Challenges of big data usage for risk management in a South African Bank

By

Boipelo Mosiane (326318)

A Dissertation submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business

Johannesburg, 2023

Supervisor: Prof. Nixon Ochara

ABSTRACT

Risk management is a critical component in the effective operation of corporations, particularly for the purpose of identifying, assessing, and mitigating potential risks associated with running a business. In recent years, the exponential growth of data, along with technological advancements, has opened new opportunities for organizations to enhance their risk management capabilities. Big data is said to be a game-changer that has the potential to completely alter how different industries conduct business. However, there are several difficulties in effectively employing big data in risk management processes. Using a qualitative research approach, this research report highlights the usage of big data in risk management, emphasizing the potential benefits, challenges, and critical considerations for successful adoption.

The research findings revealed that big data can enhance risk identification accuracy, proactive risk mitigation, strategic decision-making, and overall organizational resilience. The challenges hindering the adoption of big data in risk management are addressed, including skill gaps, data quality, technology infrastructure, talent acquisition, and bureaucratic barriers. The study highlights issues preventing widespread integration of big data in the risk community, particularly data trust and collaboration barriers between risk and technology teams. The research report recommends that the bank creates a robust talent acquisition strategy for analytics experts and prioritize retaining them to safeguard data resources. It also suggests fostering a learning-friendly environment for big data topics through accessible certifications and learning programs. Additionally, the research report emphasizes the need for addressing data quality issues in risk management, proposing solutions like RPA to improve data capture processes and enhance data accuracy and trust.

KEYWORDS

Big data| Big data Analytics| Risk Management| Banking| Big data Application

DECLARATION

I, Boipelo Mosiane, declare that this research report is my own work except as indicated in the references and acknowledgments. It is submitted in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Name: Boipelo Mosiane Signature: 

Signed at: Fourways, Johannesburg

On the 29 day of June 2023

DEDICATION

This work is dedicated to my late grandmother Dikeledi Seane, the ultimate Matriarch of my life. Everywhere I go, I hear people talking about how you raised and supported them. I have watched you feeding the less fortunate and sacrificing your last for the next person. I have done this to empower myself so that I can continue your beautiful legacy. Those valuable teachings I will always remember.

ACKNOWLEDGEMENTS

I wish to convey my heartfelt gratitude to the remarkable individuals who have been instrumental in my journey. Firstly, to give God honour and praise for continually preparing a path for my success.

Prof. Nixon Ochara, my supervisor, for the direction and assistance he gave me under very strict deadlines. I appreciate you pushing me to my very limit and never settling for the minimum.

My greatest support system, my sister, Reabetswe Mosiane. While I balanced the demands of work and education, she managed our household with remarkable perseverance and dedication. I appreciate your efforts.

To my son, you're understanding and curiosity during those long hours of studying mean the world to me. Your hugs and kisses have been my motivation, and I hope that my journey paves a path for your future.

My friends, Mmathapelo Molefe and Keletso Lefutswe, have been my constant companions on this journey. They've listened, checked in on me, and offered consistent support. Keletso, your boundless enthusiasm, and encouragement have propelled me forward, thanks for being my biggest cheerleader! I am here because of your affirmations.

Tebogo Maphothoma, thank you for your guidance, and for helping shape my research, you were always just a call away when I was stuck. You indeed ran so that I walk.

I also want to thank Wits Business School, all the lecturers, and the rest of the staff for their support in ensuring we get to this point, despite all the hurdles.

Special thanks to the participants of my study. Your curiosity and willingness to engage have contributed immeasurably to the success of my research. Your involvement has not gone unnoticed, and I am deeply appreciative.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

Acronym	
4IR/ Industry 4.0	4th industrial revolution
BCBS 239	Basel Committee on Banking Supervision- standard 239 (Principles for effective risk data aggregation and risk reporting)
BDA	Big data analytics
IoT	Internet of things
IT	Information Technology

CHAPTER 1. INTRODUCTION

1.1 Statement of purpose

The purpose of this qualitative study is to explore the challenges of using big data in banking risk management.

1.2 Background of the study

The banking industry is without a doubt the anchor that supports economies all over the world. Following the replacement of the barter system with the monetary system, banks now play an important role in a variety of areas such as investment and capital management, as well as project financing (Shakya & Smys, 2021). Banks operate in an ever-evolving environment that is influenced by stringent regulatory requirements, emerging risks, and intense competition (McKinsey, 2015). They have evolved into highly complex fields, encompassing not only traditional banking and trading systems, stocks, and futures markets, but also emerging digital currencies, payment systems, online shopping and a variety of smart systems loosely termed the Internet of things (IoT) (Cheng et al., 2021).

The banking sector is considered to be one of the fastest changing and highly regulated sectors in the world. Because banks are acting as a middleman while providing financial services, they are by virtue of performing these transactions exposing themselves to various kinds of risks (Turgut, 2019). A risk is the potential loss that may be incurred by a bank because of a particular event occurring (Hessami, 2010). Financial risks in the big data era have transcended traditional data, algorithms, and systems, posing more challenges than ever before (Cheng et al., 2021).

There have been reports of large losses caused by the occurrence of the global financial crisis, resulting in the need to strictly regulate the banking sector (Dvorski Lacković et al., 2020). According to the Basel Committee on Banking Supervision standard 239 (BCBS239) released in 2013, one of the most significant lessons learned from the global financial crisis that began in 2007 (and more prevalent in 2008), was that banks' information technology (IT) and data architectures were not adequate to support the management of financial risks (Principles for effective risk data aggregation and risk reporting, 2013). This essentially means that banks did not have the capability to track risky products and could not report what went wrong at a point in time. To mitigate these impacts, the South African Reserve Bank and related bodies dictate that a bank must have a sound risk management function.

Emerging risks such as cybercrime and escalated levels of fraud in banking also pose a major threat and potential for bank losses. Cybercrime has severe impacts on society, ranging from its ability to aid corruption, money laundering, military espionage, and terrorism, to undermining the technological and socio-economic development of any country (Haru, 2021). Moreso, financial fraud and money laundering rely on cutting-edge techniques and methods of concealment, making detection difficult (Cheng et al., 2021). To be at the forefront of these issues, banks will need to intensify their response to the risky environment by leveraging of technologies such as big data.

More data has been created in recent years when compared to history and looking at the current rate, about 2.5 quintillion bytes of data is said to be produced each day (Tang et al., 2020), creating what is called "big data". The concept of "big data" was initially coined by Roger Mogoulas of O'Reilly Media in 2005 (Wahab et al., 2021). This term aimed to describe exceptionally large datasets that would have posed significant challenges for processing using conventional management tools (Wahab et al., 2021). In the age of cloud computing, cyber-physical systems, the Internet of Things (IoT), and big data analysis, each user's click on search engines, social networks, and e-commerce portals generates data that represents an added value to companies (Arena & Pau, 2020). The main question, however, is whether banks are leveraging this "big data" that is being generated daily in their risk management processes.

Regulatory requirements are going to increase in the future, while cybercrime and other emerging risk types seep into the market at a rapid pace. It will be worthwhile for banks to take advantage of the data that is already sitting in their reserves to manage this explosion. Adopting big data technologies such as big data analytics can help banks avoid and pre-empt situations like the financial crisis. Big data analytics is a technology-driven ecosystem that facilitates the extraction of insights from data in a manner that is both interpretable and suitable and can lead to improved risk decision-making and well-informed choices (Elgendy et al., 2022). Using big data can be beneficial not only for regulatory purposes but for managing losses, that in turn will result in improved profit margins.

As seen with other industries, the use of big data can be a powerful tool in banking risk management. However, to understand that power and contribution, it is imperative to understand the challenges preventing the use of big data in the first place. This study investigated the big data landscape relative to banking risk management as well as the perceived reasons for the limited usage of big data in banking risk management. It is envisaged that the finding from the study can help chief risk officers and bank managers gain a better understanding of the impediments when applying big data capabilities as well as factors affecting the full adoption of big data in their risk management processes. The study could also help them come up with better risk management strategies as they reflect on these challenges.

1.3 Context of the study

This study was conducted in the context of a bank in South Africa. Compared to other first-world economies, the country has a robust financial infrastructure with competitive, if not cutting-edge, regulatory, and supervisory regimes (Masindi & Singh, 2022).

As defined by the Banks Act 94 of 1990, a bank's primary business includes soliciting or advertising for, and accepting, deposits from the general public

("Regulated Institutions", n.d.). The South African banking industry currently consists of eighteen (18) registered banks, four (4) mutual banks, five (5) cooperative banks, thirteen (13) foreign bank local branches, and twenty-eight (28) foreign banks with permitted local representative offices (Statista, number-of-banks-south-africa-by-type/, 2023). Four significant banks dominate the South African financial sector: First National Bank, Nedbank, ABSA, and Standard Bank. These four banks manage about 84% of total deposits and assets in South Africa (Westhuizen, 2013). Although not in scope for this research, there is recognition of emerging digital banks such as Tyme bank and Discovery bank, as well as the emergence of fintechs.

Because the banking sector contributes significantly to the global economy and its development, it is regarded as a pillar of economic growth and therefore authorities make every effort to ensure that it is well established and stable (Masindi & Singh, 2022) through stringent regulation. The South African reserve bank mentions primary regulations such as the Banks Act, the Financial Sector Regulation Act 9 of 2017, the Mutual Banks Act 124 of 1993, the Co-operative Banks Act 40 of 2007, and the Co-operatives Act 14 of 2005. Secondary legislation includes prudential and joint standards, bank regulations, cooperative bank regulations, mutual bank regulations, directives, circulars, and guidance notes (Regulated-institutions, n.d.). Banks strive to adhere to global regulatory and supervisory standards such as Basel frameworks and other requirements imposed by global and regional entities to which countries belong (Masindi & Singh, 2022).

