Understanding Big Data Analytics (BDA) for quality decision making in the IT sector – A systematic literature review

Kabir Bhatia

A dissertation submitted to
Auckland University of Technology
in partial fulfilment of the requirements for the degree
of
Master of Business (MBus) – Information Systems

2025

Department of Business Information Systems
Faculty of Business, Economics and Law

Supervisor: Ranjan Vaidya

Abstract

Big Data Analytics (BDA) has become an innovative influence in the IT sector, allowing companies to make data-driven, high-quality decisions. This dissertation studies the factors impacting BDA adoption and analyzes the primary tools and technologies employed to enhance quality decision-making. The study uses a systematic literature review to integrate findings from previous research, offering an in-depth understanding of the role, constraints, and advantages of BDA.

The study defines key BDA tools, comprising data storage and management systems (e.g. Amazon Redshift, MongoDB), distributed computing frameworks (e.g., Apache Hadoop, Spark), AI-driven analytics instruments (e.g., TensorFlow, Scikit-learn), and business intelligence platforms (e.g., Tableau, Power BI). These technologies allow businesses to manage extensive, complex datasets effectively, derive significant insights, and improve operational efficiency.

In addition to technology, non-technological elements like leadership commitment, budget limitations, workforce competencies, and organizational culture substantially impact BDA implementation. The study indicates that businesses with strong data governance, multidisciplinary teamwork, and investment in staff training are more effective in utilizing analytics. However, challenges including increased implementation costs, a lack of competent staff, and reluctance to change persist as barriers to extensive adoption.

Keywords: *Big Data Analytics, Quality Decision Making, Tools and Technology*

Table of Contents

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor used artificial intelligence tools or generative artificial intelligence tools (unless it is clearly stated, and referenced, along with the purpose of use), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning

26th February 2025

Kabir Bhatia

Acknowledgments

I wish to convey my heartfelt appreciation to all who have assisted me during my research work. I would like to express my gratitude to my supervisor, Ranjan Vaidya, for his direction, insightful feedback, and recommendations. I am extremely thankful for his crucial support throughout my dissertation, from developing the research question to reviewing the final draft. I extend my sincere gratitude to my colleagues, Hans Higgins, Maiya Sargent, and Philip Thorogood, for covering my responsibilities during my study leave, allowing me to devote my whole attention to this research.

Chapter 1: Introduction

Over the past few decades, the IT industry has grown and changed in ways never seen before because of developments in data gathering, storage, and processing power. In today's age and with the advancements, in technology we see a growth in the amount of data being produced daily by various companies and sectors worldwide. Computers have significantly increased in power, networking is widespread, and algorithms have been created that can link datasets to allow more extensive and deeper studies than were previously achievable (Provost & Fawcett, 2013). Many businesses are now turning to BDA using it to gain insights and make decisions. The term "big data" is widely used in both industry and academics (Weerasinghe, 2019). The idea of Big Data has become an essential component of the modern technology environment due to the explosion of data generation from a variety of sources, including as social media platforms, Internet of Things (IoT) devices, e-commerce transactions, and organizational databases.

Big data has become an increasingly popular term in computer science research and industry due to the exponential increase and accessibility of data (Kumar, 2015). The term "big data" refers to extensive, complex, and rapidly growing databases that ordinary data processing methods struggle to manage and analyze. Big data is defined as data characterized by high volume (the amount or quantity of data), high velocity (the rate at which data is created), and high variety (the different types of data); requiring innovative methods and technologies for its capture, storage, and analysis; and utilized to improve decision-making, generate insights, and optimize processes (Watson, 2014). This development requires the invention of specific tools, technologies, and processes for data processing, storage, and analysis. Big data emerged to characterize datasets that are excessively huge, complex, and rapidly developing, making them unmanageable and unsolvable by conventional data management tools. To address complex business intelligence requirements, big data environments require specialized software and technology platforms linked into a single process (Phillips-Wren et al., 2015).

Utilizing tools and methods, in BDA helps in examining complex datasets to reveal patterns and correlations essential, for businesses to stay ahead in today's data centric environment. Over the decade or so, Business analytics has become a focus, within enterprise systems. Nowadays, many views BDA as a new development that is set to revolutionize the fields of analytics and business intelligence significantly (Danielsen et al., 2021). Liu et al. (2018) believe that one of the main functional areas in many firms is anticipated to be business analytics, a subset of business intelligence that helps companies compete successfully in the market. A standard definition states that "Business Intelligence (BI) covers a wide range of applications, technologies, and processes for collecting, storing, accessing, and analyzing data to assist business users in making informed decisions" (Watson, 2009; Watson, 2014). According to this definition, BI is seen as a broad word including all applications that help with decision-making, and has been adopted in both industry and, increasingly, academia (Watson, 2014). For companies BDA has emerged as an asset and has consistently ranked high among the priorities of top-level executives in recent years. The use of BDA has become crucial for managing complex systems, forecasting market trends, and maximizing performance in the IT industry, where technical innovation drives growth. Businesses are examining and storing data from sources such, as social media platforms devices, like log files or GPS technology with varying volume speed and types to make informed decisions (Watson, 2014).

BDA has emerged as a key tool for businesses looking to maximize the value of their data assets. The volume of data has increased due to, among other variables, the more detailed gathering of information (Waller & Fawcett, 2013). Businesses can enhance customer experiences, streamline operations, find new revenue sources, and make well-informed decisions by utilizing these huge databases. According to Wang and Huang (2015; Weerasinghe, 2019) businesses will only be able to manage uncertainty and generate value from big data with such advanced analytics tools. For example, businesses can use consumer behavior data to customize their advertising strategies and use predictive maintenance to forecast equipment failures. IT infrastructure, human capital, and business architecture are said to be the three main elements that influence BDA capacity. A firm's capacity to acquire ambidextrous skills and techniques—which lead to a competitive advantage position—is strongly influenced by BDA (Bedeley & Nemati, 2014). BDA has the potential to significantly change how companies perform and compete. As a result, a lot of businesses are investing more in big data and are keen to implement analytics to get the most out of their investment (Duan et al., 2019).

1.1 Problem Statement

However, many businesses are facing challenges, in understanding and managing the amount of data coming from sources both internal and external in different formats (Garmaki et al., 2023). The tools and methods for storing, analyzing, and visualizing data have advanced significantly within BDA. Researchers, however, have focused less on whether and under what circumstances the development of BDA solutions creates value for businesses (Jensen et al., 2019).

BDA has emerged as a crucial capability for businesses, especially within the IT sector, where creativity and flexibility are important for success. Despite its cutting-edge potential, many organizations have difficulties in properly utilizing BDA for quality decision- making. Although BDA can serve as a valuable asset, its execution poses considerable challenges (Jensen et al., 2021). A global survey revealed that 43% of enterprises derive minimal to no benefit from BDA, and over half of these programs fail to meet their strategic goals (Mithas et al., 2013; Jensen et al., 2021).

The use of big data is a process by businesses to discover innovative methods to enhance efficiency and minimize risk, while more efficiently addressing client needs (Baig et al., 2019). The utilization of big data has unlimited potential; however, it can only be utilized by the technology, resources, and skills accessible for BDA (Sivarajah et al., 2017).

The IT sector has various advanced tools and technologies; however, the efficient use of BDA is frequently prevented by multiple challenges, including technological, organizational, and human issues. The vast quantities of data generated and managed by corporations and research institutions are getting closer to the limits of their data infrastructure (Kadadi et al., 2014). Some experts believe that large databases are not always complex, whereas small datasets are not necessarily simple, hence contributing to the complexity of big data and highlighting that a dataset's complexity is essential in determining its size of big data (Sivarajah et al., 2017).

These challenges highlight the necessity for a thorough understanding of the factors influencing BDA acceptance and applications. An immature and insufficient BDA infrastructure clearly leads to numerous issues in areas such as data management (Hirschlein et al., 2022). The successful application of various techniques and technologies depends on overcoming these barriers. Research on this subject is divided, concentrating on single elements like technology or skills, without a comprehensive analysis of their connection. This research aims to fill this gap by performing an in-depth literature review to identify the tools and technology utilized in BDA, along with the aspects that affect its implementation for quality decision-making in the IT sector. The study aims to analyse existing research to offer relevant insights and recommendations for IT organizations to enhance their utilization of BDA and address ongoing difficulties. Despite significant research on Big Data and its uses, very few studies fully assess the interaction of tools, technologies, and factors that influence of quality decision-making in the IT sector. Therefore, the research questions that direct this dissertation are:

Question 1. Which BDA technologies and tools have been highlighted in academic studies focused on the IT industry?

Question 2. Within these academic studies, what are the factors that impact the applications of BDA for quality decision-making in the IT sector?

The first question aims to examine the tools and technologies that assist businesses in the efficient storage, processing, and analysis of Big Data. The study will concentrate on classifying these tools according to their characteristics, including data storage (e.g., Hadoop), processing tools (e.g., Apache Spark), tools for visualization (e.g., Tableau), and advanced analytics applications (e.g., TensorFlow for machine learning). It aims to provide a thorough analysis of the cutting-edge technologies facilitating BDA adoption within the IT sector.

The later question studies the business, technological, and human elements that affect the effective implementation and utilization of BDA for decision-making. It seeks to identify barriers such as issues with data quality, integration problems, insufficient skills, disagreements to change, and economic constraints. It also examines the factors that promote successful BDA adoption, including leadership acceptance, data governance procedures, and innovations in automation and machine learning. The research questions aim to address significant gaps in the current literature.

The dissertation attempts to provide practical solutions for IT businesses aiming to enhance their data-driven decision-making methods, assist practitioners and policymakers in comprehending the factors that improve the impact of BDA adoption and establish a foundation for future research that combines technological, organizational, and human factors in BDA studies.

1.2 Structure of the Dissertation

This section breaks down an outline for the reader to follow, specifying the topics and objectives of each chapter. This dissertation includes six chapters – Introduction, Research Method, Literature Review, Findings and Discussion, Conclusion.

This Introduction provides the context for the research by presenting the background of BDA within the IT sector. It provides the relevance of the study in the environment of BDA, presents the research questions, and outlines the dissertation structure. Chapter 2 defines the research method used for the study. It presents the details of the systematic literature review (SLR) method, and the rationale for its suitability in the current research. The text presents the details of the literature gathering method, such as the criteria for the selection of research articles and the databases searched. The reasons for inclusion and deletion are thoroughly examined to guarantee transparency. Additionally, the chapter describes the data analysis method used to draw the research findings.

The next chapter i.e. chapter 3 presents the literature review, the literature review offers an in-depth study of current research related to the two research questions presented above. It analyzes essential subjects and technologies employed in BDA, the organizational and technological factors influencing its adoption, and the importance of data analytics in IT decision-making. This chapter also outlines research gaps, providing a basis for the theoretical framework of the study and explaining the research topics.

Chapter 4 explains the Findings and Discussions derived from the systematic literature review. The results are organized according to the main themes identified during the literature review. It offers an analysis of the tools and technologies typically employed in BDA, and the organizational and technological factors that impact the application of BDA for making quality decision-making in the IT sector. In this chapter, the results are compared with current literature to highlight consistencies, discrepancies, and emerging behaviors. The findings of the research for IT companies and decision-makers are analyzed, especially regarding the enhancement of BDA adoption and the optimization of decision-making processes. This chapter also addresses the research questions, so strengthening the study's contributions to knowledge and practice in BDA.

The last chapter (Conclusion) summarizes the main findings of the study, highlighting their relevance to the research questions. IT organizations are offered realistic guidelines to optimize their utilization of BDA for strategic decision-making. This chapter also mentions the limitations and discusses Implications for practice. The chapter outlines the studies restricts and proposes directions for future research, providing insights into areas requiring more investigation to enhance comprehension in the field of BDA.

The reference section contains all sources cited in the dissertation, adhering to the required APA 7th academic citation format.

Chapter 2: Research Method

Research Method is an area of study that examines the scientific processes involved in conducting research (Patel & Patel, 2019). Research method is a systematic plan or strategy for addressing a research problem (Kothari, 2004). The implementation of the plan might include data collection and analysis, involvement of participants, and utilization of tools or technology (Hudson & Ozanne, 1988; Kivunja & Kuyini, 2017). The method discusses both the outcomes of scientific investigation, and the process involved (Patel & Patel, 2019).

2.1 Systematic Literature Review

A systematic literature review (SLR) was chosen as the research method for this study. Systematic literature reviews (SLRs) integrate scientific evidence to address a specific research issue transparently and accurately, aiming to encompass all published information on the subject and evaluate the quality of this evidence (Nightingale, 2009). It is an organized and planned process for finding, selecting, and critically assessing current research relevant to a certain subject, compared to conventional literature reviews, which may involve subjective study selection, a systematic literature review refers to a well-established procedure to guarantee thorough and fair analysis. The principal aim of a systematic literature review (SLR) is to deliver a methodical examination of existing research to find patterns, problems, and new trends, making it especially beneficial in domains undergoing swift technological progress, such as BDA within the information technology sector. A systematic method ensures that only high-quality, peer-reviewed, and relevant studies inform the conclusions of this dissertation, considering the extensive research on BDA applications, tools, and decision-making processes.

