS.
Sarstedt, M., & Mooi, E. (2014). A Concise Guide to Market Research The Process, Data, and Methods Using IBM SPSS Statistics. Retrieved from http://link.springer.com/book/10.1007/978-3-642-53965-7
Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. *Handbook of market research, 26*, 1-40.

Schmitt, P., Thiesse, F., & Fleisch, E. (2007). Adoption and diffusion of RFID technology in the automotive industry. *Auto-ID Labs White Paper# WP-BIZAPP-041*.

Selznick, P. (1948). *Foundations of the theory of organization*. Indianapolis, Ind.: Bobbs-Merrill, College Division.

Sh. Hajirahimova, M., & S. Aliyeva, A. (2017). About Big Data Measurement Methodologies and Indicators. *IJMECS International Journal of Modern Education and Computer Science, 9*(10), 1-9.

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *JBR Journal of Business Research, 69*(10), 4552-4564.

Shmueli, G., Sarstedt, M., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., . . . Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *Eur. J. Mark. European Journal of Marketing*.

Siegel, E. (2016, September 21, 2016). 9 Bizarre and Surprising Insights from Data Science. Retrieved from https://blogs.scientificamerican.com/guest-blog/9-bizarre-and-surprising-insights-from-data-science/

Siepmann, L., & Nicholas, K. (2018). German Winegrowers' Motives and Barriers to Convert to Organic Farming. *Sustainability, 10*(11), 4215.

Simsek, Z., & Veiga, J. F. (2001). A primer on internet organizational surveys. *Organizational research methods, 4*(3), 218-235.

Siriweera, T. H. A. S., Paik, I., Zhang, J., Kumara, B. T. G. S., & Services, I. I. C. o. W. (2016). Big Data Analytic Service Discovery Using Social Service Network with Domain Ontology and Workflow Awareness. 324-331.

Skaale, D. K., & Rygh, E. (2018). *Big Data Technology Adoption Through Digitalization in Yara International ASA.* University of Stavanger, Norway,

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*: Sage.

Smith, J. A., & Shinebourne, P. (2012). *Interpretative phenomenological analysis*: American Psychological Association.

Soon, K. W. K., Lee, C. A., & Boursier, P. (2016). A study of the determinants affecting adoption of big data using integrated Technology Acceptance Model (TAM) and diffusion of innovation (DOI) in Malaysia. *Int. J. Appl. Bus. Econ. Res. International Journal of Applied Business and Economic Research, 14*(1), 17-47.

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *jroyastatsocise2 Journal of the Royal Statistical Society. Series B (Methodological), 36*(2), 111-147.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of personality assessment, 80*(3), 217-222.
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.
Swanson, E. B. (1994). Information systems innovation among organizations. *Management Science, 40*(9), 1069-1092.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed. ed.). Boston: Pearson/Allyn & Bacon.
Tambe, P. (2014). Big data investment, skills, and firm value. *Management Science, 60*(6), 1452-1469.
Tornatzky, L. G., & Fleischer, M. (1990). *The process of technological innovation*. Massachusetts: Lexington Books.
Triandis, H. C. (1980a). *Handbook of cross-cultural psychology. 6 6*. Boston, Mass.: Allyn and Bacon.
Triandis, H. C. (1980b). Reflections on Trends in Cross-Cultural Research. *Journal of Cross-Cultural Psychology Journal of Cross-Cultural Psychology, 11*(1), 35-58.
University, D. (2019). Getting Started. Retrieved from https://offices.depaul.edu/research-services/research-protections/irb/Pages/getting-started.aspx
Van der Meulen, R., & Woods, V. (2015). Gartner survey shows more than 75 percent of companies are investing or planning to invest in big data in the next two years. *Gartner< http://www. gartner. com/newsroom/id/3130817>, haettu, 5*, 17.
Van Hoye, G., Van Hooft, E. A., & Lievens, F. (2009). Networking as a job search behaviour: A social network perspective. *Journal of Occupational and Organizational Psychology, 82*(3), 661-682.
Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly MIS Quarterly, 27*(3), 425.
Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS quarterly*, 21-54.
Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186-204.
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.

Verma, S., & Bhattacharyya, S. S. (2017). Perceived strategic value-based adoption of Big Data Analytics in emerging economy: A qualitative approach for Indian firms. *J. Enterp. Inf. Manage. Journal of Enterprise Information Management, 30*(3), 354-382.

Viechtbauer, W., Smits, L., Kotz, D., Budé, L., Spigt, M., Serroyen, J., & Crutzen, R. (2015). A simple formula for the calculation of sample size in pilot studies. *Journal of Clinical Epidemiology, 68*(11), 1375-1379.

Wagner, J. (1997). The Unavoidable Intervention of Educational Research: A Framework for Reconsidering Researcher-Practitioner Cooperation. *Educational researcher, 26*(7), 13-22.

Walter, C. (2005). Kryder's law. *Scientific American, 293*(2), 32-33.

Wang, G., Dou, W., Zhu, W., & Zhou, N. (2015). The effects of firm capabilities on external collaboration and performance: The moderating role of market turbulence. *Journal of Business Research, 68*(9), 1928-1936.

Wang, L., Yang, M., Pathan, Z. H., Salam, S., & Shahzad, K. (2018). Analysis of Influencing Factors of Big Data Adoption in Chinese Enterprises Using DANP Technique. *Sustainability, 10*(11).

Westervelt, R. (2017). IDC White Paper: Information-Centric Security: Why Data Protection Is the Cornerstone of Modern Enterprise Security Programs, March 2017. *URL: symantec. com› content/dam*.

Wong, K. K.-K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin, 24*(1), 1-32.

Wong, S., & Gray, J. (2019). *Barriers to implementing Building Information Modelling (BIM) in the Malaysian construction industry.* Paper presented at the IOP Conference Series: Materials Science and Engineering.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of computer-mediated communication, 10*(3), JCMC1034.

Xian, H. (2013). Scholarly collaboration in engineering education: From big-data scientometrics to user-centered software design. Retrieved from http://docs.lib.purdue.edu/dissertations/AAI3613513

Xiaoqiang, M., Xiaoyi, F., Jiangchuan, L., Hongbo, J., & Kai, P. (2017). vLocality: Revisiting Data Locality for MapReduce in Virtualized Clouds. *IEEE Network, 31*(1). doi:10.1109/MNET.2016.1500133NM

Xin, Q. (2005). *Understanding and coping with failures in large-scale storage systems.* Available from http://worldcat.org /z-wcorg/ database.

Yadegaridehkordi, E., Hourmand, M., Nilashi, M., Shuib, L., Ahani, A., & Ibrahim, O. (2018). Influence of big data adoption on manufacturing companies' performance: An integrated DEMATEL-ANFIS approach. *TFS Technological Forecasting & Social Change, 137*, 199-210.

Yardley, L. (2000). Dilemmas in qualitative health research. *Psychology and health, 15*(2), 215-228.

Yeh, C.-H., Lee, G.-G., & Pai, J.-C. (2015). Using a technology-organization-environment framework to investigate the factors influencing e-business information technology capabilities. *Information Development, 31*(5), 435-450.

Yianilos, P. N., & Sobti, S. (2001). The evolving field of distributed storage. *IEEE Internet Computing, 5*(5), 35-39.

Yin, R. K. (2003). *Case study research : design and methods*. Thousand Oaks, Calif.: Sage Publications.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology, 9*(2), 79-94.

Zanabria, V., & Mlokozi, D. (2018). Big data analytics for achieving smart city resilience Key factors for adoption. Retrieved from http://lup.lub.lu.se/student-papers/record/8950489