The banking industry is currently facing various kinds of pressures other than the changing technological environment. There has been changes in the regulatory environment, product offerings, and number of participants that resulted in increased market competition (Westhuizen, 2013). Increased competition due to international banks entering the South African banking market, financial services provided by numerous companies like smaller banks and fintechs, who have targeted the low income and previously unbanked market, and the launch of new banking products are some of the challenges facing larger banks as it stands

(Westhuizen, 2013). All these changes put pressure on banks' lending activity, and as a result, their profitability.

This study was conducted focusing on one of the largest banks in South Africa (with branches in various African countries and the United States). The financial institution's total assets were valued at around 1.64 trillion South African rand (approximately 95 billion US dollars) as of 2021 (Statista, 2023), as well as headline earnings of about R20 billion (Businesstech, 2023). The financial institution currently has over 60 000 employees, justifying its classification as a large bank.

The South African banking sector faces a lot of different risks making it a good use case for this study. Banks in South Africa are subject to risks including those associated with changes in interest rates, commodity prices, exchange rates, political risks, and higher risks of defaulters due to the economic pressures and high unemployment rates. By examining enormous volumes of market data, big data can be used to assist in tracking and forecasting market developments.

1.4 Research problem

Since the global financial crisis, banks have placed greater emphasis on risk management, with a continuous focus on risk detection, measurement, reporting, and management. Risk management is not only one of the most regulated functions of the bank (Dvorski Lacković et al. 2020) but is also the basis where banks make their investment decisions and business enhancement strategies. Banks are primarily exposed to credit, market, and operational risks; additional risk categories include liquidity, business, and reputational risk. Historically, credit risk has been the biggest risk that banks have to deal with and the one that typically requires the most capital (Kazi Imran Moin, 2012).

According to McKinsey & Co., risk functions at banks will need to change significantly from their current state (McKinsey, 2015). The main concern for banks, and perhaps the need for enhanced risk management is to guarantee the safety of customers' electronic transactions due to the increased use of electronic banking services. In the absence of security, there is a risk of deterioration of the brand and customers' trust in the bank (reputational risk), in addition to financial losses (Kian and Obaid, 2022). Changes in risk management are anticipated to be driven by the expansion and depth of regulations, changing customer expectations, and the evolution of risk categories (Chitra, 2013).

The volume of data collected by financial institutions has increased significantly in recent years. A great deal of unstructured data is being produced and gathered often as a result of a strong push towards the digitalization of services and stricter regulatory reporting requirements. Despite the growing amount of data, tools, and insights, decision-makers are still not fully harnessing the power of current technologies (Elgendy et al., 2021) such as big data. According to the Ernst&Young Global Forensic Data Analytics Survey, 72 percent of participants believed that big data could play an important role in fraud detection and prevention, while only 7 percent were aware of specific big data technologies and only 2 percent used them.

Risk management is critical for all types of businesses, and as such there are numerous benefits of using big data in the banking sector such as the ability to detect potential risks like slow payers or poor investment (Eltweri, Alessio, & Osama, 2021). In the reviewed literature, many authors agreed that quick identification and quantification of emerging risks, as well as transparency when reporting, were critical components of risk management (Khamitkar & Rafi, 2020). This is the area where big data usage could be prevalent in accordance with the important role that risk management plays in the bank. However big data does not seem to be explored for risk management use and is especially very limited in non-financial and operational risk.

McKinsey (2015) stated that some banks will fall behind if they don't invest in the systems and infrastructure required to use big data for complex investigation and analysis. And although studies have been done on the modelling and fraud detection as well as practical application of big data analytics in certain risk management processes (Dvorski Lacković et al., 2020), they still don't extensively cover the usage thereof. As such, the problem remains that there is very minimal usage of big data in risk management that can hinder banks from making informed decisions and strategies, making it difficult for them to survive the ever-changing environments and tough competition.

1.5 Research questions

i. What is the perceived role of big data analytics in improving risk identification, assessment processes and mitigation strategies within the South African bank?
ii. What are the key challenges faced by the bank in effectively implementing and utilizing big data technologies for risk management?
iii. What factors hinder the adoption and integration of big data technologies and practices in the bank's risk management processes?

1.6 Rationale

The reason for undertaking this study was to understand why big data is not used extensively in risk management practices, and why it remains unrecognized as a tool in risk management frameworks across banks. It has been observed that banks throughout the world are looking for a digital convergence strategy to reach clients successfully in offering services and products, rather than only focusing on asset quality and regulatory compliance (Shakya & Smys, 2021), therefore it is key to close the gap in the usage of big data for risk management.

Furthermore, increasing the usage thereof can help address the problems and rivalry facing the industry. Big data analytics approach can be used to collect data helpful for fraud detection, risk management, and customer satisfaction (Shakya & Smys, 2021), helping bank managers, policymakers, and bank decision-makers gain a better understanding of their risk environment as well as coming up with informed strategies that will benefit their organizations. By addressing the research problem, banks can improve customer journeys through the leveraging digital platforms such as social media, sensor technologies, mobile devices, and scientific instruments to create engagement or promotion with customers about services (Shakya & Smys, 2021).

The study was worth conducting, because it helped spark awareness and put a spotlight on how risk management is traditionally conducted and help open possibilities of the potential that big data has in risk management. The findings can also contribute to the general body of knowledge by helping fellow researchers and field practitioners with further understanding of how big data can be applied in banking risk management, as well as the challenges experienced thereof.

1.7 Delimitations of the study

i. The study was limited to financial risk management in the context of a bank.
ii. The study was a case focusing on only one of the commercial banks within the South African context.
iii. The study included heads of risk management and risk specialists.
iv. The study focused only on the basis that big data is a large data set that can be mined to aid decision-making and was not intended to study the technical aspects such as data modelling and big data infrastructure.

1.8 Operating definitions

Risk Management

A risk is defined as the potential loss that may be incurred by a bank because of a particular event occurring (Hessami, 2010). The term "risk" typically refers to the possibility of financial loss or uncertainty in financial outcomes (Ragavan, 2003). Risk Management is a measure that is used for identifying, analysing, and then responding to a particular risk (Kanchu & Kumar, 2013). According to Mawutor (2014), Risk management constitutes the fundamental operation of a bank, encompassing the identification, quantification, and control of risk in order to facilitate a comprehensive comprehension of risk by the managers of said institutions. It is the process of aligning risk-related choices with the strategic direction and goals of the organization. Sufficient capital is available to undertake risks, and the risks taken are compensated by the expected payoff (Mawutor, 2014). For the purpose of this study, risk management will be taken to mean activities undertaken to curb the potential losses that may be incurred by a bank as a result of a particular event occurring.

Big Data

According to Maja & Letaba (2022), the term "big data" has gained widespread popularity and is commonly used to describe vast quantities of electronic data, including both structured and unstructured formats. Unstructured data is data that is difficult to process and evaluate using conventional data management systems because it does not follow a clear, ordered format (Maja & Letaba, 2022) such as customer surveys and social media data. Another source says that big data is a large collection of data that is difficult to store, process, analyse, and comprehend using traditional database processing tools (Grander & Silva, 2021). Upon examining various definitions provided by scholars, it can be concluded that the defining characteristic of big data is its vast and varied nature (Guptal, Verma, Kalra, & Kumar, 2014). Looking at the variations presented, big data will be taken to refer to datasets that are too vast for conventional data-processing methods and thus require new technologies (Misut & Jurik, 2021).

Big Data Analytics

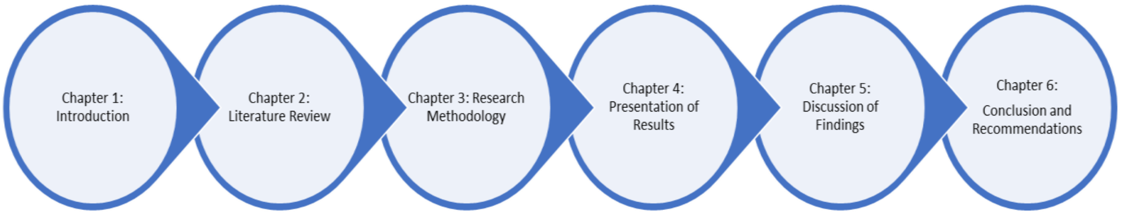
Big data analytics refers to the use of sophisticated analytical methods on large-scale datasets (Elgendy & Elragal, 2014). The use of analytics predicated on extensive data samples facilitates the identification and exploitation of business transformations (Elgendy & Elragal, 2014). Various authors write about big data analytics as the strategy of analysing a large volume of data gathered from various sources in an unstructured, semi-structured, or structured form using various analytic methods. However, for the purpose of this study, a definition that will be applied is big data as a holistic approach to managing, processing, and analysing big data sets by applying advanced analytics techniques (Elgendy et al., 2022), owing to the fact that the research is centred around conducting an analysis on actual risk data.

1.9 Assumptions

i. The participants in the research study provided a truthful and accurate representation of the risk management landscape.
ii. It is assumed that the existing risk management frameworks are applicable to the context of this study despite the unique adaptation of risk framework adapted at the bank being studied.
iii. This study assumes that the absence of investment in the necessary systems and infrastructure for using big data may result in certain banks encountering difficulties and lagging behind.

1.10 Chapter Outline

This study is divided into six chapters. The first section presented the background of the study and motivation for undertaking the specified topic. The second chapter looks at the literature, by taking the reader through important definitions and origin of concepts, and further dive deeper into the status of risk management and big data in the banking sector as well as what literature says about the two subjects. The third chapter focuses on the research methodology and how the overall research was conducted. The fourth chapter is a presentation of the results as well as the process that was followed for analysis. This chapter is then followed by chapter 5, which is a discussion of the results and chapter 6 which is the conclusion as well as recommendations for future research. All supporting documents can be found in the appendices section.



CHAPTER 2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 Introduction

This chapter sets out an overview of the literature relating to the challenges of using big data in risk management, where work from various researchers was collected and analysed. The section begins by exploring risk management and big data respectively as bodies of knowledge, introducing the concepts and their different constructs. The relationship between the two disciplines is further established by looking at literature pertaining to the role that big data plays in risk management, as well as how it has been applied in different sectors. It further looks at the challenges as well as benefits thereof. Furthermore, the gaps found in the literature are explored and analysed, providing more justification for pursuing this study.