This research was conducted predominantly utilizing a qualitative method by carefully assessing every relevant source and analyzing the findings. The method employed for addressing the research question remained primarily inductive. A thematic analysis was utilized to examine the qualitative data gathered from secondary literature on the application of BDA and how businesses can use BDA to make a quality decision focusing in the IT sector.

2.1.1 Methods

This section offers a comprehensive description of the methods employed for data collection and analysis, including the processes utilized to gather and transform raw data into evidence relevant to the research questions. The SLR describes both the content and quality of existing knowledge, effectively expressing the importance of previous studies to the readers (Okoli & Schabram, 2010). This dissertation is directed by the research questions mentioned above to maintain the focus and structure of the SLR. These questions correspond with the study's aims, simplifying the structure of the review process and concentrating on the most appropriate academic findings.

For inclusion and exclusion of the articles, a thorough search approach was utilized to find appropriate literature. Data were gathered from several library databases (listed below) to mitigate the problems of incomplete, dated, incorrect or biased information. The primary focus remained on academic articles, theses, and conference papers on the factors impacting BDA application in the IT sector. The following electronic databases were searched for related studies:

- Science Direct
- Scopus
- AIS eLibrary
- Google scholar

The data collection methods included searches for keywords, titles, authors, and citations. The search phrases covered combinations of the following keywords:

- Big data analytics
- decision making
- IT sector
- BDA tools
- Data visualisation
- predictive analytics
- decision making
- data analytics
- data driven decision making.

Boolean search strategies were utilized to widen or limit the range of search results. For example, the combinations of the following keywords were searched. “Big Data Analytics” AND “IT Sector” AND “Decision Making” “BDA Tools” AND “Decision Making” “Data Analytics” AND “Data Visualisation” AND “Data Driven Decision Making”

References cited in each article were regarded as potential information sources and investigated as necessary to expand the search parameters.

To guarantee the selection of relevant and high-caliber studies, defined inclusion and exclusion criteria were utilized. The literature search focuses on research published in the time frame of last 15 years (2009–2024) to assure the inclusion of the newest developments in BDA. The articles published in languages other than English, studies not relevant to the research questions and articles that are not accessible were excluded. Greater emphasis was placed on papers that provide proof of observations and discussions with stakeholders. Non-journal literature from computer magazines did not provide detailed instances and hence were excluded from search results. All references were maintained in the Endnote software application. Duplicate entries were identified and eliminated. The following step involved evaluating the gathered documents and eliminating those that weren't related to BDA applications within the IT sector or tools and technology that facilitate quality decision-making. In addition to the language and the relevant timeframe, additional filters such as document kinds and subject categories were applied to the sources. Document types were limited to journals or articles, thesis and conference papers and topic areas included computer science, social sciences, economics, business management and Finance. Articles by the same author related to a different case study were reevaluated for potential inclusion. This procedure considerably lowered the quantity of articles. Literature that lacked actual data or case studies were excluded from research.

Source	Number of Articles	Relevant	Accepted
Science Direct	2511	22	5
Scopus	1271	23	8
AIS eLibrary	3501	26	5
Google Scholar	5677	35	7

Table 1: Article Inclusion Summary

After thorough inclusion and exclusion, 25 articles (Table 1) obtained from the comprehensive literature search formed the data set for this investigation. This was the necessary minimum quantity of articles within the allotted timeframe for this research. The 25 selected papers were used to construct a literature matrix. Table 2 presents the first five entries from the literature matrix and the full matrix is included in appendix 1. The subsequent column headings are as follows:

- Serial Number (S.N.)
- Title of the article
- Author
- Publication
- Year
- Type
- Reference

S.N.	Title	Author	Publication	Year	Type	Reference
1	Toward an understanding of big data analytics and competitive performance	Danielsen, F., Olsen, D. H., & Framnes, V. A.	Association for information systems	2021	Journal	(Danielsen et al., 2021)
2	Big data analytics capability and contribution to firm performance: the mediating effect of organizational learning on firm performance	Garmaki, M., Gharib, R. K., & Boughzala, I.	Emerald Insight	2023	Journal	(Garmaki et al., 2023)
3	Tutorial: Big Data Analytics: Concepts, Technologies, and Applications	Watson, H. J	Association for information systems	2014	Journal	(Watson, 2014)
4	MANAGING BIG DATA ANALYTICS PROJECTS: THE CHALLENGES OF REALIZING VALUE	Jensen, M. H., Nielsen, P. A., & Persson, J. S	Association for information systems	2019	Journal	(Jensen et al., 2019)
5	The Challenges of Business Analytics: Successes and Failures	Liu, Y., Han, H., & DeBello, J.	Proceedings of the 51st Hawaii International Conference on System Sciences	2018	Journal	(Liu et al., 2018)

Table 2: Literature Matrix

2.1.2 Data Analysis

Following the selection of relevant papers, data analysis was performed using a thematic analysis method. This involved carefully classifying and integrating results from several research to discern patterns, themes, and limitations concerning BDA and its impact on decision- making in the IT industry. Thematic analysis (TA) comprises methods for identifying and comprehending patterns of meaning in qualitative data (Braun & Clarke, 2006). The data analysis procedure included the comprehensive examination (i.e. providing codes and generating themes) of 25 articles that contained applicable information on BDA and its influence on decision-making within the IT sector, as determined during the data collecting process.

The study was performed in four primary stages: data extraction, thematic coding, synthesis and interpretation, and validation and reliability assessments. Every phase was important in organizing the insights obtained from the literature, allowing to answer the research questions concerning the tools, technologies, and factors affecting BDA in decision-making within the IT industry.

After extracting the relevant information (by providing the codes mentioned in Appendices), the next stage involved to identify themes, a qualitative method used to determine and classify main topics within the literature. Thematic coding was conducted in three primary phases. The first phase is open coding which involves an initial examination of the collected data, during which significant keywords, phrases, and concepts were emphasized without the application of defined categories. The second phase was axial coding and during this stage, the accepted ideas were categorized into broader themes according to their similarities and interactions. The third phase is selective coding that involves refining the themes to ensure they precisely captured the primary points from the literature. Technological factors, Non-Technological Factors (Organizational Factors and Human Factors) and Ethical and Privacy Factors are one of the main themes identified from this method.

Following to thematic coding, the next stage is synthesizing the findings to identify significant insights. This stage was essential for comparing results across many studies, finding common trends, underlining best practices, and finding research gaps It starts with identifying common trends in BDA adaption. Research frequently indicates that businesses showing strong leadership support, a well-defined data strategy, and investment in current BDA tools are more effective in utilizing analytics for decision-making. Numerous studies highlighted the significance of a specialized analytics team, the addition of real-time analytics, and the assurance of data quality to maximize the effectiveness of BDA. Exploring Gaps in the Literature - Although numerous studies have addressed technological and organizational factors, there is a lack of study regarding the adoption of BDA by small and medium-sized organizations (SMEs) in comparison to large corporations. This gap indicates possibilities for future investigation.

To guarantee the accuracy and dependability of the results the findings were compared with prior systematic literature reviews to confirm consistent trends and themes.

On top of that, a quality assessment of the included studies was conducted utilizing specified evaluation criteria, including effectiveness to BDA applications in the IT sector.

Chapter 3: Literature Review

This chapter presents a descriptive account of the main themes related to the research questions. The discussion is organized around many themes, first with an overview of BDA, including its definition, scope, and development. The chapter presents the main tools and technologies employed in BDA, focusing advancements in data storage, processing, visualization, and machine learning. Additionally, it presents the factors impacting the adoption and implementation of BDA, classified into technological, organizational, and human elements. The important role of BDA in improving quality decision-making processes within the IT sector is also described as a thematic category in this chapter. The chapter identifies research gaps and potential methods as important themes, highlighting areas requiring more research to enhance BDA processes inside IT firms.

3.1 BDA: Key Concepts and Terminologies

3.1.1 BDA Definition

Business Analytics is a component of Business Intelligence that empowers organizations to compete effectively in the marketplace and is projected to become a main functional area for several enterprises (Liu et al., 2018). (Favaretto et al., 2020; Dick, 2022) define big data as a wide range of concepts, including the collecting and consolidation of extensive data sets, as well as a variety of complex digital methods aimed at finding patterns associated with human behaviour. BDA significantly transforms various industries by allowing companies to use extensive structured and unstructured data for quality decision-making. Its scope includes multiple sectors, including customer analytics, operational efficiency, risk management, and fraud detection. Businesses globally are employing BDA to enhance their business operations, secure competitive advantages, and improve consumer experiences. Big data refers to extensive data collections that include significant volume, diversity, and complexity, presenting challenges in storage, analysis, and visualization for subsequent processes or outcomes (Sagiroglu et al., 2013). Business analytics have been the most important enterprise system topic of the last decade, and currently, BDA is regarded as an innovative technology set to fundamentally transform analytics and business intelligence (Danielsen et al., 2021). Businesses are gradually adopting analytics to optimize the value generated by their operations, as these technologies become critical distinctive advantages across many industries (Abbasi et al., 2016). A significant application of BDA is in the analysis of customer behavior and customized marketing strategies. Corporations such as Amazon and Netflix use machine learning algorithms and recommendation systems to examine consumer preferences and consumption behaviours. Also, we know that Apache Hadoop is one of the primary open-source tools utilized in the market that utilizes distributed architecture for managing data processing challenges (Nandimath et al., 2013). BDA is an advanced procedure for examining wide and complicated data sets to find fundamental patterns, correlations, market trends, consumer preferences, and other critical business insights. Enterprises are gathering, storing, and analyzing data differentiated by high volume, velocity, and variety from diverse sources, including social media, devices, log files, video, text, images, and GPS (Watson, 2014). Employing BDA to enhance comprehension of customer behavior, market trends, operational performance, and risk management allows companies to improve their decision-making processes. Effective BDA requires various elements, including a clear business requirement, a dedicated sponsor, alignment of IT and business strategy, a culture of data-driven decision-making, a strong data architecture, appropriate analytical tools, and proficient staff in analytics. Additionally, numerous studies have demonstrated the substantial potential benefits of BDA (Watson, 2014). Structured, semi-structured, and unstructured data sources may provide significant insights and value and are present within big data ecosystems. While this has generated fresh opportunities for political reform, it may also provide obstacles in the management and organization of these huge amounts (Chatfield et al., 2015).

3.1.2 Data Warehouses

A data warehouse is a collection of combined current and historical data, organized in a simplified style for convenient end-user access (Ang & Teo, 2000). Large amounts of organized and semi-structured data from many sources can be effectively stored, managed, and analyzed using a data warehouse (Borra, 2024). A data warehouse, or a smaller-scale data mart, is a precisely organized collection of data designed to ease decision-making (Wixom et al., 2001). A data analysis and reporting system, suitable for structured data. It provides a clear and uniform structure, allowing businesses to extract useful knowledge effectively. Big data comprises several types of information, including structured, semi-structured, and unstructured data, stored within the Data Warehouse (DW) (Dick, 2022). Examples are Snowflake and Amazon Redshift. A wide range of commercial goods and services are currently readily available with all major database management system developers providing offers in these areas (Chaudhuri & Dayal, 1997). A plethora of data warehousing methods and tools exists to accommodate the expanding market (Sen & Sinha, 2005). Information integration constitutes an essential part of a Data Warehouse. When data transitions from the application-oriented operational setting to the Data Warehouse, potential inconsistencies and repetitions must be addressed to ensure the warehouse provides an integrated and reconciled perspective of the organization's data (Calvanese et al., 2001). To ensure and uphold data quality, data warehouses must be constantly updated. The unified technique merges the extraction, transformation, and loading (ETL) processes into a single step to transfer data from sources to the data warehouse (Yang et al., 2011).

3.1.3 Data Lakes

The volume and variety of data have significantly exceeded the capabilities of manual analysis and, in certain instances, exceeded the limits of traditional databases (Provost & Fawcett, 2013). A data lake is a centralized repository that allows storing various types of data (structured, unstructured, and semi-structured) at scale. Organizations seek to gain new insights and derive value from data by transferring it to a centrally managed scalable storage repository called data lake (Mathis, 2017). The data lake may assist the removal of organizational barriers and system complexities (Mathis, 2017). In contrast to conventional data warehouses, data lakes do not require the processing or structuring of data prior to storage, making them optimal for advanced analytics. Data lakes are becoming popular for the management of Big Data and data analytics (Hai et al., 2023). According to Fang (2015) the idea of a data lake is gaining popularity as a method to structure and develop advanced systems to address modern big data concerns. Despite growing interest from academics and industry, considerable misunderstanding exists around the definition, functions, and accessible technologies of data lakes (Hai et al., 2023). The existing research on data lakes is unclear and insufficient, and the recommended implementation methods fail to address all aspects of data lakes or offer a thorough design and implementation plan (Giebler et al., 2019). A broad and coherent understanding of the challenges and solutions associated with data lakes remains unclear (Hai et al., 2023).