2.2 Overview of Risk Management in the banking sector

In a study conducted to highlight the latest amendments proposed by the Basel Committee for managing banking risks, Turgut (2019) defines risk management as the systematic development and implementation of a strategy to deal with potential business losses. The author further states that the primary goal of risk management in banking is to manage an institution's exposure to losses or risks and to protect the value of its assets. According to International Organization for Standardization (ISO) risk management guidelines, risk management is a group of activities coordinated to direct and control an organization with regard to risk (ISO, 2018). It encompasses proactively creating strategies, planning, leading, and organising the range of hazards that are woven into the fabric of an organization's daily and long-term operations (Kanchu & Kumar, 2013).

According to BCBS (2001), risk management comprises four distinct processes. The first process involves identifying events and categorizing them into market, credit, operational, or other risks, which are then further divided into specific sub-categories. The second process involves assessing these risks using data and risk models. The third process involves monitoring and reporting the risk assessments in a timely manner. Finally, that senior management is responsible for controlling and managing these risks (Rosman, 2009).

There are various definitions of the term “risk”, and for the purpose of this study, a risk is defined as the potential loss that may be incurred by a bank because of a particular event occurring (Hessami, 2010). There are various types of risks that the bank gets exposed to while operating. The management of the bank is taking on more risk in an effort to boost profits for its stockholders. The risks that banks confront are numerous and include foreign exchange risk, country or sovereign risk, interest rate risk, market risk, credit risk, off-balance-sheet risk, liquidity risk, and bankruptcy risk. The performance of a bank is largely dependent on how well these risks are managed. Furthermore, banks are the focus of regulatory attention due to these risks and their function in financial systems (Saunders, 2006).

Although not an exhaustive list given the rapid changes in the banking environment, a traditional bank will at minimum have the following risk types to manage, as indicated by Dvorski Lacković et al., (2020):

- Credit risk – Risk of loss as a result of borrower default or rating deterioration.
- Operational risk – Losses that occur as a result of internal processes, people, systems, or external events.
- Market risk – the risk of loss due to changes in the price or the exchange rate of an instrument.
- Liquidity risk – Losses that occur because of the inability to sell an instrument on the market quickly enough (market liquidity) or because the financial institution is unable to pay its liabilities on time (funding liquidity).
- Interest rate risk in banking book – Losses resulting from the effects of changing interest rates on banks' economic capital.

Banking institutions are vulnerable to various hacks and breaches, and this is not a new fact (Srivastava & Gopalkrishnan, 2015). Regulation requires banks to define the overall risk management framework from which to manage its risks, as the lack thereof can have significant impact on risk appetite definition, strategic decision making, stress tests and scenario analysis, and risk reporting (Dvorski Lackoviæ et al., 2016). McKinsey (2015) suggests that the following expected capabilities are crucial for banks to be compliant with the regulators:

I. Accurate and complete data sourcing from catalogued sources received in a timely manner.
II. Data storage and processing must be effective and able to accommodate historical analysis.

2.3 The concept of Big Data

Big data is a new technology paradigm for data that is generated at a high velocity, volume, and variety. Big data is envisioned as a game changer capable of changing the way firms operate in a variety of industries (Lee, 2017). Below we look at the evolution of big data as well as the concepts, methodologies, and application of big data.

2.3.1 *The evolution of Big Data*

The concept of big data arose from the need to understand trends, preferences, and patterns within the massive database generated when people interact with various systems and each other (More & Moily, 2021). The word "big data" is said to have first appeared as a concept in 1974, then again in editorials in 2006 and 2007, and from 2008 began to appear regularly in scientific articles (Picciano, 2012). While big data has only recently emerged, the act of acquiring and storing enormous volumes of data dates back to the early 1950s when the first commercial mainframe computers were introduced, and decreased between the early 1950s and the mid-1990s due to the high cost of computers, storage, and data networks (Lee, 2017). Amid this claim, the concept of 'big data' is new and has very obscure roots (Gandomi & Haider, 2015).

According to Lee (2017), big data has progressed through three distinct stages:

- Big data1.0 is said to have occurred between 1994 and 2004 and coincides with the introduction of e-commerce in 1994 when online enterprises were the primary providers to web content. Web mining techniques were used to evaluate users' online actions, which are classified into three categories: web usage mining, web structure mining, and web content mining. Structure mining studies the structure of a website, whereas usage data captures user identity and browsing behaviour.
- Big data 2.0 cycling from 2005 to 2014 was driven by Web 2.0 and the social media phenomenon. Social media analytics is the study and interpretation of human activity on social media platforms, providing insights and drawing conclusions based on a consumer's interests, online surfing patterns, friend lists, moods, occupation, and opinions.
- Big data 3.0 that is said to have started in 2015 onwards, including data from big data 1.0 and big data 2.0, refers to a technological environment in which Internet of Things (IoT) applications generate data in the form of images, audio, and video. With the rapid expansion of the Internet of Things, connected devices and sensors will overtake social media and e-commerce websites as the key providers of big data.

According to Figure 1 below, showing the frequency of interest in big data over time, the term gained popularity around mid-2011 and skyrocketed in 2013. The hype can be attributed to companies such as IBM and other top technology corporations that invested in developing the specialist analytics sector (Gandomi & Haider, 2015).

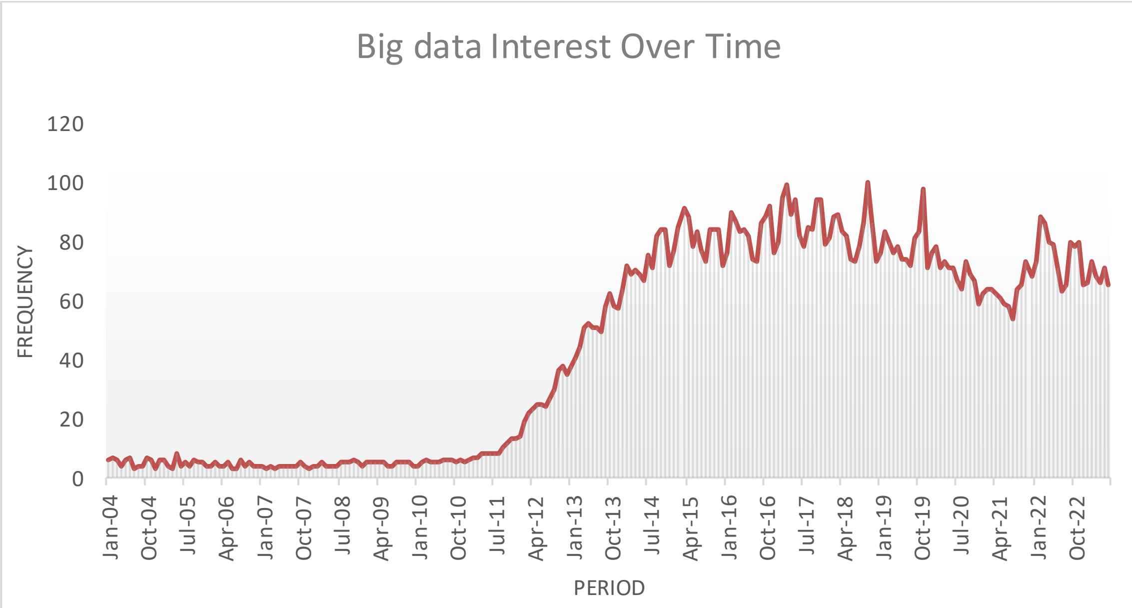


Figure 1: Google trends: Frequency distribution of documents containing the term "big data".

It is said that the market for big data is expanding and was projected to attain a value of $67 billion by 2021, with the leading position being held by Data Analytics software. The prevalence of electronic information storage exceeds 98% globally, and the corporate sector has embraced this form of data management due to its potential advantages. Conventional methods for data analysis have become obsolete and necessitate the adoption of advanced and intricate techniques (Adil & Abdelhadi, 2022).

2.3.2 *Defining Big Data*

Big data is defined as a large collection of data that is difficult to store, process, analyse, and comprehend using traditional database processing tools (Grander et al., 2021). It is further defined by More and Moily (2021) as processes and tools involved in managing and utilizing large data sets. Shabbir and Gardezi (2020) define big data as massive amounts of numerous observational data used in decision-making as well as real-time information, non-traditional media data, IT-driven data, social media, and large volume data. The sources of traditional data and big data are distinct from one another, as shown in Table 1 below. Data volume and data source are the two factors that set these two apart:

Parameters	**Traditional Data**	**Big Data**
Type of Data	Structured	Unstructured
Volume of Data	Terabyte	Petabytes and Exabytes
Architecture	Centralised	Distributed
Relationship between data	Known	Complex
Sources of Data	• Documents • Finances • Stock Records • Personnel Files	• Photos • Audio and Video • 3D Models • Simulations • Location Data

Table 1: Big data vs. traditional data (Kiziltan, 2018)

Figure 2 below shows various sources of big data. Internal sources contain data produced within an organisation, consisting of both history and real time data, whereas external sources are data generated outside of the organisation.

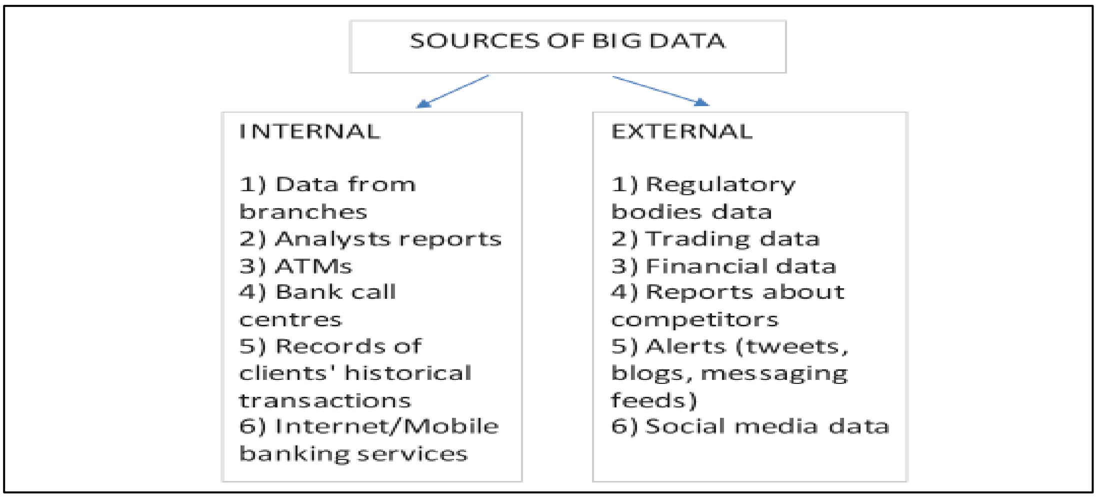


Figure 2: Sources of big data (Dvorski Lacković et al., 2020)

In an online survey of 154 global executives conducted, the following definitions of big data were found, depicted in Figure 3 below:

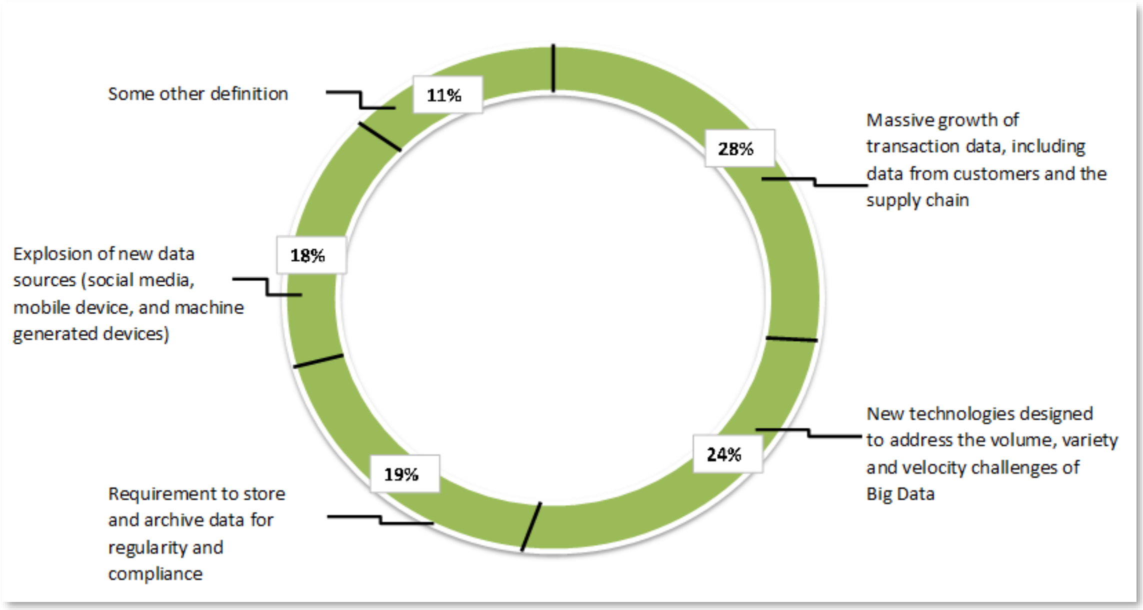


Figure 3: Big data definitions (Gandomi & Haider, 2015)

For a more comprehensive definition of big data, Dicuonzo et al., (2019) mention that, big data cannot be handled using traditional tools and techniques because of its high volume, variety, velocity, value, and veracity. Volume refers to a large amount of data, whereas variety refers to the various types of data collected from structured and unstructured data sources. Velocity refers to the rate at which data is collected, processed, updated, and analysed. The strategic and informational advantages of big data are referred to as value, while veracity refers to the dependability of the data sources (Dicuonzo et al., 2019). Variability and visualization have also been added in recent years. Variability refers to how the insights change as the interpretation of information changes or as new data is added to the mix, and visualization is the meaningful representation of data as to reveal hidden patterns and trends (Elgendy et al., 2022).

The hype around big data has also resulted in a widespread misunderstanding of what the word entails. Taking that into consideration, we look at the three-part definition established by Boyd and Crawford, that big data is made up of the following components: 1. Technology: increasing computer power and algorithmic precision to collect, analyse, link, and compare enormous data sets. 2. Analysis: the process of identifying patterns in big data sets to establish economic, social, technical, and legal claims. 3. Mythology: the widely held assumption that enormous data sets provide a higher level of intelligence and understanding that can yield previously impossible discoveries with the aura of truth, objectivity, and correctness (generating large amounts of data) (D'Ignazio & Bhargava, 2017).

2.3.3 *Application of Big Data*

For big data to be useful, there are technologies and methods that must be employed such as big data analytics to make sense of the data. Below we explore both the technologies available as well as how big data can be analyzed for useful insights. We then look at examples of sectors where these technologies and methods have been applied.

2.3.3.1 Big Data Analytics

Big data analytics is defined as the process that analyses big data sets using algorithms to extract diagrams, reports, and useful and unknown information (Dicuonzo et al., 2019). According to More and Moily (2021), big data analytics, loosely called BDA, is the complex process of examining large and diverse data sets or big data to reveal information such as hidden patterns, unknown correlations, market trends, and customer preferences that will assist organizations in making informed business decisions. Furthermore, BDA can also be defined as a set of instruments that can provide valuable intuitions for decision-making, review corporate business performance, establish competitive advantages, and therefore increase enterprise value (Dicuonzo et al., 2019).

BDA can also be looked at as a group of tools that can generate intuitions that are useful in the decision-making process, assess financial institution business performance, establish competitive advantages, and, therefore, increase enterprise value (Dicuonzo et al., 2019). Dicuonzo et al., (2019) further state that it is a method used to extract valid, useful, previously unknown, or hidden models and information from large data sets, as well as to identify important relationships between variables, ensuring a competitive advantage. The use thereof can enable the development of actionable ideas for measuring performance, gaining a competitive advantage, and serving as a new platform for productivity, innovation, and better data-driven decision-making (Elgendy et al., 2022).

According to More & Moily (2021), big data analytics is the complex process of examining large and diverse data sets or big data to reveal information such as hidden patterns, unknown correlations, market trends, and customer preferences that will assist organizations in making informed business decisions. It can make fresh discoveries that were previously concealed because of limitations in conventional data systems, and it can be viewed as a catalyst for process and performance improvement.

Several studies have examined how BDA affects organizational performance, indicating that companies that use data-driven strategies are more productive and profitable than their competitors. However, in most of the literature exploring big data application across multiple sectors, most authors focus predominantly on the use of big data analytics (BDA), particularly in healthcare and marketing (and sales), and very limited in the banking sector or use in risk management. There seem to be a consensus across consulted literature that merges the traditional use of data analytics into big data, and the keyword remains “Big”- to indicate an analysis of big data sets.

To derive value from big data sets, data has to be retrieved and analytics methods applied to it in order to draw meaningful insights. The practice of extracting intriguing, nontrivial, implicit, previously undiscovered, and possibly valuable patterns, or knowledge from massive amounts of data is known as data mining (Kazi Imran Moin, 2012). According to Braun H(2015), data mining is the practice of retrieving potentially valuable information from vast, imperfect, ambiguous, and noisy datasets.

Data mining is increasingly becoming a strategically significant field for many businesses, including the banking industry. It is a procedure that involves examining the facts from several angles and condensing it into insightful knowledge. The data could be textual, online, multimedia, geographic, or time series stored in data warehouses and data lakes (Kazi Imran Moin, 2012).

Data mining helps banks find unknown relationships in the data and uncover hidden patterns within a group (Chitra, 2013). More & Moily (2021) notes the following data mining techniques for analysing big data:

Method	Application
Classification	Utilizes a collection of pre-categorized instances to construct a model capable of categorizing a vast population of data entries. The analysis of fraud detections is highly suitable for this particular type of investigation.
Clustering	Clustering refers to the process of identifying groups of objects that exhibit similar characteristics. This technique involves the grouping of transactions that exhibit similar behaviour.
Association Rule	The main objective of association rule mining is to ascertain sets of binary variables that exhibit a high frequency of co-occurrence within a transaction database. Association rules are a type of logical statements that facilitate the identification of connections between seemingly unrelated data within a relational database or any other data repository.
Prediction	One of the techniques employed in data mining involves the identification and analysis of the associations between independent variables and dependent variables.

Regression Analysis	The process of constructing a mathematical representation to analyse the association between multiple independent variables and multiple dependent variables. There are several regression techniques that are commonly used in statistical analysis, including linear regression, multivariate linear regression, and nonlinear regression.

Table 2: Big data analytics data mining methods

Data mining provides an answer by revealing relationships and hidden patterns in this data. Banks can use data mining to evaluate loan default risks, find prospective clients for new products, anticipate customer reactions to interest rate changes, and improve the profitability of their customer relationships (Kazi Imran Moin, 2012).

There are three models of big data analytics: descriptive, predictive, and prescriptive. Descriptive analytics summarizes past performance and is limited in its predictive abilities. Predictive analytics uses a variety of statistical, modelling, data mining, and machine learning techniques to study recent and historical data. Prescriptive analytics involves predicting the probability of future outcomes and automatically acting based upon predicted outcomes. Below are examples of application of these different methods according to Kushwaha et al., (2021):

Method	Application
Descriptive Analytics	Descriptive analytics, which is typified by traditional business intelligence (BI) and visualizations like pie charts, bar charts, line graphs, tables, or generated narratives, is the analysis of data or content, to address the question "What happened?" (Or What is happening?).