3.1.4 Data Visualization

Data visualization is a vital part of the analytics process. Effective visualizations allow a scientist to comprehend their data and convey their ideas to others (Waskom, 2021). Data visualization is an essential component of BDA, offering a visual representation of data which makes complex data understandable and actionable. It connects raw data to decision-making by allowing stakeholders to rapidly and effectively evaluate analytical results. Data Visualization techniques are employed to generate tables and diagrams for BDA (Yaqoob et al., 2016). Data visualization represents a significant advancement in BDA, as companies progressively implement various business intelligence and data visualization tools to improve their management and utilization of big data. Visualization has demonstrated efficiency in both transmitting critical information inside extensive datasets and facilitating complex analyses (Keim et al., 2013; Aseeri & Kang, 2020). Most visualization techniques are predicated on the notion that the data to be depicted has no element of uncertainty. However, this is rarely the situation. In contrast to big data storage and analysis, big data visualization is the final stage in improving decision-making, since it provides an engaging and interactive method for easing this process (Aseeri & Kang, 2020). Recently, the visualization industry has addressed the difficulty of integrating uncertainty signs into visual representations (Brodlie et al., 2012). Data visualization includes the graphical or pictorial representation of data, improving understanding of the information. It clarifies facts and establishes directions of action. It will enhance any discipline that involves new techniques for conveying vast, intricate information (Sadiku et al., 2016). Popular data visualization tools are Tableau, Power BI, Qlik View, D3.js and Google Data Studio.

3.2 BDA in the IT sector

In the IT sector, big data is becoming essential day by day. IT firms employ BDA to track network traffic, identify obstacles, and enhance bandwidth allocation. Apart from these examples, the evolution of BDA shows a series of advances that have helped organizations' transition from static data reporting to dynamic, predictive insights. Transition from batch processing to real-time analytics in domains such as fraud detection and system governance. Creation of hybrid IT ecosystems that connect physical IT systems with cloud solutions to enhance cost efficiency and performance. Utilizing BDA tools, firms can apply big data to enhance customer loyalty, minimize supply chain risks, provide strategic insights, and conduct reliable market research to inform important decisions (Božič & Dimovski, 2019; Niu et al., 2021).

BDA is commonly used across multiple sectors within the IT industry, offering actionable insights that increase performance, security, and user experience. For example, BDA has been used as IT artefacts of the healthcare industry. Healthcare providers utilize big data to identify high-risk patients, predict patient outcomes, and develop customized treatment plans. Health insurance companies are using big data to save expenses and enhance treatment accessibility. Approximately 12.4% of Americans are aged 65 and over, and the United States government is spending over $2 trillion on healthcare due to ongoing illnesses (Tan et al., 2015). Healthcare organizations may now analyze the huge volume, diversity, and speed of data across various healthcare networks through BDA, which has emerged from business intelligence and decision support systems. This enables them to facilitate the use of evidence in making decisions and taking action (Wang et al., 2018). In the retail sector, Walmart improved its capacity to convert massive unstructured data into valuable insights by moving from a trial 10-node Hadoop cluster to a 250-node Hadoop cluster. This allowed the integration of data from ten different sites worldwide into a singular huge data ecosystem (Liu et al., 2018).

Modern business processes are influenced by a significant transformation (Unhelkar & Arntzen, 2020). To guide business research efforts in the big data era, information systems must lead in understanding and assessing the impacts of both technology and management (Baesens et al., 2016). The achievement of business IT alignment demands a shift in mentality regarding the acceptance of BDA and its impact on the organization (Hirschlein et al., 2022). Understanding software engineering technologies alongside numerous new innovations designed to facilitate big data is equally significant (Hassani &

Gahnouchi, 2017). Leading organizations can utilize Big Data across many industries to analyze its impact on the main operational models of those organizations (Roden et al., 2017).

3.3 Types of Analytics in BDA

BDA includes several techniques for analysis that assist businesses in processing and interpreting extensive data sets for improved decision-making. Predictive, descriptive and prescriptive analytics are crucial in comprehending past patterns and influencing future decisions. This part addresses these three analytical methods, their significance in the IT sector, and their influence on quality decision-making.

3.3.1 Predictive Analytics

The process of analyzing present and past data to forecast future results. Predictive analytics predicts potential future outcomes by addressing the question, 'What will happen?' It utilizes machine learning techniques, including classification and regression models (Wissuchek & Zschech, 2024). The long-term impact of predictive analytics, given the continuously expanding sources of big data, remains mostly untapped (Janke et al., 2015).

Academics recognize the significance of big data and predictive analytics (BDPA) in realizing economic benefits and improving firm performance (Gunasekaran et al., 2016). Predictive analysis is an essential element of BDA, combining statistical methods, machine learning algorithms, and data modeling to predict future events based on historical and present data. It allows firms to forecast trends, recognize possible threats, and make early choices that enhance company success. Big Data and Predictive Analytics (BDPA) is a complete name covering methods designed to manage large datasets defined by high volume, velocity, and variety (Duan & Xiong, 2015; Wang, Gunasekaran, Ngai, & Papadopoulos, 2016; Zhou, Chawla, Jin, & Williams, 2014 cited in Gunasekaran et al., 2016). Companies are not only storing large datasets, but they are also increasingly enhancing their big data management technologies to facilitate data integration and analysis for improved informational outputs (Aseeri & Kang, 2020). IT organizations utilize predictive models to examine user behavior, allowing targeted marketing efforts, custom recommendations, and increased user experiences. Predictive analysis functions as an efficient tool for IT firms, allowing them to spot industry trends, reduce risks, and support innovation efficiently. The addition of it into BDA frameworks has transformed decision-making processes, providing it an essential asset in the data- driven environment. Predictive analytics includes empirical techniques, including statistical methods, that provide data projections and evaluate predictive accuracy. Predictive analytics not only aid in the development of accurately applicable models, but they also significantly contribute to theory construction and validation when used with explanatory modelling (Shmueli & Koppius, 2011).

3.3.2 Descriptive Analytics

Descriptive analytics concentrates on the past and present by answering questions such as 'What is happening?' or 'What happened?' employing conventional Business Intelligence methods such as Online Analytical Processing (OLAP) or data mining (Wissuchek & Zschech, 2024). Descriptive analytics is the most fundamental kind of BDA, encompassing the summary and characterization of knowledge patterns through basic statistical techniques, including mean, median, mode, standard deviation, variance, and frequency analysis of specific instances within big data streams (Rehman et al., 2016; Sivarajah et al., 2017). Descriptive analytics use statistical methods, visualization tools, and data aggregation approaches to provide companies with a comprehensive understanding of past events, thus facilitating informed business decisions.

Descriptive analytics examines data to explain the status of a business scenario, identifying developments, patterns, and discrepancies through the generation of standard reports, ad hoc reports, and alerts (Joseph & Johnson, 2013; Sivarajah et al., 2017). Descriptive analytics are commonly used in IT companies to assess system performance, monitor client routines, and evaluate operational effectiveness. Descriptive analytics techniques are utilized to identify existing patterns inside companies, understanding past events will enable individuals and organizations to predict forthcoming occurrences (Dick, 2022). Organizations employ descriptive analytics for analyzing server logs, network traffic, and system performance statistical analysis. For example, IT firms monitor downtime events through historical data to pinpoint frequent issues and enhance system reliability. Watson (2014) claims that descriptive analytics, including reporting, dashboards, scorecards, and data visualization, have been extensively utilized for an extended period and represent the fundamental applications of conventional business intelligence. Currently, a trend in descriptive analytics is to make use of insights from predictive analytics, including predictions of future revenues, into dashboards and scorecards (Sivarajah et al., 2017). A variety of tools enable descriptive analytics within the IT sector, including SQL-based databases (e.g., MySQL, PostgreSQL) facilitate the storage and recovery of organized historical data. Tableau, Power BI, and Google Data Studio convert raw data into understandable reports and dashboards. Apache Hadoop and Spark facilitate extensive data gathering and processing. By effectively employing descriptive analytics, IT businesses can encourage a data-driven culture, thereby improving efficiency and strategic planning.

3.3.3 Prescriptive Analytics

While, descriptive analytics offers insights into past events, prescriptive analytics develops by suggesting particular actions derived from data research. Prescriptive analytics aims to determine best decisions, recommendations, or actions by answering the question, “What should be done?” (Delen & Ram, 2018; Wissuchek & Zschech, 2024). Prescriptive analytics utilizes machine learning, artificial intelligence (AI), and optimization models to enable companies to actively enhance outcomes. This form of analytics is conducted to identify the causal relationships between analytic outcomes and business process optimization strategies (Sivarajah et al., 2017). Prescriptive analytics is essential for informing IT decision-makers by providing actionable insights based on prediction models. Consequently, for prescriptive analytics, firms enhance their business process models informed by the insights obtained by predictive analytic models (Sivarajah et al., 2017). Prescriptive analytics involves optimization and randomized testing to evaluate how firms improve service levels while reducing costs (Joseph & Johnson, 2013; Sivarajah et al., 2017). A lot of IT companies employ prescriptive analytics to predict system failures prior to their occurrence. For ex: Cloud service companies such as AWS and Microsoft Azure employ AI-driven prescriptive analytics to advise hardware upgrades prior to major failures. Prescriptive solutions assist business analysts in decision-making by identifying activities and evaluating their effects on company objectives, needs, and constraints (Sivarajah et al., 2017). The complete potential of predictive analytics can only be realized when integrated with prescriptive analytics, which proactively improves decision-making processes (Wissuchek & Zschech, 2024).

To conclude, Prescriptive analytics develops informed decisions derived from predictive analytics outcomes, taking into account the optimal timing for implementing activities prior to the projected event, however, descriptive analytics can be employed post-event to examine its fundamental origins and effects, working across many time frames for reactive or long-term solutions (Wissuchek & Zschech, 2024).

3.4 Main tools and technologies that IT industry uses in BDA[1]

The IT sector significantly depends on BDA to handle and analyze extensive volumes of structured, semi-structured, and unstructured data for decision-making and efficiency in operation. The adoption of new technology for big data has enhanced performance, promoted innovation in business models, and supported service and decision-making processes (Carasso, 2012; Yaqoob et al., 2016). The review conducted revealed multiple methods and technologies developed for analysing substantial volumes of data produced by various applications (Zomaya & Sakr, 2017; Ikegwu et al., 2022). The tools and technologies employed in BDA can be divided according to their functions, including data storage, processing, machine learning, and visualization.

[1] There are many tools (such as AI based tools, Rstudio, etc) that are widely used by the industry but are not listed here as these were not discussed in the identified literature.

3.4.1 Data Storage and Management Tools

3.4.1.1 Hadoop Distributed File System (HDFS)

Hadoop is an open-source, distributed processing framework designed for the business's processing and storage of large datasets within clustered systems, using simple programming that includes the Hadoop Distributed File System (HDFS) and MapReduce computational framework (Oliverio, 2018; Ikegwu et al., 2022). Apache Hadoop facilitates the processing of large amounts of data (Yaqoob et al., 2016). Hadoop allows the study of complex data, developed in Java that supports parallel and distributed data processing and ensures reliable data storage (Sogodekar et al., 2016). HDFS simplifies the storage and distribution of data over numerous nodes, hence ensuring reliability and scalability. Furthermore, it is friendly, reliable, and economically delivers the functionality and attributes to accommodate several nodes, up to 100, in a cluster. HDFS possesses the capability to store organized, semi-structured, and unstructured data, supporting files of one terabyte or greater (Ikegwu et al., 2022).

3.4.1.2 MapReduce

We can understand MapReduce as a statistical framework for managing extensive datasets through the distribution of tasks among numerous servers. The programming framework was developed by Google for the processing of large datasets produced in applications such as search engine optimization (SEO), internet services, and advertising improvements (Sagiroglu & Sinanc, 2013; Ikegwu et al., 2022). It is the most popular and currently used method in the analysis of big data. MapReduce is a fundamental concept in BDA and plays a crucial role in the execution of data mining operations (Ibtisum et al., 2023). It accepts a <input, value> pair as input and returns an output in the form of a different <input, value> pair (Srivastava & Maurya, 2022). MapReduce is renowned for its intrinsic simplicity and scalability, making it exceptionally suitable for the effective processing of large datasets over huge clusters of conventional hardware (Ibtisum et al., 2023).

3.4.1.3 Amazon Redshift

Amazon Redshift is a rapid, fully managed, petabyte-scale data warehouse solution that facilitates the efficient and cost-effective analysis of substantial data volumes using existing business intelligence tools (Gupta et al., 2015). It is a cloud-based data warehouse suited for online analytical processing (OLAP) applications (Worlikar et al., 2021). Amazon RedShift manages the establishment and operation of a cloud-based data warehouse, upon building the data warehouse, data can be retrieved using SQL queries and analytical applications (Watson, 2014). Amazon Redshift is five times more cost-effective and three times faster than conventional data warehouses, and it is closely connected with Amazon S3 (Kulkarni, 2023). Netflix use Amazon Redshift to analyze consumer data, improve content recommendations, and conduct real-time analytics, which increases user engagement and retention.

3.4.1.4 MongoDB

NoSQL databases are designed for effectively handling unstructured and semi-structured data at scale. MongoDB is a document-oriented NoSQL database that supports flexible schema configuration. MongoDB can handle data of any structure without incurring costly data warehouse loads, regardless of its rate of changes, thus we may implement additional functionality cost-effectively without requiring a redesign of the database (Abbes & Gargouri, 2016). MongoDB provides a tree-like structure in documents by integrating references to "parent" nodes into child nodes to represent complex hierarchical data relationships (Abbes & Gargouri, 2017). MongoDB lacks support for elements typical of relational databases, such as Transactions or JOIN operations; however, it offers a more adaptable and easily manipulable structure with JSON as the data format (Heripracoyo & Kurniawan, 2016).