Discovery Analytics	The process involves the extraction of early indicators by means of text summarization, as well as the extraction of features from images and videos.
Predictive Analytics	Analytics that seek to answer the question "What is likely to happen?" using data and information, driven by sophisticated machine learning techniques and a variety of econometrics models.
Diagnostics Analytics	A type of advanced analytics where the "Why did it happen?" inquiry is addressed by looking at data or content. Techniques like drill-down, data mining, correlations, and discoveries explain it.

Table 3: Big data analysis methods

2.3.3.2 Big Data Analytics as a driver for decision making

Big data is a new driver of competitive advantage, driven by the need to analyse vast amounts of data and use intractable processes (Soltani Delgosha et al., 2021). However, the technology gap between industrial applications and decision-makers remains a challenge. Firms that bridge this gap quickly gain significant competitive advantage by uncovering information that is not immediately available to others (Adil & Abdelhadi, 2022). According to Elgendy et al., (2022), organizational performance and competitiveness have always been built on decisions. Many successes and failures have been attributed to a single, fate-changing decision over the years (Elgendy et al., 2022). With the advent of big data, the information requirements of executives have changed (Misut & Jurik, 2021) and enormous pressure has been placed on decision-makers to make the best decision possible in a timely manner (Elgendy et al., 2022). Moreover, the utilization of data-driven decision-making is widely regarded as a means to enhance the quality of decisions by integrating the intuition and expertise of human decision-makers with rigorous data analysis. This approach facilitates the adoption of more rational choices, ultimately yielding superior outcomes (Elgendy et al., 2022). With BDA, decisions in business are made using data and information rather than intuition. The decision-making process has shifted away from relying on intuition and now prioritizes data-driven approaches.

Data-driven decision-making is widely recognized as a constant process encompassing various stages, namely data collection, data transformation into meaningful information, and ultimately, knowledge acquisition. This knowledge serves as the foundation for making informed decisions, which are subsequently monitored for implementation and accompanied by feedback throughout the entire process (Misut & Jurik, 2021). Additionally, there are large amount of structured, semi-structured, and unstructured data from various sources, making it possible to use the value hidden in this data for strategic, tactical, and operational decisions in businesses in a variety of ways (Misut & Jurik, 2021). Several existing businesses as well as new businesses are investing in developing BDA capabilities, where the goal is to generate actionable insights from the volume, variety, and veracity of data, allowing individuals or organizations to make and communicate informed decisions (Kushwaha et al., 2021).

For the banking sector, big data analytics can improve the extrapolative power of risk models used by banks and big data. Big data analytics in risk management business intelligence (BI) tools can identify potential risks associated with money lending processes in banks. And with the help of big data analytics, banks can analyse market trends and decide on lowering or increasing interest rates for different individuals across various regions (More & Moily, 2021). As seen in the preceding sections, BDA can aid in faster and more informed decision-making. Not only will it bring new insights to the banks, but it can also enable them to stay a step ahead of the game with advanced technologies and analytical tools (More & Moily, 2021). Figure 4 below from Picciano (2012) shows the basic data-driven decision-making process, where information systems encompass a computerized database system capable of storing, manipulating, and reporting on a wide range of data:

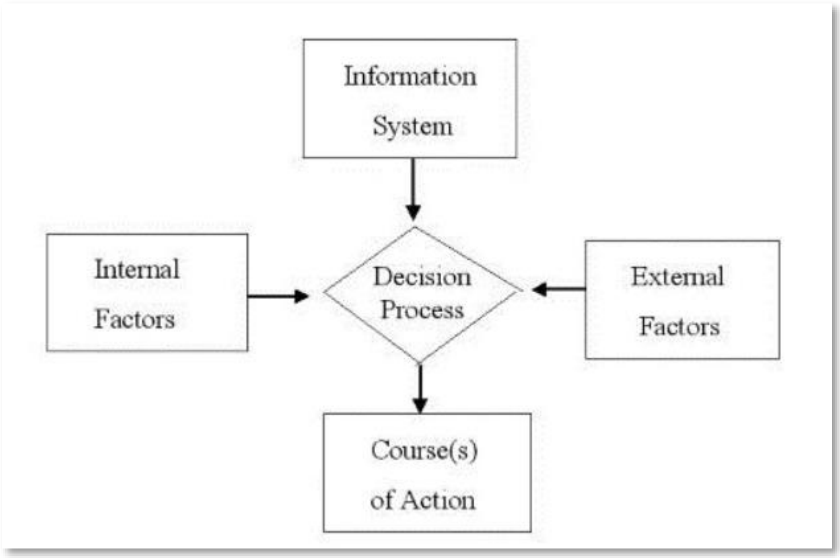


Figure 4: Data-driven decision-making (Picciano, 2012)

Data warehousing, data mining, and data disaggregation are all terms connected to big data application in data-driven decision making. The next section will evaluate these technologies in detail.

2.3.3.3 Big Data Analytics Technologies

Big data technologies are tools or technologies that are used to efficiently analyse data that has been classed as large data (Amanullah et al., 2020).

Big data is processed using various major technologies such as Hadoop, HDFS, NoSQL, MapReduce, MongoDB, Cassandra, PIG, HIVE, and HBASE that work together to achieve a common purpose such as extracting value from data previously regarded as dead (Zakir et al., 2015), as well as offer the data warehousing capabilities needed to process big data. Data warehousing is simply a database information system that can store, integrate, and retain enormous amounts of data over time (Picciano, 2012). The storage tools that are commonly used include Teradata, Hadoop, Amazon RDS, and Pentaho. Various data mining tools that are available for use include Python, IBM Cognos, Apache, R, and Google Big Query. The analytics tools utilized in this context include Apache Spark, Tableau, and SAS Visual Analytics (Maja & Letaba, 2022). According to Maja & Letaba (2022), data visualisation and data mining tools such as Tableau, Python and IBM Cognos are useful for visualising, understanding, and interpreting data trends, patterns, and anomalies.

The most widely used framework for big data is Hadoop, which mixes commodity hardware with opensource software (Zakir et al., 2015). According to Sagiroglu and Sinanc (2013), Hadoop is an open-source platform with a Java foundation. The following features are provided by Hadoop, which consists of a distributed file system, analytics and data storage platforms, and a layer that provides parallel processing, workflow, and configuration management:

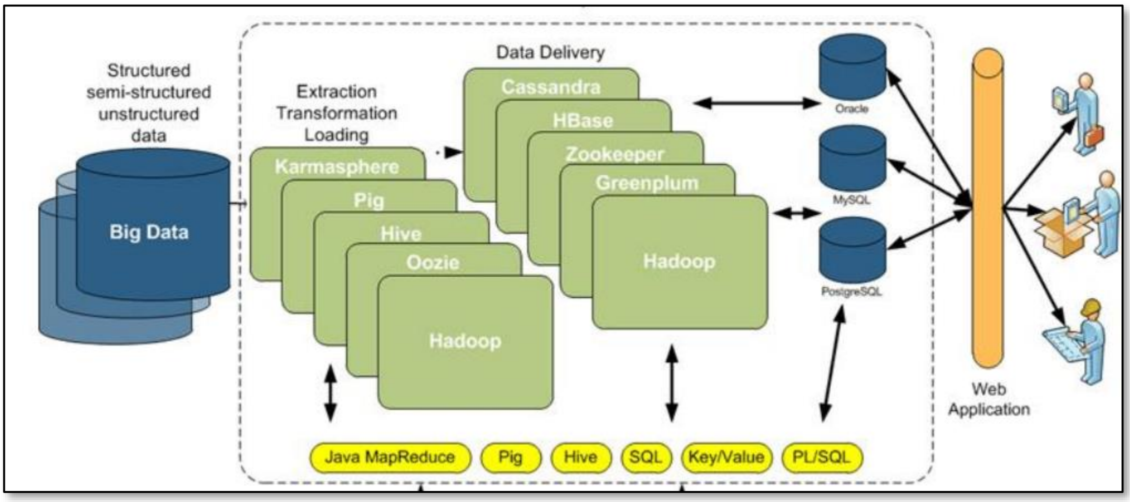


Figure 5: Hadoop architecture (Sagiroglu, 2013)

- HDFS is a highly fault-tolerant distributed file system that oversees data storage on clusters.
- MapReduce is a strong parallel programming approach for cluster-based distributed computing.
- HBase is a scalable, distributed database that allows for random read/write access.
- Pig is a high-level data processing system for evaluating data sets that are generated by a high-level language.
- Hive is a data warehouse program with a SQL-like interface and relational model.

- Sqoop is a project that allows you to transfer data between relational databases and Hadoop.
- Avro is a data serialization system.
- Flume is a dependable and distributed streaming log collection system.
- Zookeeper is a centralized service that provides group services and distributed synchronization.

These technologies do necessitate a skill set that is unfamiliar to most IT departments, who will have to work hard to integrate all important internal and external data sources (Zakir et al., 2015).

2.3.3.4 Sector application

Kushwaha et al., (2021) looked at industries where big data can be used. These are (1) Healthcare management, (2) Crisis management, (3) Governance management, (4) Smart manufacturing or industry 4.0, (5) Dynamic capabilities management, (6) Decision support systems, (7) Business models management, (8) Network management (primarily social media platform-oriented), (9) Services Management, (10) Search engine optimization (SEO), (11) Digital management, and (12) Financial management. For financial management, they found that BDA has bridged the gap through econometric models for financial studies, research, and data science implementation (Kushwaha et al., 2021), however there is not enough evidence to support the use in banking risk management. Taking into consideration the benefits presented in sectors above, it is apparent that banks can equally have an advantage if they increase the use of available data to predict, manage, and report risks (Dicuonzo et al., 2019).

Below are examples of how big data analytics has been applied in other sectors:

I. **Energy**: It was estimated that industrial production activities consume approximately 35% of the total global electricity supply and produce approximately 20% of total carbon emissions. As a result, there is a strong interest in using big data analytics to develop new approaches that allow for the realization of sustainable production cycles while minimizing energy consumption (Arena & Pau, 2020).