3.4.2 Data Processing Tool

3.4.2.1 Apache Spark

Apache Spark is a powerful platform for processing large volumes of data, known for its capacity to effortlessly integrate hybrid frameworks (Ibtisum et al., 2023). Apache Spark is a strong platform for distributed data processing and machine learning. Apache Spark is an in-memory computing platform that significantly improves Big Data processing relative to conventional batch-processing tools such as Hadoop's MapReduce. It supports multiple programming languages like Python, Java and Scala. Apache Spark also integrates with Hadoop, Kafka, and cloud computing systems. It also allows both batch and streaming data processing. The use of comprehensive optimization in Spark significantly improves its computing efficiency by enabling the prior analysis of a whole set of tasks (Ibtisum et al., 2023). Spark is a cutting-edge framework for high-performance parallel computing, designed specifically to effectively manage repetitive computational processes that continuously execute operations on the same data, such as supervised machine learning algorithms (Reyes-Ortiz et al., 2015). As we all know that Uber is a prime example that utilizes Apache Spark to analyse real-time travel data, facilitating surge pricing predictions, route optimization, and avoiding fraud. Spark facilitates immediate decision-making with millions of daily ride requests, decreasing delays and improving user experience.

3.4.3 Machine Learning Tools

Machine learning (ML) is essential in Big Data for understanding patterns, trends, and potential outcomes derived from past data. A wide range of machine learning algorithms have been developed and implemented in several programming languages over the last two decades (Hao & Ho, 2019).

3.4.3.1 TensorFlow

TensorFlow is an open-source framework developed by Google for scientific and numerical computation employing data flow graphs, which represent TensorFlow's execution model. TensorFlow is a machine learning platform designed for large-scale tasks in diverse settings (Abadi, 2016). The data flow graphs utilized in TensorFlow assist machine learning specialists in executing advanced and extensive training on their data to create deep learning and predictive analytics models, hence facilitating quality decision-making (Zaccone & Karim, 2018). TensorFlow use dataflow graphs to illustrate computation, shared state, and the operations that modify that state. It allocates the nodes of a dataflow graph over numerous machines in a cluster and within a machine across various computational devices, including multicore CPUs, general-purpose GPUs, and specialised ASICs referred to as Tensor Processing Units (TPUs) (Abadi et al., 2016). In November 2015, Google launched TensorFlow, an open-source deep learning software library for the definition, training, and deployment of machine learning models (Goldsborough, 2016). It contains an extensive and flexible ecosystem of tools, libraries, and community resources that enables academics to advance the field of machine learning and allows developers to effortlessly build and deploy machine learning-powered applications (Developers, 2022).

3.4.3.2 Scikit-learn

Scikit-lean is a Python-based machine learning library that provides algorithms for classification, regression, and clustering. The Scikit-learn library is quite beneficial for data modeling. It possesses effective tools for machine learning and graphical modeling, including classification and regression (Rahul et al., 2020). Scikit-learn streamlines machine learning model training and evaluation. Scikit-learn utilizes this extensive ecosystem to deliver cutting-edge implementations for multiple established machine learning algorithms, while maintaining a user-friendly interface that is tightly integrated with the Python language (Pedregosa et al., 2011). It benefits the widely used Python language, which is largely adopted in scientific disciplines and supported by an experienced community of contributors (Varoquaux et al., 2015). Airbnb uses Scikit-learn to forecast booking demand and dynamically modify pricing in response to consumer demand, the seasons, and competitive factors.

3.4.4 Data Visualisation Tools

We have already discussed above how data visualisation is important in the field of BDA and quality decision making. Data visualization is a key component of BDA, as it simplifies the conversion of raw data into significant insights via interactive dashboards and reports. Below are the tools we will discuss that facilitate data visualisation.

3.4.4.1 Tableau

Tableau is a robust visual analytics tool helping firms to develop interactive dashboards without requiring significant programming skills. Tableau is a well-known software tool for data visualization, extensively utilized in both corporate and educational settings, the tool is effectively designed and user-friendly (Schiller, 2017). It converts huge amounts of complex data into visual and graphical formats, facilitating the rapid and effortless identification of insights and patterns (Schiller, 2017). Tableau facilitates real-time connections to databases. It integrates with multiples other sources like Hadoop, SQL, and cloud-based databases. According to Zhang et al. (2020) Tableau is a popular business analytics tool recognized for its simple user interface and fast generation of interactive visualizations.

3.4.4.2 Microsoft Power BI

Microsoft Power BI is a business intelligence application that enables companies to generate customisable reports and real-time dashboards. Power BI is an online platform that allows data searching, transformation, visualization, and sharing of generated reports and dashboards to users within the same or various departments or organizations, as well as to the general public (Krishnan, 2017). Microsoft Power BI is a comprehensive set of business intelligence and analytics tools designed for data analysis, insight sharing, and quick response through interactive data visualizations via dashboards accessible on many devices, including applications, desktops, and mobiles (Bhargava et al., 2018). The fundamental operations of Microsoft Power BI are as follows: 1) Collect the data from the necessary data source. 2) Analyze the data using organizational connectors and gateways. 3) Construct the report employing various visuals and filters. 4) Publish the report to the web via Power BI Desktop 5) Edit the report if modifications are necessary and enable sharing via the web publishing option to create an embed URL. 6) Retrieve report data from several Microsoft applications, including Power Apps and Mobile Power BI. 7) Update the dashboard by refreshing the data through various organizational gates (Bhargava et al., 2018).

The IT sector employs a variety of BDA tools and technologies to manage, process, analyze, and visualize data. Hadoop and Spark are extensively utilized for Big Data processing, whilst data warehouses such as Amazon Redshift efficiently store structured data. Machine learning frameworks like TensorFlow and Scikit-learn offer predictive characteristics, while visualization technologies such as Tableau and Power BI assist companies in analyzing complex data.

Implementing these tools enables firms to attain a competitive advantage, improve decision-making, and discover new opportunities in the data-driven economy.

Table 3 below presents these technologies according to their classification:

Data storage and Management tools	**Data processing tools**	**Machine Learning tools**	**Data Visualisation tools**
Apache Hadoop	Apache Spark	Tensorflow	Tableau
MapReduce	Apache Flink	Scikit Learn	Microsoft Power BI
Amazon Redshift	Apache kafka	Pytorch	Matplotlib & Seaborn
MongoDB	Apache Storm		

Table 3: Tools and Technologies

This chapter helps to understand the answer of the first research question. The next chapter (Findings and Discussion) focuses on the second question and discusses the factors impacting BDA applications. Limitations and Implications for practice.

Chapter 4: Findings and Discussion

4.1 Factors Impacting BDA Application

The application of big data is a process by which firms discover innovative methods to improve efficiency and decrease risk, while better addressing client needs (Baig et al., 2019). The adoption and effective implementation of BDA rely on a variety of related variables. These requirements govern the ability of enterprises to extract meaningful information and improve decision- making abilities. This section examines three themes affecting BDA applications: Technological Factors, Non-Technological Factors (Organizational Factors and Human Factors) and Contextual Factors.

4.1.1 Technological Factors

An effective approach to understand the status of big data technologies is to compare it to the commercial adoption of internet technologies (Provost & Fawcett, 2013). Big data has greatly impacted numerous sectors in the modern world, including transportation, marketing, healthcare, manufacturing, and retail. Beyond the government sector, the financial sector stands to benefit significantly from the utilization of big data; this technology not only enables financial institutions to optimize data value but also assists them in achieving competitive advantages, reducing costs, transforming challenges into opportunities, and reducing risks in real time (Dick, 2022). Technological issues significantly influence the uptake and efficacy of BDA. For BDA to be implemented successfully in the IT industry, technological factors are essential. These include the infrastructure, design, systems, and tools needed to gather, process, analyze, and present vast amounts of data. Challenges including data integration, scalability, and quality directly affect the security and accessibility of analytical insights. Companies in all industries are certain that they may create value through analytics by utilizing this wide range of data (Fernando & Engel, 2018). There is a growing necessity to maintain data quality consistency across systems and platforms to enable effective integration of large data sources (An et al., 2016; An et al., 2018). At the corporate level, digital continuity facilitates the execution of a proactive and optimal information governance strategy to guarantee the accuracy, reliability, integrity, and usability of data in via big data integration (An et al., 2018). Furthermore, the success rate of big data projects is disturbingly low, with statistics indicating that 80 to 87 percent fail to yield sustainable outcomes (Escobar et al. 2021; Hotz 2024; Ataei 2024). A commonly seen issue is poor data quality, evident through a lack of uniformity, significant variation, and inconsistencies and incompleteness in the data (Raut et al., 2021, Sivarajah et al., 2017, Alharthi et al., 2017, Dai et al., 2020, Khan et al., 2019, Moktadir et a., 2019, Che et al., 2020, Halaweh et al., 2015; Hirschlein et al., 2022). Poor data governance methods, a lack of standardization, or old systems are frequently the cause of problems with data quality. Meanwhile, there are a lot of contextual and technological difficulties in combining different data sources, such as third-party APIs, IoT sensors, social media feeds, and internal databases. Although the quality of the data is a significant factor (Ataei and Staegemann, 2023; Ataei, 2024). System quality is assessed through system reliability, accessibility, adaptability, integration, response time, and privacy (Aseeri & Kang, 2020). The integration of data from various sources, combining organized, semi-structured, and unstructured formats, continues to provide an important problem. However, the expected benefits of using integration technology comes from the belief that integrating disparate information processing leads to unified business processes, hence enhancing corporate performance. (Lee & Biran, 2002). The quality of data is essential for effective analytics. The flaws of methods of analysis applicable to various datasets are directly linked to the previously noted limitation of BDA tools (Hirschlein et al., 2022). Insufficient, inconsistent, or incorrect information might result in inaccurate insights and poor decision-making.

According to prior research on BDA, the technological challenges associated with BDA can be stated in three key aspects (Hyun et al., 2020). First, BDA transitions to an innovative approach for corporate data management, departing from conventional database platforms (Davenport et al., 2012; Hyun et al., 2020). Secondly, BDA improves companies' market intelligence by accessing information from external sources, enabling them to respond more effectively to market fluctuations. (Roberts and Grover 2012; Hyun et al., 2020). Third, BDA gathers data from several divisions, enabling knowledge-based workers to engage with data scientists to enhance decision-making through more effective data use (Davenport et al. 2012; Wang et al. 2019; Hyun et al., 2020).

In this study, system quality is assessed by enabling big data tasks, including storage, analysis, and visualization, alongside big data security and privacy, which will ultimately enhance decision-making (Aseeri & Kang, 2020). High-quality data are essential for the analysis and utilization of big data, as well as to safeguard the data's value. At present, there is a lack of thorough study and research about quality standards and assessment procedures for big data (Cai & Zhu, 2015). The study indicates that the majority of research efforts focused on operational and assurance costs are incurred, research and development, and the generation of data products (Wang et al., 1995). Ultimately, data are the outcome of a process of production, and the execution of this process significantly influences data reliability (Veregin, 1999). This lack of agreement makes it more difficult to assess and enhance data integrity, which reduces the basis for reliable analytics. To address the issues related to BDA, we should examine both technological and non-technological factors in its application and utilization (Aseeri & Kang, 2020). The technological factors relate to big data analytics tasks (BDATs) including storage, analysis, and visualization, which are crucial for extracting value from BDA and enhancing decision- making processes (Aseeri & Kang, 2020). Although the literature on BDA is fast developing with a focus on technical factors, there is a lack of research empirically founded in information systems theories that integrate social and technological factors to enhance understanding of the influence of these components (Gupta and George, 2016; Aseeri & Kang, 2020).

4.1.2 Non-Technological Factors

For Big Data Analytics (BDA) to be implemented and used in businesses effectively, non-technical aspects are just as important as technological factors. These include human and organizational aspects, which frequently serve as either catalysts or barriers to the adoption of BDA solutions. Even though businesses may have cutting-edge technology, the full potential of BDA may not be realized if there are insufficient organizational support systems or human resources.

4.1.2.1 Organizational Factors

The application of BDA offers several advantages; yet organizations often face various challenges and obstacles that may hinder successful implementation. Despite the substantial advantages of BDA, organizations are encountering difficulties in implementing the full potential of data (Vasilyeva and Richardson, 2022). Organizations seeking to utilize BDA as a method of value creation face a variety of potential issues during the implementation process (Hirschlein et al., 2022).

The potential of big data is endless; but it is limited by the technology, resources, and skills accessible for BDA (Sivarajah et al., 2017). To fully use the potential of BDA, it is essential to address these challenges, encompassing technical, organizational, and cultural dimensions. Although BDA may serve as a valuable intellectual asset, its execution poses considerable challenges (Tardio et al., 2015; Jensen et al., 2021). Modern business executives face challenges such as intense competition, increased client expectations, rising material and labor prices, and reduced product lifecycles (Al-Dmour et al., 2023). It is essential to acknowledge that scalable infrastructure is required to manage the vast volume, velocity, and variety of big data. A strong organizational foundation, consisting scalable infrastructure, supporting culture, and clear leadership, is necessary for the successful implementation of BDA. Constructing and sustaining the infrastructure necessary for BDA can be costly. On the other hand, (Maroufkhani et al., 2020) says that BDA significantly transforms business operations, small and medium-sized organizations (SME's) should carefully consider the adoption of such analytics. The adaption of big data enables organizations and industries to stay ahead of their competitors (Baig et al., 2019). Currently, the quality of decisions taken by the upper management enables the enhancement of the firm's performance (Aseeri & Kang, 2020).

On the other hand, if we talk about manufacturing sector, one of the examples is of a worldwide technology firm established a BDA framework under the direction of its CEO, leading to enhanced supply chain management and a 20% decrease in operational expenses. The manufacturing industry will derive greater benefits from big data-driven analysis with the complementary support of related emerging technologies within the context of Industry 4.0 (C. Li & Chen, 2021. The accessibility of data increases the efficiency of the organization's decision-making process (Aseeri & Kang, 2020). Having a data-driven mindset is an organizational change essential for the implementation of BDA. This involves creating an environment that values data insights, increasing understanding of data, and facilitating data-driven decision-making. Increasing informal proof indicates that companies can leverage BDA to gain competitive advantages (Davenport 2006; Chen et al., 2021). Executive funding and transformational leadership are essential factors because they offer the resources, vision, and change management assistance required for analytics projects to be successful. Furthermore, promoting data literacy throughout the company guarantees that workers at all levels can participate in analytics-driven innovation and interact meaningfully with data. Organizations still have to overcome major challenges to organization such uneven data ownership, confusing governance frameworks, and a lack of internal champions, even if BDA's strategic value is becoming more widely acknowledged. These elements highlight how crucial it is to integrate analytics into the organization's strategy framework and integrate BDA projects with more general business objectives.

4.1.2.2 Human Factors

The third identified feature outlines challenges concerning skills and knowledge at the employee level, particularly in the creation, development, and management of BDA-related expertise (Hirschlein et al., 2022). The success of BDA applications is heavily influenced by human parameters, such as worker competence and readiness for change. Other studies argue that the security of an information system is influenced by IT employees (Betz, 2016; Aseeri & Kang, 2020). Moreover, BD systems, similar to other IT systems, are susceptible to the influence of the IT workers who operate and maintain them (Russom, 2011; Aseeri & Kang, 2020). Information security is essential not just in the field of Information Systems but also for protecting Big Data (Jung, 2017; Aseeri & Kang, 2020). A close collaboration and shared understanding between business and IT professionals is essential to ensure success in BDA projects (Vidgen, 2017; Hirschlein et al., 2022). On the other hand, data integrity enhances decision-making, enables effective decisions, and eventually strengthens organizational planning and control processes (Aseeri & Kang, 2020). Organizations need a targeted talent management strategy to recruit and retain proficient BDA professionals, including data scientists and engineers (Hirschlein et al., 2022). Despite increasing demand, multiple companies encounter an absence of skilled employees possessing experience in machine learning, statistical modeling, and data engineering. Providing training programs and certifications enables firms to enhance internal abilities and address skill shortages. The consequent shortage in data literacy presents further issues (Hirschlein et al., 2022). Recently a technology company allocated resources to enhance the employee's skills, leading to a 15% increase in analytics-driven revenue within one year. We define BDA capability as the organizational proficiency in deploying and managing BDA-related resources, which enables an organization to address the BDA deployment gap (Hirschlein et al., 2022). Therefore, it is essential to establish a comprehensive BDA competency that covers both technical and managerial fields, requiring the development of a centralized educational program (Hirschlein et al., 2022).

4.1.3 Ethical and Privacy Factors

BDA offers numerous benefits for organizations; yet it also poses significant ethical, privacy, and legal concerns. Organizations must consider these elements to maintain compliance with laws, promote ethical data consumption, and maintain public trust. Companies using Big Data have identified individuals for products they were previously unaware of needing, ignored citizens during street servicing, informed friends and family of pregnancies or engagements, and imposed higher charges on customers based on their computer type. These companies generate extensive, complicated data sets and employ new predictions and generalizations (Martin, 2015).

The unfair treatment of specific groups could result from algorithms and data that reflect and reinforce existing biases. Ensuring equal choices involves the implementation of unbiased mechanisms for data collecting and analysis. Hiring algorithms should be monitored to prevent discrimination against applicants based on gender, ethnicity, or other protected characteristics. For big data to operate ethically, data owners—the people whose data is utilized—must, at a minimum, be provided with a transparent consciousness of how their data is employed or sold. Instances of the unethical application of BDA and its resultant effects on reputation and the economy, include the public outrage about Target's alleged pregnancy prediction of a teenager (Duhig, 2013) and the US Federal Trade Commission's inquiry into the US$1 billion acquisition between DoubleClick and Abacus, which raised apprehensions regarding the possibility for enhanced customer behavior tracking capabilities (Someh et al., 2016).

Businesses must maintain transparency on the algorithms they employ and their data utilization practices. This involves holding individuals accountable for the decisions provided by automated systems and explaining the reasoning behind those decisions. This can be achieved through comprehensive documentation and effective stakeholder communication. A fundamental principle of ethics is obtaining individuals' informed consent prior to the collection and use of their data. This ensures individuals retain authority over their personal data and are informed of its intended application. Consent documents must be accessible, brief, and clear. Organizations should, when possible, anonymize or de-identify personal data to safeguard individual privacy. To prevent the re-identification of individuals from the data, it is necessary to delete or hide identifiable information. This can be achieved using tactics such as data masking, anonymization, and aggregation. The ethical considerations focus on discrimination and limit personal agency (Schmidt et al., 2021). The risks to privacy decrease when only the data necessary for a specific purpose is gathered. Organizations should ensure that the data they collect is relevant and suitable for their requirements while avoiding a collection of unnecessary data. To uphold this ideal, data collection techniques must be subject to regular audits and evaluations. Ensuring privacy involves the implementation of robust security measures to protect data from breaches, unauthorized access, and cyberattacks. This includes employee training on data security best practices, encryption, access limits, and regular security audits. BDA presents numerous advantages for companies; nevertheless, it also raises significant concerns regarding privacy, ethics, and legal implications that must be carefully addressed. Organizations must adopt a responsible approach to data usage, ensuring compliance with regulations, protecting personal information, and resolving ethical concerns. By doing so, they may develop the trust of key stakeholders and utilize big data in a manner that honors individual rights while remaining beneficial.

4.2 Decision Making in the IT sector

4.2.1 Importance of Data-Driven Decision-Making

Organizations have recognized the necessity of employing data scientists, educational institutions are rapidly developing data science educational programs, and publications are promoting data science as a desirable—indeed, "attractive"—career option (Provost & Fawcett, 2013). We are in the age of big data, and enterprises are trying to become more data-driven to enhance their decision-making efficiency (Jia et al., 2015). The growing volume of data and the increasing importance of data-driven decision-making in digital business strategies have elevated the concept of a data-driven culture, particularly in encouraging the use of technologies like BDA and artificial intelligence (Anton, Oesterreich, & Teuteberg, 2021; Enholm et al., 2021; Mikalef et al., 2018; Anton et al., 2023). In recent years, an increasing number of firms are thinking about strategies to operate more intelligently, swiftly, and efficiently by utilizing appropriate data to facilitate effective decision-making (Davenport, 2006; Jia et al., 2015). This is commonly referred to as data-driven decision making or evidence-based decision making, which prioritizes conclusions based on data analysis rather than solely on intuition (Provost & Fawcett, 2013; Jia et al., 2015). In this way, A data- driven business orientation is essential for encouraging innovation and competitiveness across all sectors (Gupta & George, 2016; Upadhyay & Kumar, 2020; Anton et al., 2023). Based on this comprehension of data, the idea of data-drivenness can be studied (Fischer et al., 2022). Recently, enhanced capabilities for gathering large-scale, crowd-sourced data have enabled companies to access a wider and more diverse array of data sources, facilitating the discovery of unique behavioural trends (Brynjolfsson et al. 2015; Rhyn & Blohm, 2019). Despite significant advancements in recent years in establishing the technical foundations for processing large-scale crowd data (Chen et al. 2012; Rhyn & Blohm, 2019), the decision-making processes within organizations that leverage these creative abilities to source and analyze data regarding individuals' actual behaviors, opinions, or choices remain poorly recognized (Rhyn & Blohm, 2019). Decision-making refers to the collection of data-processing activities and evaluative patterns by which key actors assess information and select actions to address an organizational issue (e.g., creating a new product based on insights or data on behavior from a crowd) (Rhyn & Blohm, 2019). Researchers at the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST) say that data-driven decision making and the utilization of data for ongoing enhancement are the dominant principles (Mandinach et al., 2006). The integration of BDA in decision-making enables IT organizations to enhance processes, reduce risks, and secure a competitive edge.

In addition to the factors impacting BDA appications discussed above businesses demand scalable storage architectures that can store organized, semi-structured, and unstructured data. Conventional relational databases are unsuitable due to their limits in scalability and real-time processing capabilities. Data warehouses maintain structured and refined data optimised for reporting and analysis, whereas data lakes store raw, semi-structured, and unstructured data. Amazon Redshift, Google BigQuery, and Snowflake are commonly used for their scalability, adaptability, and cost-effectiveness. NoSQL Databases like MongoDB, Cassandra, and Apache HBase are prominent options for handling high-velocity, schema-less data, allowing businesses to store complex and rapidly expanding datasets. Due to the vast amount of data, businesses require efficient data processing systems which allow faster computations. Distributed computing frameworks enable businesses to process extensive datasets simultaneously, lowering computation time and increasing real-time decision-making. Apache Hadoop and MapReduce are generally employed for batch processing of extensive datasets in scattered environments. Apache Spark is an enhanced alternative to Hadoop, delivering real-time data streaming and in-memory computation functionalities. Tensorflow Is an extensively employed for deep learning applications, allowing firms in developing AI-driven predictive models which in turn helps companies for quality decision making. The growing implementation of edge computing in domains such as healthcare and manufacturing facilitate real-time decision-making near data sources.

The accessibility of simple analytics tools allows companies to get insights from data effectively. The capacity to view and analyze complex datasets is essential for decision-makers. Business Intelligence (BI) tools like Tableau and Power BI are the most popular solutions for interactive dashboards, real-time analytics, and business intelligence reporting. Machine Learning Libraries like Scikit-learn, TensorFlow, and PyTorch are frequently employed in predictive analytics, fraud detection, and customer behavior analysis. BiQuery ML is a cloud-based platform that combines machine learning functions with SQL queries, enabling businesses to conduct AI-driven analytics without requiring considerable coding skills. The successful adoption of BDA is motivated by powerful leadership and alignment with organizational objectives. In the absence of leadership commitment and strategic vision, analytics efforts frequently fail to produce significant results. Organizations with appointed Chief Data Officers (CDOs) and/or Chief Technical Officers (CTOs) are more inclined to achieve success in BDA implementation. Data-centric cultures improve strategic decision-making, promoting innovation and operational efficiencies. The absence of definitive ROI evaluation is a common barrier, resulting in hesitation towards BDA investment.

Although BDA provides considerable benefits, the substantial expenditures associated with infrastructure and qualified personnel are challenges for some businesses. Cloud-based solutions such as AWS, Google Cloud, and Microsoft Azure decrease infrastructure expenses. Open-source frameworks like as Hadoop, Spark, and Scikit-learn offer affordable substitutes for private analytics solutions. Organizations that prioritize staff training and upskilling have increased success rates in the implementation of BDA. A significant barrier in the implementation of BDA is a lack of competent staff skilled in data science and machine learning. Significant demand for competence in Python, R, TensorFlow, and Spark. Companies that participate in employee upskilling experience increased adoption rates of BDA technologies. The culture of an organization significantly influences BDA adoption, since employees might reject the shift from conventional decision-making to data-driven techniques. Employees having minimal understanding of technology frequently oppose the implementation of analytics. Thorough training programs increase acceptance rates. Companies possessing powerful data-driven cultures show faster and improved decision-making.

Even though BDA offers considerable advantages for businesses, it also causes ethical and privacy issues that require resolution. These factors affect the adoption, regulation, and reliability of analytics-driven decision-making in the IT sector. The utilization of big data raises ethical concerns, especially around bias, discrimination, and data exploitation. Businesses must make sure that their BDA processes adhere to ethical principles to avoid unforeseen outcomes. Numerous AI and machine learning methods repeat bias from prior data, resulting in unfair decision-making (e.g., biased recruitment models, unfair loan approvals). Black-box AI models in BDA often show a lack of transparency, affecting the justification and auditing of actions. Companies are progressively exploiting user data without clear authorization, which raises ethical concerns. The extensive gathering, storing, and processing of data have rendered data privacy an essential problem. Enterprises must adhere to privacy requirements while guaranteeing data security. Numerous enterprises gather and scrutinize Personally Identifiable Information (PII) without sufficient anonymization. Companies must comply with privacy regulations set by the Government. Growing cyber-attacks on BDA systems risk organizations to financial losses and damage to reputation. The legal structure regarding big data is continually changing, ensuring that enterprises remain informed about compliance requirements. Conflicts emerge around the ownership of data collected by third-party services. AI-generated analytics create unique intellectual property; however legal systems struggle with establishing ownership rights. Specific restrictions limit the transfer of data across national borders (e.g., GDPR's data localization restrictions).