II. **Healthcare**: The global pandemic, COVID-19, generated a massive and diverse amount of data, which is rapidly increasing. This data can be used in a variety of ways, including diagnosis, risk score estimation or prediction, healthcare decision-making, and the pharmaceutical industry (Alsunaidi et al., 2021).

III. **Marketing and sales**: The study considered was around the use of big data analytics to improve customer loyalty programs using software as a service (SaaS) to collect and analyse data from an e-commerce platform to provide users of the portal with a list of recommended products (Arena & Pau, 2020). The adoption of big data analysis in retention strategies is worthwhile not only for the attainment of new customers but also to enhance the profits generated by pre-existing clients, as it allows the industries to adapt their productive strategies directly to the needs of consumers, thus seeking to maximize revenues (Arena & Pau, 2020).

Looking at the finance sector specifically, the significance of big data emerges most prominently in the banking and insurance sectors because of intense supervisory authorities (e.g., European Central Bank, European Security, and Markets Authority), which impose stringent capital regulations (Dicuonzo et al., 2019). The cases of large bank losses and the occurrence of the global financial crisis necessitated stricter regulation of the banking sector (Lacković et al., 2020). According to Dvorski Lackoviæ et al., (2016), big data is applicable on multiple levels in the banking business, including retail (branch and mortgage banking, credit card processing), commercial (credit risk analysis, non-credit products, client management and sales, middle market lending), capital markets (trading and sales, underwriting and structured finance, no depository credit institutions), and asset management (wealth management, asset investment management, asset issuer services, global asset reporting, and investment deposit analytics) (Dvorski Lackoviæ et al., 2016).

According to Shakya and Smys (2021), using big data capabilities has the potential to reduce overhead costs associated with the reporting, auditing, and transaction verification processes. Dicuonzo et al., (2019) state that, banks that use BDA have an advantage of 4% in market share over banks that do not. This shows that big data usage can provide numerous benefits to both banks and their customers. The following are some of the globally recognized benefits of using big data for banking according to More and Moily (2021):

I. **Customer Segmentation**: This segmentation enables banks to provide or deliver exactly what their customers require according to customer bases that are divided into groups of people with similar interests relevant to marketing and business, such as age, gender, financial situation, interests, and spending habits.

II. **Risk Management**: Early risk detection is a large and important part of risk management, and BDA can help with risk management as well as fraud detection. The location and presentation of big data on a single large scale make it easier to reduce the number of risks to a manageable number. This reduces the likelihood of data loss or ignoring fraud in banking transactions.

III. **Fraud Detection and Prevention**: As one of the major issues confronting the financial sector, using big data can help assure banks that no unauthorized transactions or access will be made from their systems, providing a level of safety and security that will raise the overall financial industry's security standard.

Indriasari et al., (2019) further state that accurate customer analytics, risk analysis, and fraud detection are some of the ways that the banking business can profit from the big data analytics sector. These methods can result in transactions that are more intelligent, which can help businesses control risks and offer more individualized services, providing them a competitive advantage (Indriasari et al., 2019). Banks utilize big data primarily to improve customer service, risk assessment, decision-making assistance, evaluate and assess new chances for profitability, invest in new markets, shorten the time needed to enter the market, and invest in blockchain projects (Soltani Delgosha, et al., 2021).

The banking industry is one of the corporate sectors that invests the most in big data technologies (Al-Dmour et al., 2022). However, the usage or application specifically in banking risk management remains poorly explored as reflected by the limited literature on the subject.

2.4 Challenges of Big Data usage in Risk Management

According to Forrester Research, firms use less than 5% of their accessible data effectively (Zakir et al., 2015). Despite the boost big data has received in the recent past, the problem of applying it for risk management practices is still underexplored. Most of the past studies focus on BDA algorithms, while few of the other studies relate the adoption of big data to human skills (Kushwaha et al., 2021). To date, only a few rigorous and large-scale studies have been conducted to investigate how the adoption of big data can improve risk management. According to Dvorski Lacković et al., (2020), there is a number of challenges experienced with the use of big data techniques, namely:

I. Having to deal with the issue of larger than usual data storage resulting from the need to keep historical data for many years. This historical data is used for trend analysis and other complex analytics, which poses a significant challenge to the storage system's reliability.
II. Working with multiple data types that for example come in unstructured formats, as most big data today is generated from multiple sources.
III. Creating a high-performance computing system to process distributed historical and incoming data.

Bedeley and Iyer (2014) identified one of the challenges of big data usage as legacy. They claim that even though a particular system may have been chosen to replace a number of legacy systems in a functional domain, the implementation won't be effective unless the replacement system is able to identify, document, and meet all of the specific important criteria of the legacy systems. They further argue that the challenge is not only the collection and management of vast amounts of data in various types that is a challenge, but also extracting meaningful value from it. Here, the authors only speak of technological legacy, and do not go further to address legacy from culture, change readiness, and transformation. The state of readiness is impacted by perceptions that the planned change is necessary, meaningful, and suitably supported by the context in which it will occur (Shah et al., 2017). If there is a people legacy, the change in systems may not necessarily mean the adoption of big data practices, this may very well be the impediment to the usage and not system legacy.

In the data collection from a study based on a bank located in Italy, Dicuonzo et al., (2019) found that:

I. The data analysis tools available to the bank aided in decision-making by proposing operational and strategic solutions.
II. Based on their interviews, the bank's interest in large amounts of data and the resulting information has led to numerous investments in advanced IT technologies.
III. Although investments were made in the technologies required to use big data, new skill sets were required from risk managers, who needed to have stronger quantitative skills in data management and analysis.

This, however, only discusses a fraction of big data aspects and fails to provide a comprehensive representation of application in risk management. These ideas, however, seems to focus more on the theoretical impediments of the usage and not necessarily what may be the reasons in practise.

Most of the studies done relating to big data analytics applications are more qualitative in nature and are mostly literature reviews and case studies. Although there have been extensive studies in healthcare and marketing, there is still very few studies done on the application in banking, and more specifically risk management.

Existing research on the application of big data in risk management in banks has mainly focused on the following categories:

I. The drivers and challenges of using big data in finance.

II. Framework for big data usage
III. The use of big data in specialized risk management processes such as modelling and fraud detection

According to the review of the literature, there is a significant gap in the literature that examines why the banking industry is not yet leveraging big data potential to curb potential risks and losses. The existing body of literature offers insightful information on a number of big data analytics topics related to the financial sector, including risk management, fraud detection, customer analytics, and operational effectiveness. But even with the clear benefits of using big data in these domains, there is a noticeable lack of comprehensive research and knowledge on the obstacles or difficulties banks face when implementing and utilizing big data.

In the absence of big data, financial institutions might find it difficult to carry out thorough and timely risk assessments. Due to a lack of timely insights into developing risks, the institution may become exposed to unanticipated events or abrupt market shifts. Banks that do not use big data may find it difficult to comply with the intricate and constantly changing financial regulations, which could result in fines and compliance violations. Financial institutions risk losing market share and competitive advantage if they don't use big data in risk management and trail behind their competitors. However, these issues are still not addressed in literature.

This study seeks to bridge the gap of knowledge by investigating reasons why big data has not been widely adopted in risk management and the challenges faced in this regard.

2.5 Theoretical underpinnings

This section explores the theoretical framework that will be adopted to guide the study, as well as the conceptual framework that was developed to show the relationship between concepts being studied.

2.5.1 Theoretical Framework: Risk Management Framework

To further unpack the role of big data in risk management, the theoretical framework is explored based on the claims that big data can be useful for the management of risk in banks, as seen in functions such as health, marketing and sales. The primary justification for this claim is that using big data analytics for risk management allows banks to manage their risks at an early stage (Dvorski Lacković et al. 2020). After reviewing the literature, it was discovered that most of the big data usage in risk management is in fraud detection, market risk and credit risk, but very limited in non-financial and operational risk. Currently, the utilization of big data for credit investigation has become prevalent in the consumer finance sector, as well as other forms of internet-based consumer finances (Wang, 2021). According to Wang (2021), Credit risk big data offers various capabilities such as personal credit investigation services that rely on personal credit reports, risk profiles, and bank account verification.

Rosman (2009), suggests a framework that investigates the correlation between risk management procedures and the four fundamental components of the risk management process which are: Identifying, analysing, measuring, and defining the intended risk level through risk control and risk transfer as a widely acknowledged risk management approach (Rosman, 2009), depicted in Figure 6 below. The framework posits a positive correlation between risk management practices and the risk management process aspect. Secondly, it proposes an examination of the risk management processes category that significantly impacts the implementation of risk management (Rosman, 2009).

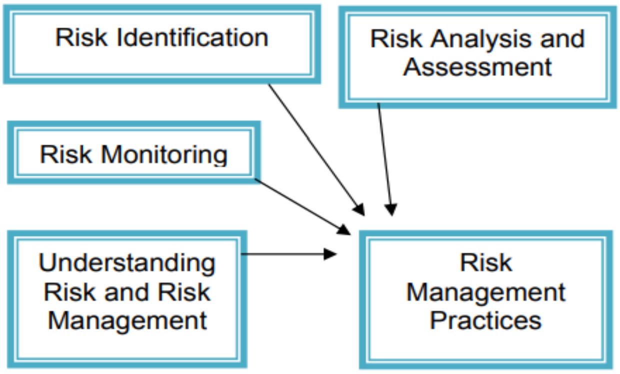


Figure 6: Risk Management Framework (Rosman, 2009)

However, this framework only covers risk management activities and not how capabilities such as big data can be applied to enhance the activities.

Tekiner & Keane (2013) suggest a big data framework used to establish a connection between strategic decision-making and practical situations. This was in comparison with established methodologies such as TOGAF, Zachman, and Gartner which offer approaches for overseeing extensive information systems. This framework is made up of three (3) stages: Stage 1- Multiple Data Sources - Choose the Right Data; Stage 2- Data Analysis and Modelling; Stage 3- Data Organization and Interpretation. Stage 1 encompasses the acquisition of data and its subsequent filtration through the utilization of metadata and various processes. This procedure confers added value and facilitates the attainment of a competitive edge for organizations. In Stage 2, the provided data is utilized for the purpose of applying analytics and predictive models, with a specific emphasis on present and forthcoming associations (Tekiner & Keane, 2013).

Dicuonzo et al., (2019) suggest a framework that shows ways that big data can be used in relation to these risk management activities as depicted in Figure 7 below:

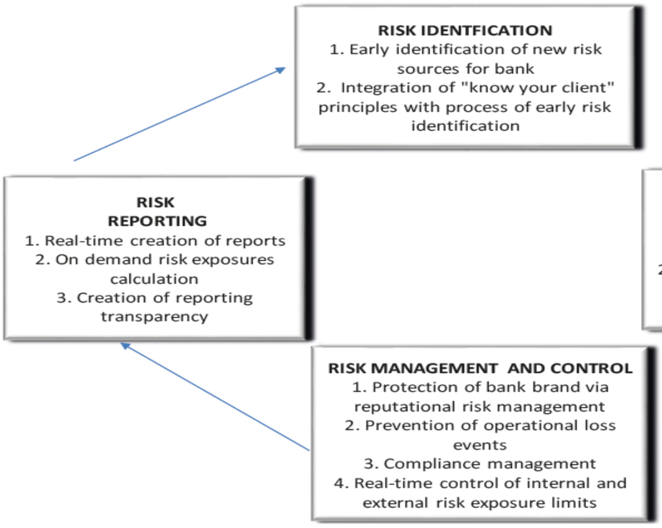


Figure 7: Risk Management Framework (Dvorski Lacković et al., 2020)

1. Risk identification is about the use of big data to identify new sources for early risk identification as well as in-depth customer knowledge.
2. Risk assessments are performed as underlying data analysis through the development and calculation of various risk indicators, real-time simulation of risk indicators, and predictive analysis for all risk typologies.
3. Risk management and control is the implementation of controls across business processes to mitigate and manage reputational risk, curb operational losses, compliance management, and real-time financial risk control.
4. Reporting is about generation of real-time reports, on demand risk exposure calculations, increased transparency, and real-time on demand stress testing.

This framework was adopted for this study when investigating the challenges of using big data for risk management in banking. This is because it is the only framework found to cover both big data and risk management and demonstrate the application thereof. Most frameworks looking at big data and risk management focus on data governance and the technical big data frameworks e.g., MapReduce. The available risk management frameworks also follow the generic, widely accepted four-way framework as can be seen in both Risk Management Frameworks by Dvorski Lacković et al., (2020) and Rosman (2009).

The chosen framework is deemed fit as it aligned with this study and the underlying research questions because it depicts the application of big data in risk management and helped underpin where challenges can be in every stage of application.

2.5.2 *Conceptual Framework*

There is an explosion of data growth that can benefit organisations with gaining new industry insights, help with regulatory compliance as well as the reduction of inherent risk through application of big data capabilities. It was established from literature that there is potential for big data to be applied risk management processes and Figure 8 is a depiction the proposed conceptual framework that shows how BDA techniques can be applied to risk data to form part of the risk management processes.

Jabareen (2009) posit that a conceptual framework serves to delineate the essential factors, constructs, or variables, and assumes interrelationships among them. The conceptual model presented for this study is constructed based on the literature and the theoretical framework presented, combining ideas from literature as well as the frameworks presented under theoretical underpinnings. This conceptual framework does not seek to provide a causal relationship but rather offers an explanatory approach (Jabareen, 2009) to the research problem as extrapolated in literature. This will be used to explore where the challenges are mostly experienced at various stages of the risk management framework.

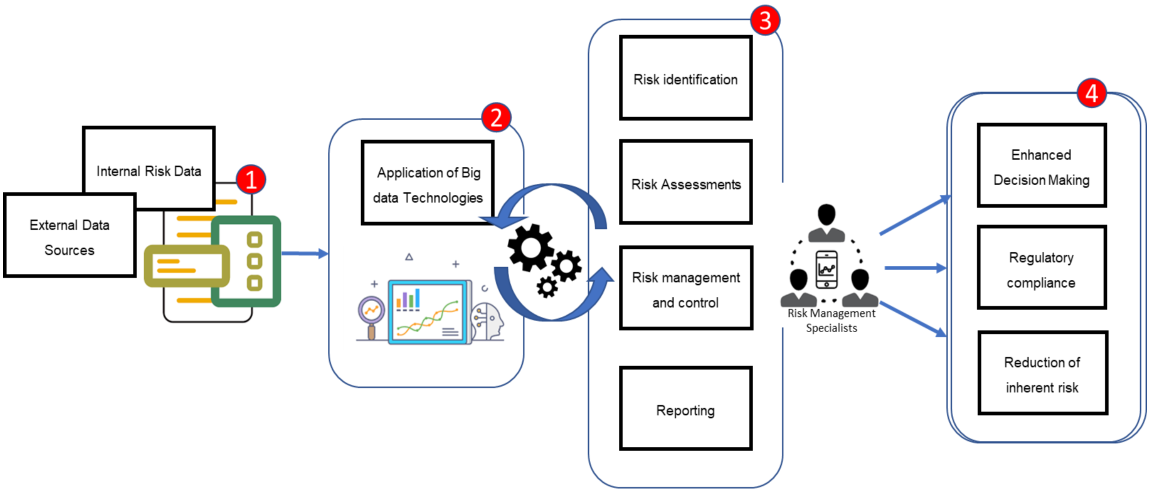


Figure 8: Conceptual framework (Researcher's own interpretation)

The conceptual framework is made up of four main stages or touch points.

1. The collection of data from various sources, made up of both internal and external sources (see Figure 2: Sources of big data.
2. The second stage is about ingesting the data sources. Internal sources refer to data that is generated within an organization, including both historical and real-time data. Conversely, external sources are comprised of data that originates outside of the organization.
3. Stage three has to do with application of big data to risk management activities suggested by both Dvorski Lacković et al., (2020) and Rosman (2009). This framework shows the incorporation of big data in risk management that is expected to result in advantages such as decreased response time to newly identified risks faced by banks. The aforementioned objective is attained by means of a direct identification of potential risks, a real-time assessment of their likelihood and impact, a daily monitoring of the extent of exposure to risks, and an improved level of transparency in the process of reporting.

4. The last stage shows the outcomes and benefits post processing the data.

The utilization of both internal and external data sources can result in a significant increase in value, leading to a synergistic learning effect (Sagiroglu, 2013). The use of big data derived from external sources can furnish vital non-financial corroboration that can be employed to evaluate financial statements. For example, external data is useful in instances where a product garners negative feedback on social media platforms, despite an increase in sales as reflected in financial records, it may be prudent to conduct a more thorough inquiry (Balios et al., 2020).

Data from internal and external sources is fed into big data technologies and advanced analytical tools that are utilized by business professionals to facilitate effective decision-making in the implementation of strategies. The processing of big data involves the utilization of several key technologies. Frequently used storage tools encompass Teradata, Hadoop, Amazon RDS, and Pentaho and data mining tools such as Python, IBM Cognos, Apache, R, and Google Big Query can be used, as well as analytics tools such as Apache Spark, Tableau, and SAS Visual Analytics (Maja & Letaba, 2022). These technologies collaborate to achieve a shared objective of extracting value from data that was previously considered obsolete (Zakir et al., 2015).

Automated tools collect and normalize diverse data, process real-time data, analyse high-value systems, and implement advanced monitoring systems, with active controls for user authentication, data transfers, and decision-making (Sagiroglu, 2013). The data sources go through a process of competitive intelligence that is an iterative procedure of converting data into valuable and profitable knowledge (Adil & Abdelhadi, 2022). The process entails converting information into implicit knowledge that can be utilized by decision-makers to implement strategic actions in response to alterations in the environment (Adil & Abdelhadi, 2022).

These big data analytics tools are applied to risk management data and activities in order to draw meaningful insights, enhance decision making as well as aid in increased compliance. This helps with enhanced decision-making by providing valuable insights, generating new business models, products, and services, creating knowledge, implementing management strategies, innovating products, and services, enabling dynamic market analysis, and assisting managers in decision-making through customized dashboards (Adil & Abdelhadi, 2022).

This conceptual framework helps underpin where challenges appear in the data lifecycle and risk management activities, and possible recommendations, risk specialists will know where most challenges are experienced in the process of big data application.

2.6 Conclusion of Literature Review

In this chapter, the literature pertaining to impediments to big data usage was analysed. Risk management for banks was discussed, as well as the role it plays in banking. It was deduced that most of the literature focuses on big data usage in fraud detection, market risk, and credit risk, with limited non-financial and operational risk. Further to that, big data concepts were introduced, the application thereof analysed to establish applicability in banking risk management. Different technologies and analytics methods used in big data were also explored. Thereafter, the concept of data driven decision making was introduced as the main benefit of using BDA, and the gaps found when linking big data usage to risk management.

BDA has gained significant attention in recent years, but its application in risk management practices is underexplored, with most studies focusing on algorithms and big data technologies. Despite extensive research in healthcare and marketing, there is a lack of rigorous and large-scale studies on the application of BDA in banking, particularly risk management. Most studies on big data and big data analytics applications are qualitative and mainly look at literature and case studies. Dvorski Lackoviae et al., (2016) state that organizations should incorporate big data analytics into their processes for decision-making, but there is a significant gap in literature on how the banking industry is leveraging big data potential to curb risks and losses.

Although the literature is not extensive for this topic, a relationship between the two concepts was established. As mentioned by Dvorski Lackoviæ et al., (2016), big data technology is one of the most discussed topics in the technology landscape and research areas and has been mainly used by departments such as marketing and sales, however its potential for risk management does not seem to be gaining much needed traction. Big data usage can be a powerful tool in today's banking risk management, however, to understand that power and contribution, it is imperative to understand the challenges as they pertain to risk management.