4.3 Limitations

With the thorough methodology used in this dissertation, a few limitations must be highlighted to present a fair assessment of the research's conclusions. These limitations may affect how the results are interpreted and how broadly they can be applied. They are methodological, contextual, and practical in character.

First of all, because this study is grounded in a systematic literature review (SLR), it is inevitably restricted to secondary data and already published studies. Systematic reviews are useful for synthesizing existing information, but they are not able to collect primary data or validate findings in real time. As a result, the study might not accurately reflect the most recent developments in BDA technology or organizational procedures, especially those that were published outside of academic databases or after the deadline for article inclusion.

Secondly, only peer-reviewed journal articles and conference papers in English were eligible to be included in the literature review. This could have led to the exclusion of crucial ideas from non-academic or international contexts where English is not the primary language of publication, as well as useful information from industry studies, white papers, and non-English publications. This could result in some publication and linguistic bias, which would reduce the range of perspectives considered.

Furthermore, the study mainly examined BDA adoption from the perspectives of organizational, technological, contextual, and human factors. Despite being thorough, this complex framework could still oversimplify the complex relationships between various elements in real-life scenarios.

Lastly, the dynamic and quickly changing nature of big data techniques and technology is a significant restriction. As new technologies gain hold, some of the findings given in this study may become obsolete due to the rapid development of new tools, frameworks, and analytical models. Future trends or new disruptions in the BDA field cannot be considered by the study due to the lack of continuous analysis and real-time data.

In acknowledging these limitations, it is important to point out that they do not affect the research's significance. Instead, they draw attention to ways that future studies can expand on the existing findings through cross-sectoral comparisons, empirical research, or the incorporation of new issues like sustainability, data governance, and ethical AI into big data projects.

4.4 Implications for Practice

The findings of the research have important implications for several practical applications stakeholders engaged in the strategic use and deployment of BDA in the IT industry. This section provides important considerations for IT managers, developers, and policymakers looking to improve data-driven decision-making in their businesses by converting the insights from the systematic literature study into helpful guidance.

4.4.1 Implications for IT Managers

The implementation of BDA is led by IT managers, who oversee integrating technological capabilities with strategic business goals. The study emphasizes how crucial it is to use cloud-native platforms like Azure Synapse or Google BigQuery, as well as scalable and compatible BDA tools like Hadoop and Spark, that support the data kinds and volumes of the company. Data management systems that ensure data quality, integrity, and consistency across departments must be given top priority by managers. To enable smooth data integration from internal and external sources.

4.4.2 Implications for Developers and Data Professionals

The research's practical implications for developers, data engineers, and analysts focuses on technical skill development, tool selection, and system architecture. The study highlights how important it is for experts to keep up with new programming frameworks and BDA tools that support batch and real-time analytics.

Developers also need to be prepared to deal with the difficulties of data integration, especially in circumstances where semi-structured or unstructured data is present. This includes becoming proficient with technologies for secure API integrations, schema matching, and data cleaning. Human elements like speaking with company stakeholders and working with subject matter experts are equally important.

4.4.3 Implications for Policy Makers

This study emphasizes the significance of creating uniform regulations for data security, connectivity, and quality for regulators and policymakers. Policies that address data autonomy, ethical data use, and cross-border data flows are urgently needed as BDA applications expand into fields like cloud computing and the Internet of Things. Clearer guidelines and incentives for implementing BDA technology should be made available by governmental and regulatory organizations, particularly to SMEs that might not have the financial or technical resources to carry out large deployments.

Additionally, funding talent development programs and digital literacy programs at the national level could help in closing the skills gap in the analytics sector. Furthermore, policymakers must encourage public-private partnerships that share best practices, create industry standards, and make high-quality, anonymous information more accessible for research and innovation.

Chapter 5: Conclusion

This dissertation focused on the fundamental significance of BDA in enhancing quality decision-making in the IT sector. The study conducted a systematic literature review to identify the main tools and technologies utilized in BDA and analyze the factors affecting its adoption. The results indicate that the adoption of BDA is impacted by technological, non-technological factors and ethical and privacy factors with storage and processing technologies, machine learning models, and visualization tools being essential for facilitating data- driven decision-making.

This research identified from a technological standpoint that hybrid storage solutions, which include data warehouses and data lakes, enhance cost efficiency while maintaining data accessibility for various analytical tasks. Companies must adopt automated data pipeline technologies, to guarantee data consistency and optimize data flexibility. Businesses need to focus on real-time data processing frameworks such as Apache Spark for managing time-sensitive BDA. AI-powered analytical workflows utilizing TensorFlow, and Scikit Learn can improve prediction capacities and automate process of decision-making. Self-service BI tools must be prioritized to help non-technical employees to participate in decision-making processes. Analytics dashboards driven by AI can improve decision-making quality and operational efficiency.

In addition to technology, non-technological factors such as leadership commitment, budget limitations, workforce competencies, and organizational culture were identified as equally important in the adoption of BDA. Organizations should incorporate Key Performance Indicators (KPIs) and Return on Investment (ROI) evaluation structures into their analytical process. Multidisciplinary collaboration among IT, business, and data teams is crucial for the success of BDA programs. A hybrid cloud strategy, which combines on-site and cloud technologies, can enhance cost- effectiveness and data security. Investing in open-source tools can substantially reduce software license costs while keeping strong analytical skills. Partnerships with universities and certification programs help mitigate workforce shortages. Gamification and practical case studies help expedite employee adoption of BDA tools. Management must promote a culture of data literacy and ongoing learning.

The research additionally studied the ethical and privacy factors which are gaining importance as companies use extensive data sources. Establishing an ethics committee within firms helps ensure monitoring of data-driven decision-making processes. Anonymization and encryption methods, such as differential privacy, must be implemented when dealing with sensitive data. Businesses must adopt rigorous security protocols, including multi-factor authentication (MFA) and blockchain technology, to ensure secure data exchanges. Failure to follow with privacy laws may lead to lawsuits and monetary sanctions, making regulatory compliance essential. Legal teams must collaborate closely with data scientists to guarantee that AI models adhere to regulatory requirements. Data governance policies should clearly identify ownership, retention periods, and disposal protocols to prevent legal conflicts. Companies should implement Privacy-by-Design rules to include legal compliance within data analytics methods.

This research highlights the importance for firms to adopt a comprehensive strategy to BDA implementation, integrating advanced technology with powerful organizational strategies and ethical issues. Despite significant progress in big data tools and technologies, companies must emphasize data literacy, governance, and ethical AI utilization to optimize the advantages of analytics-driven decision-making.

5.1 Research Gaps and Future Research Directions

Even though the literature on Big Data Analytics (BDA) is expanding, there are still a number of unresolved issues that need to be addressed. These gaps show areas where theoretical advancement, evidence-based support, and practical application are still lacking. They were found through the systematic review and analysis of academic papers.

The insufficient integration of socio-technical viewpoints in BDA research is one of the most noticeable gaps. A significant portion of the literature still places a strong emphasis on technology tools, platforms, and architectures—often at the expense of organizational and human factors. There is a lack of empirical research based on strong information systems (IS) theories that combine cultural readiness, leadership, and talent development. Multidisciplinary techniques that consider the interactions and effects of technological, organizational, and social aspects on BDA outcomes should be used in future study.

Another underexplored area is the empirical validation of BDA success metrics in different organizational contexts. Although the reviewed literature identifies key factors such as data quality, infrastructure, and human expertise, there remains a lack of consensus on how to measure the success of BDA initiatives. Furthermore, there is little understanding of the dynamic interplay between these factors over time, particularly in longitudinal studies. Investigating the evolution of BDA maturity models across sectors and firm sizes—especially among small and medium-sized enterprises (SMEs)—could provide valuable insights.

Last but not least, there is still an absence of practitioner-oriented and qualitative research. Most of the research, which concentrate on measures related to organizational preparation or tool performance, mainly use quantitative approaches. Nonetheless, qualitative information collected from case studies, interviews, or ethnography may highlight the complicated facts of BDA implementation, such as intra-organizational conflicts, leadership dynamics, and change resistance. Maintaining the relevance and usefulness of BDA research requires bridging the gap between academic research and practical practitioner experience.

By filling in these research gaps, BDA's theoretical understanding and practical applicability would both increase, empowering stakeholders to make more strategic and informed decisions about the implementation and application of analytics.

References

Abadi, M. (2016). TensorFlow: Learning functions at scale. *In Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming,* 1–1. https://doi.org/10.1145/2951913.2976746

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., . . . Zheng, X. (2016). TensorFlow: a system for large-scale machine learning. *Operating Systems Design and Implementation,* 265–283. https://doi.org/10.5555/3026877.302689

Abbasi, A., Sarker, S., & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems, 17*(2), I–XXXII. https://doi.org/10.17705/1jais.00423

Abbes, H., & Gargouri, F. (2016). Big Data Integration: A MongoDB Database and Modular Ontologies based Approach. *Procedia Computer Science*, 96, 446–455. https://doi.org/10.1016/j.procs.2016.08.099

Al-Dmour, H., Saad, N., Basheer Amin, E., Al-Dmour, R., & Al-Dmour, A. (2023). The influence of the practices of big data analytics applications on bank performance: filed study. *VINE Journal of Information and Knowledge Management Systems, 53*(1), 119-141.

An, X., Bai, X., Chen, H., & Han, W. (2018). Effective big data integration in the development of smart cities in China: a digital Continuity approach. *CONF-IRM,* 8. https://aisel.aisnet.org/confirm2018/8/

Ang, J., & Teo, T. S. (2000). Management issues in data warehousing: insights from the Housing and Development Board. *Decision Support Systems, 29*(1), 11–20. https://doi.org/10.1016/s0167-9236(99)00085-8

Anton, E., Duong, T., Aptyka, M., & Teuteberg, F. (2023). Beyond Digital data and Information Technology: Conceptualizing Data-Driven Culture. *Pacific Asia Journal of the Association for Information Systems, 15*, 1–36. https://doi.org/10.17705/1pais.15301

Duan, Y., Ramanathan, R., & Cao, G. (2019b). A Systematic literature review on the applications of big data analytics – Identifying influential factors and impact. *Americas Conference on Information Systems*. http://dblp.uni-trier.de/db/conf/amcis/amcis2019.html#DuanRC19

Aseeri, M. M., & Kang, K. (2020). Technological and human factors for supporting big data analytics in Saudi Arabian higher education. *Americas Conference on Information Systems*. https://opus.lib.uts.edu.au/bitstream/10453/146300/2/AMCIS%20PAPER%202020.pdf

Ataei, Pouya; Thorpe, Stephen; Regula, Sri; and Staegemann, Daniel, "Why Big Data Projects Fail? A Systematic Literature Review" (2024). *ACIS 2024 Proceedings*. 35. https://aisel.aisnet.org/acis2024/35

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational issues of big data and analytics in networked business. *MIS Quarterly, 40*(4), 807–818. https://doi.org/10.25300/misq/2016/40:4.03

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. *Information Processing & Management, 56*(6), 102095. https://doi.org/10.1016/j.ipm.2019.102095

Bedeley, R., & Nemati, H. (2014). Big data analytics: a key capability for competitive advantage. *Americas Conference on Information Systems*. https://aisel.aisnet.org/amcis2014/Posters/BusinessIntelligence/7

Bhargava, M. G., Kiran, K. T. P. S., & Rao, D. R. (2018). Analysis and Design of Visualization of Educational Institution database using Power BI Tool. *Global Journal of Computer Science and Technology*. https://computerresearch.org/index.php/computer/article/download/1776/1760

Borra, P. (2024). An overview of cloud data warehouses: Amazon Redshift (AWS), Azure Synapse (Azure), and Google BigQuery (GCP). *International Journal of Advanced Research in Computer Science, 15*(3), 23–27. https://doi.org/10.26483/ijarcs.v15i3.7099

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brodlie, K., Allendes Osorio, R., & Lopes, A. (2012). A Review of Uncertainty in Data Visualization. *Springer, London,* 81–109. https://doi.org/10.1007/978-1-4471-2804-5_6

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal, 14*(0), 2. https://doi.org/10.5334/dsj-2015-002

Calvanese, D., De Giacomo, G., Lenzerini, M., Nardi, D., & Rosati, R. (2001). DATA INTEGRATION IN DATA WAREHOUSING. *International Journal of Cooperative Information Systems, 10*(03), 237–271. https://doi.org/10.1142/s0218843001000345

Chatfield, A. T., Reddick, C. G., & Al-Zubaidi, W. H. A. (2015). Capability challenges in transforming government through open and big data: tales of two cities. *International Conference on Information Systems*, 1. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1507&context=icis2015

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record, 26*(1), 65–74. https://doi.org/10.1145/248603.248616

Chen, D. Q., Preston, D. S., & Swink, M. (2021). How big data analytics affects Supply Chain Decision-Making: An Empirical analysis. *Journal of the Association for Information Systems, 22*(5), 1224–1244. https://doi.org/10.17705/1jais.00713

Danielsen, F., Olsen, D. H., & Framnes, V. A. (2021). Toward an understanding of big data analytics and competitive performance. *Scandinavian Journal of Information Systems, 33*(1), 6. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1635&context=sjis

Developers, T. (2022). TensorFlow. *Zenodo*.

Dick, A. N. S. (2022). BIG DATA: DESCRIPTIVE ANALYTICS AND PERFORMANCE OF COMMERCIAL BANKS IN PORT HARCOURT, RIVERS STATE, NIGERIA. *BW Academic Journal,* 21-21.

Duhigg, C. (2013). 24. How companies learn your secrets. *In Columbia University Press eBooks* (pp. 421–444). https://doi.org/10.7312/star16075-025

Fang, H. (2015). Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem.*2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 820–824. https://doi.org/10.1109/cyber.2015.7288049

Fernando, F., & Engel, T. (2018). Big Data and Business Analytic Concepts: A Literature Review. *Americas Conference on Information Systems*. https://aisel.aisnet.org/amcis2018/DataScience/Presentations/8

Fischer, H., Wiener, M., Strahringer, S., Kotlarsky, J., and Bley, K. (2022). From Knowing to Data-Driven Organizations: Review and Conceptual Framework. *ACIS 2022 Proceedings*. 42. https://aisel.aisnet.org/acis2022/42

Garmaki, M., Gharib, R. K., & Boughzala, I. (2023). Big data analytics capability and contribution to firm performance: the mediating effect of organizational learning on firm performance. *Journal* *of Enterprise Information Management, 36*(5), 1161–1184. https://doi.org/10.1108/jeim-06-2021-0247

Giebler, C., Gröger, C., Hoos, E., Schwarz, H., & Mitschang, B. (2019). Leveraging the Data Lake: Current state and challenges. *In Lecture notes in computer science* (pp. 179–188). https://doi.org/10.1007/978-3-030-27520-4_13

Goldsborough, P. (2016). A tour of tensorflow. *arXiv preprint arXiv:1610.01178.* https://doi.org/10.48550/arXiv.1610.0117

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2016). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, 70, 308–317. https://doi.org/10.1016/j.jbusres.2016.08.004

Gupta, A., Agarwal, D., Tan, D., Kulesza, J., Pathak, R., Stefani, S., & Srinivasan, V. (2015). Amazon Redshift and the Case for Simpler Data Warehouses. *In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1917–1923. https://doi.org/10.1145/2723372.2742795

Hai, R., Koutras, C., Quix, C., & Jarke, M. (2023). Data Lakes: A Survey of Functions and Systems. *IEEE Transactions on Knowledge and Data Engineering, 35*(12), 12571–12590. https://doi.org/10.1109/tkde.2023.3270101

Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics, 44*(3), 348-361. https://doi.org/10.3102/1076998619832248

Hassani, A., & Gahnouchi, S. A. (2017). A framework for Business Process Data Management based on Big Data Approach. *Procedia Computer Science*, 121, 740–747. https://doi.org/10.1016/j.procs.2017.11.096

Heripracoyo, S., & Kurniawan, R. (2016). Big Data Analysis with MongoDB for Decision Support System. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 14(3), 1083. https://doi.org/10.12928/telkomnika.v14i2.3115

Hirschlein, N., Meckenstock, J., & Dremel, C. (2022). Towards bridging the gap between BDA challenges and BDA capability: a conceptual synthesis based on a systematic literature review. *Proceedings of the . . . Annual Hawaii International Conference on System*

Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2022.748

Hudson, L. A., & Ozanne, J. L. (1988). Alternative ways of seeking knowledge in consumer research. *Journal of Consumer Research, 14*(4), 508. https://doi.org/10.1086/209132

Hyun, Y., Kamioka, T., Park, J., & Chang, Y. (2020). Why Big Data Analytics Competency for Organizational Agility: A View of IS Resources. *PACIS 2020 Proceedings*, 114. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1113&context=pacis2020

Ibtisum, N. S., Bazgir, N. E., Rahman, N. S. M. A., & Hossain, N. S. M. S. (2023). A comparative analysis of big data processing paradigms: Mapreduce vs. apache spark. *World Journal of Advanced Research and Reviews, 20*(1), 1089–1098. https://doi.org/10.30574/wjarr.2023.20.1.2174

Ikegwu, A. C., Nweke, H. F., Anikwe, C. V., Alo, U. R., & Okonkwo, O. R. (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing, 25*(5), 3343-3387 https://doi.org/10.1007/s10586-022-03568-5

Janke, A. T., Overbeek, D. L., Kocher, K. E., & Levy, P. D. (2015). Exploring the potential of predictive analytics and big data in emergency care. *Annals of Emergency Medicine, 67*(2), 227–236. https://doi.org/10.1016/j.annemergmed.2015.06.024

Jensen, M. H., Nielsen, P. A., & Persson, J. S. (2019). MANAGING BIG DATA ANALYTICS PROJECTS: THE CHALLENGES OF REALIZING VALUE. *European Conference on Information Systems*. https://aisel.aisnet.org/ecis2019_rp/47/

Jensen, M. H., Nielsen, P. A., & Persson, J. S. (2021). Improving the impact of big data analytics projects with benefits dependency networks. Scandinavian *Journal of Information Systems.* https://vbn.aau.dk/en/publications/improving-the-impact-of-big-data-analytics-projects-with-benefits

Jia, L., Hall, D. J., & Song, J. (2015). The conceptualization of data-driven decision making capability. *Americas Conference on Information Systems.* https://aisel.aisnet.org/amcis2015/BizAnalytics/GeneralPresentations/4

Jiwat, R., & Zhang, Z. (2022). Adopting big data analytics (BDA) in business-to-business (B2B) organizations – Development of a model of needs. *Journal of Engineering and Technology Management*, 63, 101676. https://doi.org/10.1016/j.jengtecman.2022.101676

Kadadi, A., Agrawal, R., Nyamful, C., & Atiq, R. (2014). Challenges of data integration and interoperability in big data. *2021 IEEE International Conference on Big Data (Big Data)*. https://doi.org/10.1109/bigdata.2014.7004486

Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. *International Journal of Higher Education, 6*(5), 26. https://doi.org/10.5430/ijhe.v6n5p26

Kothari, C. (2004). *Research methodology: Methods and techniques*. https://ndl.ethernet.edu.et/bitstream/123456789/88770/1/2004%20Kothari_%20Research%20Methodology%20Methods%20and%20Techniques.pdf

Krishnan, V. (2017). Research Data Analysis with Power BI. *INFLIBNET Centre*. http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/2116/1/24.pdf

Kulkarni, A. (2023). Amazon Redshift: Performance tuning and optimization. *International Journal of Computer Trends and Technology, 71*(02), 40–44. https://doi.org/10.14445/22312803/ijctt-v71i2p107

Kumar, G. B. (2015). An encyclopedic overview of 'Big data' analytics. *International Journal of Applied Engineering Research, 10*(3). https://www.researchgate.net/publication/280721972

Lee, Y., & Biran, D. (2002). REDEFINING INTEGRATION FOR DATA QUALITY PRACTICE: IMPLEMENTING INTEGRATION TECHNOLOGY FOR DISTRIBUTED BUSINESS STRATEGY. *Americas Conference on Information System*, 180. http://aisel.aisnet.org/amcis2002/180

Li, C., & Chen, Y. (2021). A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal*.

Liu, Y., Han, H., & DeBello, J. (2018). The Challenges of Business Analytics: Successes and Failures. *Proceedings of the . . . Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System*

Sciences. https://doi.org/10.24251/hicss.2018.105

Mandinach, E. B., Honey, M., & Light, D. (2006). A theoretical framework for Data-Driven decision making. *Annual Meeting of the American Educational Research Association, San Francisco, CA,* 39–52. http://cct.edc.org/sites/cct.edc.org/files/publications/DataFrame_AERA06.pdf

Maroufkhani, P., Ismail, W. K. W., & Ghobakhloo, M. (2020). Big data analytics adoption model for small and medium enterprises. *Journal of Science and Technology Policy Management/Journal of Science & Technology Policy Management, 11*(4), 483–513. https://doi.org/10.1108/jstpm-02-2020-0018

Martin, K. E. (2015). Ethical issues in the big data industry. *Social Science Research Network*. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2611767_code883690.pdf?abstractid=2598956&mirid=1

Mathis, C. (2017). Data lakes. *Datenbank-Spektrum, 17*(3), 289-293. https://doi.org/10.1007/s13222-017-0272-7

Nandimath, J., Banerjee, E., Patil, A., Kakade, P., Vaidya, S., & Chaturvedi, D. (2013). Big data analysis using Apache Hadoop. *In 2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*, 700–703. https://doi.org/10.1109/iri.2013.6642536

Nightingale, A. (2009). A guide to systematic literature reviews. *Surgery (Oxford), 27*(9), 381–384. https://doi.org/10.1016/j.mpsur.2009.07.005

Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. (2021). Organizational business intelligence and decision making using big data analytics. *Information Processing & Management, 58*(6), 102725. https://doi.org/10.1016/j.ipm.2021.102725

Okoli, C., & Schabram, K. (2010). A Guide to conducting a Systematic Literature Review of Information Systems Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1954824

Patel, M., & Patel, N. (2019). Exploring Research Methodology: *Review article. Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.3987854

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830. https://doi.org/10.5555/1953048.2078195

Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business Analytics in the context of Big Data: A Roadmap for research. *Communications of the Association for Information Systems*, 37. https://doi.org/10.17705/1cais.03723

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

Rahul, K., Banyal, R. K., Goswami, P., & Kumar, V. (2020). Machine learning algorithms for big data analytics. *In Advances in intelligent systems and computing* (pp. 359–367). https://doi.org/10.1007/978-981-15-6876-3_27

Reyes-Ortiz, J. L., Oneto, L., & Anguita, D. (2015). Big Data Analytics in the Cloud: Spark on Hadoop vs MPI/OpenMP on Beowulf. *Procedia Computer Science*, 53, 121–130. https://doi.org/10.1016/j.procs.2015.07.286

Rhyn, M., & Blohm, I. (2019). Patterns of Data-Driven Decision-Making: How Decision-Makers leverage crowdsourced data. *International Conference on Information Systems*. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1648&context=icis2019

Roden, S., Nucciarelli, A., Li, F., & Graham, G. (2017). Big data and the transformation of operations models: a framework and a new research agenda. *Production Planning & Control, 28*(11–12), 929–944. https://doi.org/10.1080/09537287.2017.1336792

Sadiku, M., Shadare, A. E., Musa, S. M., Akujuobi, C. M., & Perry, R. (2016). Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT), 2*(12), 11-16.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *In 2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE

Schiller, S. (2017). Storytelling with Tableau: A Hands-on Workshop. *MWAIS 2017 Proceedings*. https://aisel.aisnet.org/mwais2017/1/

Schmidt, K., Ullrich, A., & Eigelshoven, F. (2021). FROM EXPLOITATIVE STRUCTURES TOWARDS DATA SUBJECT-INCLUSIVE PERSONAL DATA MARKETS – A SYSTEMATIC LITERATURE REVIEW. *European Conference on Information Systems.* https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1059&context=ecis2021_rp

Sen, A., & Sinha, A. P. (2005). A comparison of data warehousing methodologies. *Communications of the ACM, 48*(3), 79–84. https://doi.org/10.1145/1047671.1047673

Shmueli, N., & Koppius, N. (2011). Predictive Analytics in Information Systems research. *MIS Quarterly, 35*(3), 553. https://doi.org/10.2307/23042796

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research,* 70, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Sogodekar, M., Pandey, S., Tupkari, I., & Manekar, A. (2016). Big data analytics: hadoop and tools. *2016 IEEE Bombay Section Symposium (IBSS),* 1–6. https://doi.org/10.1109/ibss.2016.7940204

Someh, I. A., Breidbach, C. F., Davern, M. J., & Shanks, G. G. (2016). Ethical Implications of Big Data Analytics. *European Conference on Information Systems*. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1016&context=ecis2016_rip

Srivastava, N., & Maurya, P. (2022). Big data analytics in Agriculture using MapReduce. *In Algorithms for intelligent systems* (pp. 407–414). https://doi.org/10.1007/978-981-16-6460-1_31

Tan, C., Sun, L., & Liu, K. (2015). BIG DATA ARCHITECTURE FOR PERVASIVE HEALTHCARE: A LITERATURE REVIEW. *European Conference on Information Systems.* https://doi.org/10.18151/7217494

Unhelkar, B., & Arntzen, A. A. (2020). A Framework for Intelligent Collaborative Enterprise Systems. Concepts, opportunities and challenges. *Scandinavian Journal of Information Systems, 32*(2), 6. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1640&context=sjis

Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn: Machine learning without learning the machinery. *GetMobile: Mobile Computing and Communications, 19*(1), 29-33. https://doi.org/10.1145/2786984.2786995

Vasilyeva, O. & Richardson, A. (2022), Big Data and Data Analytics for Enhanced Decision-Making in the Public Sector. *ICIS 2022 Proceedings.* 14. https://aisel.aisnet.org/icis2022/data_analytics/data_analytics/14

Veregin, H. (1999). Data quality parameters. *Geographical information systems*, 1, 177-189. https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch12.pdf

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics, 34*(2), 77–84. https://doi.org/10.1111/jbl.12010

Wang, R., Storey, V., & Firth, C. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering,* *7*(4), 623– https://doi.org/10.1109/69.404034

640.

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019

Waskom, M. (2021). seaborn: statistical data visualization. *The Journal of Open Source Software, 6*(60), 3021. https://doi.org/10.21105/joss.03021

Watson, H. J. (2014). Tutorial: Big Data Analytics: Concepts, Technologies, and Applications. *Communications of the Association for Information Systems,* 34. https://doi.org/10.17705/1cais.03465

Weerasinghe, K. (2019). Transformation through Big Data Analytics: A Qualitative enquiry in healthcare. *ACIS 2019 Proceedings,* 62. https://aisel.aisnet.org/acis2019/62/

Wissuchek, C., & Zschech, P. (2024). Prescriptive analytics systems revised: a systematic literature review from an information systems perspective. *Information Systems and e-Business Management.* https://doi.org/10.1007/s10257-024-00688-w

Wixom, B. H., & Watson, H. J. (2001). An Empirical Investigation of the Factors Affecting Data Warehousing Success. *MIS Quarterly, 25*(1), 17–41. https://doi.org/10.2307/3250957

Worlikar, S., Arumugam, T., & Patel, H. (2021). Amazon redshift cookbook : Recipes for building modern data warehousing solutions. *Packt Publishing, Limited.*

Yang, W., Qu, Y., & Fairley, R. (2011). Improving the Data Warehouse Architecture Using Design Patterns. *MWAIS 2011 Proceedings,* 17. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1010&context=mwais2011

Zaccone, G., & Karim, R. (2018). Deep learning with TensorFlow : explore neural networks and build intelligent systems with Python. *Packt Publishing Ltd.* https://cds.cern.ch/record/2622338

Appendices

Appendix 1 Literature Matrix

S.N.	Title	Author	Publication	Year	Type	Refernce
1	Toward an understanding of big data analytics and competitive performance	Danielsen, F., Olsen, D. H., & Framnes, V. A.	Association for information systems	2021	Journal	(Danielsen et al., 2021)
2	Big data analytics capability and contribution to firm performance: the mediating effect of organizational learning on firm performance	Garmaki, M., Gharib, R. K., & Boughzala, I.	Emerald Insight	2023	Journal	(Garmaki et al., 2023)
3	Tutorial: Big Data Analytics: Concepts, Technologies, and Applications	Watson, H. J	Association for information systems	2014	Journal	(Watson, 2014)
4	MANAGING BIG DATA ANALYTICS PROJECTS: THE CHALLENGES OF REALIZING VALUE	Jensen, M. H., Nielsen, P. A., & Persson, J. S	Association for information systems	2019	Journal	(Jensen et al., 2019)
5	The Challenges of Business Analytics: Successes and Failures	Liu, Y., Han, H., & DeBello, J.	Proceedings of the 51st Hawaii International Conference on System Sciences	2018	Journal	(Liu et al., 2018)
6	An encyclopedic overview of 'Big data' analytics	Kumar, G. B.	International Journal of Applied Engineering Research	2015	Journal	(Kumar, 2015)
7	Critical analysis of Big Data challenges and analytical methods	Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V.	Science Direct	2016	Journal	(Sivarajah et al., 2017)
8	Big data analytics: a key capability for competitive advantage	Bedeley, R., & Nemati, H. (2014)	Association for information systems	2014	Journal	(Bedeley & Nemati, 2014)

9	Big data adoption: State of the art and research challenges	Baig, M. I., Shuib, L., & Yadegaridehkordi, E.	Science Direct	2019	Journal	(Baig et al., 2019)
10	Adopting big data analytics (BDA) in business-to-business (B2B) organizations – Development of a model of needs	Jiwat, R., & Zhang, Z.	Science Direct	2022	Journal	(Jiwat & Zhang, 2022)
11	Organizational business intelligence and decision making using big data analytics	Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C	Science Direct	2021	Journal	(Niu et al., 2021)
12	Data Lakes: A Survey of Functions and Systems	Hai, R., Koutras, C., Quix, C., & Jarke, M.	IEEE Explore	2023	Journal	(Hai et al., 2023)
13	Managing data lakes in big data era: What's a data lake and why has it become popular in data management ecosystem	Fang, H.	IEEE Explore	2015	Journal	(Fang, 2015)
14	Big data and predictive analytics for supply chain and organizational performance	Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S.	Science Direct	2016	Journal	(Gunasekaran et al., 2016)
15	A Review of Uncertainty in Data Visualization	Brodlie, K., Allendes Osorio, R., & Lopes, A.	Springer, London	2012	Journal	(Brodlie & Lopes, 2012)
16	Predictive Analytics in Information Systems research	Shmueli, N., & Koppius, N.	MIS Quarterly	2011	Journal	(Shmueli & Koppius, 2011)
17	Big data analytics adoption model for small and medium enterprises	Maroufkhani, P., Ismail, W. K. W., & Ghobakhloo, M	Emerald Insight	2020	Journal	(Maroufkhani et al., 2020)
18	Why Big Data Analytics Competency for Organizational Agility: A View of IS Resources	Hyun, Y., Kamioka, T., Park, J., & Chang, Y.	Association for information systems	2020	Journal	(Hyun et al., 2020)
19	Big Data and Data Analytics for Enhanced Decision-Making in the Public Sector	Vasilyeva, O. & Richardson, A.	Association for information systems	2022	Journal	(Vasilyeva & Richardson, 2022)

20	Towards bridging the gap between BDA challenges and BDA capability: a conceptual synthesis based on a systematic literature review	Hirschlein, N., Meckenstock, J., & Dremel, C.	Proceedings of the 51st Hawaii International Conference on System Sciences	2022	Journal	(Hirschlein et al., 2022)
21	The conceptualization of data-driven decision making capability.	Jia, L., Hall, D. J., & Song, J.	Association for information systems	2015	Journal	(Jia et al., 2015)
22	From Knowing to Data-Driven Organizations: Review and Conceptual Framework	Fischer, H., Wiener, M., Strahringer, S., Kotlarsky, J., and Bley, K	Association for information systems	2022	Journal	(Fischer et al., 2022)
23	Big Data Research in Information Systems: Toward an Inclusive Research Agenda	Abbasi, A., Sarker, S., & Chiang, R.	Association for information systems	2016	Journal	(Abbasi et al., 2016)
24	A framework for Business Process Data Management based on Big Data Approach	Hassani, A., & Gahnouchi, S. A.	Science Direct	2017	Journal	(Hassani & Gahnouchi, 2017)
25	Ethical Implications of Big Data Analytics.	Someh, I. A., Breidbach, C. F., Davern, M. J., & Shanks, G. G.	Association for information systems	2016	Journal	(Someh et al., 2016

Appendix 2 Thematic Coding

S.N.	Title	Author	Codes	Themes
1	Toward an understanding of big data analytics and competitive performance	Danielsen, F., Olsen, D. H., & Framnes, V. A.	• Organisational Culture • Cost Constraints • Skills Gap • Leadership Alignment • Quality Decision Making	• Non-Technological factors impacting BDA application • Decision making in the IT sector
2	Big data analytics capability and contribution to firm performance: the mediating effect of organizational learning on firm performance	Garmaki, M., Gharib, R. K., & Boughzala, I.	• Organisational Culture • Cost Constraints • Machine Learning • Data Storage • Data Processing • Quality Decision Making • Data-driven Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Decision making in the IT sector
3	Tutorial: Big Data Analytics: Concepts, Technologies, and Applications	Watson, H. J	• Data Storage • Data Processing • Machine Learning • Data Visualization • Data Security • Data Privacy • Organisational Culture • Cost Constraints • Skills Gap • Legal Compliance • Quality Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector

			• Data-driven Decision Making	
4	MANAGING BIG DATA ANALYTICS PROJECTS: THE CHALLENGES OF REALIZING VALUE	Jensen, M. H., Nielsen, P. A., & Persson, J. S	• Data Storage • Data Processing • Organisational Culture • Skills Gap	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application
5	The Challenges of Business Analytics: Successes and Failures	Liu, Y., Han, H., & DeBello, J.	• Data Visualization • Organisational Culture • Leadership Alignment • Data Security • Quality Decision Making • Data-driven Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector
6	An encyclopedic overview of 'Big data' analytics	Kumar, G. B.	• Data Storage • Data Processing • Machine Learning • Data Visualization • Cost Constraints • Quality Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Decision making in the IT sector
7	Critical analysis of Big Data challenges and analytical methods	Sivarajah, U., Kamal, M. M., Irani,	• Data Storage • Data Processing • Machine Learning	• Technological factors impacting BDA application

		Z., & Weerakkody, V.	• Data Visualization • Cost Constraints • Skills Gap • Data Security • Data Privacy • Legal Compliance • Quality Decision Making	• Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector
8	Big data analytics: a key capability for competitive advantage	Bedeley, R., & Nemati, H. (2014)	• Skills Gap • Organisational Culture	• Non-Technological factors impacting BDA application
9	Big data adoption: State of the art and research challenges	Baig, M. I., Shuib, L., & Yadegaridehkordi, E.	• Data Storage • Data Processing • Machine Learning • Data Visualization • Data Security • Data Privacy • Cost Constraints • Skills Gap • Organisational Culture • Data Security • Data Privacy • Quality Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector
10	Adopting big data analytics (BDA) in business-to-business (B2B) organizations – Development of a model of needs	Jiwat, R., & Zhang, Z.	• Leadership Alignment • Cost Constraints • Skills Gap • Organisational Culture	• Non-Technological factors impacting BDA application • Decision making in the IT sector

			• Quality Decision Making • Data-driven Decision Making	
11	Organizational business intelligence and decision making using big data analytics	Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C	• Machine Learning • Data Visualization • Leadership Alignment • Cost Constraints • Organisational Culture • Data Security • Data Privacy • Quality Decision Making • Data-driven Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector
12	Data Lakes: A Survey of Functions and Systems	Hai, R., Koutras, C., Quix, C., & Jarke, M.	• Data Storage • Data Processing	• Technological factors impacting BDA application
13	Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem	Fang, H.	• Data Storage • Data Processing	• Technological factors impacting BDA application
14	Big data and predictive analytics for supply chain and organizational performance	Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S.	• Data Storage • Data Processing • Data Visualization • Organisational Culture • Leadership Alignment • Quality Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Decision making in the IT sector

15	A Review of Uncertainty in Data Visualization	Brodlie, K., Allendes Osorio, R., & Lopes, A.	• Data Storage • Data Processing • Machine Learning • Data Visualization • Leadership Alignment • Data Security • Ethics • Quality Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application Decision making in the IT sector
16	Predictive Analytics in Information Systems research	Shmueli, N., & Koppius, N.	• Data Storage • Data Processing • Data Security • Data Privacy	• Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application
17	Big data analytics adoption model for small and medium enterprises	Maroufkhani, P., Ismail, W. K. W., & Ghobakhloo, M	• Data Storage • Data Processing • Leadership Alignment • Cost Constraints • Skills Gap • Organisational Culture • Data Security • Data Privacy	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application
18	Why Big Data Analytics Competency for Organizational Agility: A View of IS Resources	Hyun, Y., Kamioka, T., Park, J., & Chang, Y.	• Data Storage • Data Processing • Cost Constraints	• Technological factors impacting BDA application

			• Skills Gap • Organisational Culture • Quality Decision Making	• Non-Technological factors impacting BDA application • Decision making in the IT sector
19	Big Data and Data Analytics for Enhanced Decision-Making in the Public Sector	Vasilyeva, O. & Richardson, A.	• Data Storage • Data Processing • Leadership Alignment • Skills Gap • Organisational Culture • Ethics • Data Security • Data Privacy • Legal Compliance • Quality Decision Making • Data-driven Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector
20	Towards bridging the gap between BDA challenges and BDA capability: a conceptual synthesis based on a systematic literature review	Hirschlein, N., Meckenstock, J., & Dremel, C.	• Data Storage • Data Processing • Data Visualization • Skills Gap • Organisational Culture • Data Security • Data Privacy • Quality Decision Making • Data-driven Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector

21	The conceptualization of data-driven decision making capability.	Jia, L., Hall, D. J., & Song, J.	• Data Storage • Data Processing • Machine Learning • Data Visualization • Data Security • Data Privacy • Quality Decision Making • Data-driven Decision Making	• Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector
22	From Knowing to Data-Driven Organizations: Review and Conceptual Framework	Fischer, H., Wiener, M., Strahringer, S., Kotlarsky, J., and Bley, K	• Data Storage • Data Processing • Leadership Alignment • Organisational Culture • Quality Decision Making • Data-driven Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Decision making in the IT sector
23	Big Data Research in Information Systems: Toward an Inclusive Research Agenda	Abbasi, A., Sarker, S., & Chiang, R.	• Data Storage • Data Processing • Machine Learning • Data Visualization • Data Security • Data Privacy • Ethics • Legal Compliance • Leadership Alignment • Organisational Culture • Cost Constraints • Quality Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Ethical and Privacy factors impacting BDA application • Decision making in the IT sector

			• Data-driven Decision Making	
24	A framework for Business Process Data Management based on Big Data Approach	Hassani, A., & Gahnouchi, S. A.	• Data Storage • Data Processing • Machine Learning • Organisational Culture • Quality Decision Making	• Technological factors impacting BDA application • Non-Technological factors impacting BDA application • Decision making in the IT sector
25	Ethical Implications of Big Data Analytics.	Someh, I. A., Breidbach, C. F., Davern, M. J., & Shanks, G. G.	• Ethics • Data Security • Data Privacy • Legal Compliance	• Ethical and Privacy factors impacting BDA application