The theoretical frameworks were explored looking at the role of big data in risk management in banks. A framework by Rosman (2009) investigates the correlation between risk management procedures and the four fundamental components of the risk management process, highlighting a positive correlation between risk management practices and the process aspect. A framework by Dvorski Lacković et al., (2020) was adopted as it links big data and risk management, demonstrating the application thereof. Despite some efforts to evaluate and develop a theoretical framework for deriving organizational value from big data Analytics (BDA), these endeavours have predominantly adopted a limited viewpoint cantered on information systems and technology. Lastly, a conceptual framework was established. This conceptual model is based on literature and theoretical frameworks, offering an explanatory approach to the research problem. It explores how big data is applied in risk management from source, helping underpin the challenges experienced at various stages of the risk management framework.

The literature highlights a gap in understanding why the banking industry hasn't fully utilized big data for risk management. This study attempts to bridge this gap by investigating why big data is not widely used in risk management practises.

CHAPTER 3. RESEARCH METHODOLOGY

3.1 Introduction

The purpose of this chapter is to describe the methodology that was chosen and used in this study. Based on the literature studied in chapter 2, qualitative methodology was determined to be the best fit for answering the primary and secondary research questions formulated post concluding the literature review.

3.2 Research approach

According to Öhman (2005), a qualitative research methodology is a methodology that is focused on exploring individuals' own experiences, presented in thoughts, ideas, feelings, attitudes, and perceptions (Öhman, 2005). It is commonly used to dive deeper and get the quality of the subject as opposed to the quantity (Öhman, 2005). Qualitative research examines subjects in their natural environment to understand their meaning and the interpretations attached by individuals or groups (Hammarberg et al., 2016). It is said that qualitative research was less popular in past decade, and as a result it became appreciated for its ability to dive deeper into a subject with a much richer interpretation (Njie & Asimiran, 2014). This helps develop new knowledge that is not based on a predetermined hypothesis when compared to a quantitative method. In that regard, the use of qualitative methodology for this research study is the most suitable method as limited research has been conducted on the subject.

Öhman (2005) mentioned that the advantages of using qualitative methods are that they provide a detailed account of the participant's thoughts about the subject, their opinions, and experiences. He further mentioned that they allow for better insight and depth into the subject matter that has not been broken down to a desired level, with limited pre-existing knowledge.

This method was chosen as it was deemed the most suitable for conducting a deep dive and answering the research questions presented. This is because the topic of big data usage has not been widely researched and is lacking in literature, and a more in-depth knowledge from subject matter expert is required. This study collected individual perceptions of what role big data plays in risk management, while understanding why it has not yet been employed in the underlying risk processes. As can be seen in the literature review, the knowledge base in this subject is very limited and still needs to be explored further to deduce more theories and understanding. Using this method assisted with further establishing the real relationship of the two phenomena. This method also helped the participants introspect on the subject and sparked a discussion that may help the participants leverage of existing technologies, instead of continuing a perpetual chase on traditional methods.

This type of research approach was selected based on the assumption that subject matter experts that were part of the study have sufficient knowledge and are fully equipped to respond to the research questions posed. It was assumed that this approach is best to garner more in-depth responses to the questions and help gain a deeper understanding of the problem.

3.3 Research design

The research design that was used is a case study. A case study is a research design in which the researcher delves deeply into a program, an event, a process, or one or more individuals (Creswell, 2018). This design encourages deeper investigation into risk management processes in a way that is not influenced by prior knowledge, assisting the study in unearthing concepts and theories that may have received little to no research attention in the past. Because the purpose of this research was to achieve new insights into challenges faced with big data usage and adoption, the use of a case study was deemed more appropriate design, helping narrow down into one bank as opposed to many since there are few earlier studies as can be seen with the literature review concluded. A case study was also deemed fit due to timing constraints of carrying out the research.

3.4 Data collection methods

This study was primary research and gathered information directly from the subjects. This was because of a need to carry out an in-depth study to answer the posed research questions. Data collection for this study was conducted using two types of research methods. The first method used was semi-structured interviews and the second was focus groups, responding to a list of questions that were designed to achieve specified research objectives. This was to get a true sense of the environment, based on the population types interviewed.

I. Interviews

Interviews can be characterized as a social interaction where individuals engage in a collaborative manner to generate accounts or narratives pertaining to their past or future actions, experiences, emotions, and thoughts (Seale et al., 2004). Interviews as a data collection method were selected because they help draw deeper insights from subject matter experts and allow for probing for clarity and further information. The questions are open-ended, which allows the interviewer to add new questions during the interview and restructure based on the responses of the participants.

II. Focus group

Powell & Single (1996) define a focus group as a group of individuals selected and assembled by a researcher to discuss and comment on, from personal experience, the topic that is the subject of the research. The benefit of using focus groups is that unlike in individual interviews where it is easy for the researcher to control the interview, in a focus group the shape of the discussion can be controlled or shaped by the participants, helping the researcher to gather a larger amount of information in a shorter period. Particularly in the interest of this research, focus groups were used because of the seniority differences between the different participants, which, some were decision-makers. This approach helped establish if there was a degree of consensus on the topic between the heads of risks and the more specialist/ analyst roles.

3.5 Population and sample

3.5.1 *Population and case site*

The bank chosen for this case study is a commercial bank that was established in 1991, based in Johannesburg, South Africa. The bank is made up of over sixty thousand employees with branches in various African countries as well as the United States. As a provider of financial services, it operates in segments such as personal and business banking, corporate and investment banking, wealth and investment management, and bank assurance. It forms part of the four major banks in South Africa as stipulated in the Statistica report (Statista, 2022), classified according to headline earnings and asset value. Within the institution, there are various support functions such as the risk management office, IT development office, human resources, and finance.

The target population only included risk management professionals across different risk types within the bank. This demographic was chosen based on the study focusing on risk management, and risk specialists and managers being deemed the correct target population. The study also focuses on banks and the financial sector, so it was fit for the population to be employed in the banking sector. Because the study also focuses on big data, it was deemed fit that the population must exist in an environment where there are large systems and large data sets are collected.

3.5.2 *Sample*

For this study, the sample was made up of a total of twenty-three (23) participants, five (5) of whom were heads of risk across risk types that exist within the bank and 18 risk specialists. According to Dworkin (2012), the recommended number of participants for a qualitative study is anywhere between five (5) and fifty (50) participants. Furthermore, Guest and Johnson (2006) recommend a sample size of twelve (12) as appropriate for studies aimed at understanding the experiences of a homogeneous group. Therefore, the sample size of twenty-three

(23) participants was adequate in this regard to achieve the research objectives. The final number of participants was determined by saturation, as similar ideas, themes, and concepts were starting to come up repeatedly. Saturation is defined as the point at which the data collection process no longer provides any new or relevant data (Dworkin, 2012).

There were four focus groups and five individual interviews making up a total of nine (9) interview sessions. Table 4 below shows the composition of the sample: the number of participants per risk type in the interview sessions. The sample was made up of one (1) financial and three (3) nonfinancial risk types. This was to ensure coverage of all risk types and to get a sense of whether they share similar or different experiences on big data usage.

Risk Type	**No of Participants**
Operational risk	10
Technology Risk	6
Credit risk	5
Enterprise risk	2
Total number of participants	**23**

Table 2: Sample Distribution by Risk Type

3.5.3 *Sampling method*

The sample for this study was selected using the non-probability sampling method, purposive sampling. The purposeful or purposive sampling method is used to uncover detailed information about the issue under investigation (Jahja et al., 2021). It entails identifying and selecting individuals or groups of individuals who are knowledgeable about a phenomenon of interest (Etikan, 2016). In this regard, the sample is chosen by the researcher with the understanding that participants can contribute knowledge that can be used to answer research questions (Jahja et al., 2021). This method was chosen because the aim of the study is not to make a statistical inference and rather to gain a deeper knowledge and understanding of the impediments of big data usage in banking risk management. Although this study will allow for a subjective selection of participants, it will allow for subject matter experts to be chosen, supporting the collection of suitable data that will aid in answering the research questions posed.

3.6 The research instruments

Semi-structured, virtual interviews were used as a data collection tool for both individual and focus group sessions. Interviews allow for the collection of data from participants while also allowing for the emergence of related issues, information, or new insights (Jahja et al., 2021).

An interview guide attached in Appendix F was used and consisted of a total of sixteen (16) questions, four (4) of which were demographic-type questions. The set of questions was designed in such a way that answers each research question.

The first three interview guide questions addressed the first research question: RQ1- What is the perceived role of big data in risk management? The subsequent nine questions answered the last two research questions namely RQ2- What hinders the usage and adoption of big data technologies in risk management practices and RQ3- What challenges are experienced when using big data in risk management? The questions for last two research questions did not always follow chronological order and were based on how the participants answered and the follow-up question required for that answer. The interview guide was mainly used as a guide to ensure consistency, but participants were allowed to respond as they deemed fit.

3.7 Procedure for data collection

Prior to interviews and focus groups being set up, a letter of consent and permission to conduct research within the organization was obtained from the people risk representative and a compliance officer for the group. Furthermore, an attestation to comply with the group compliance policy was completed.

In the identification phase, the participants were identified by first reading the enterprise risk framework to see the list of risk types within the bank. A list of the risk types was extracted and using the combination of the internal organization structure and directory, a list of heads of the risk types as well as their reporting lines was compiled. It was then decided that the interviews will be divided into two (2) groups- individual interviews and Focus groups. Individual interviews were conducted with heads of risk and focus groups were conducted with risk specialists. This was to get views from both strategic and operational lenses and so that the less senior individuals are not influenced by the ideas and thoughts of the heads, as that can be the case with focus groups.

According to Powell & Single (1996), a focus group should be made up of individuals who share essential qualities relevant to the study, that are strangers to each other so that they are not inhibited or deferential to intra-group differences (such as occupational seniority). The focus group participants were selected based on being risk specialists and/or reporting to the selected heads of risk. They were then grouped together to form a focus group according to the risk type they fall in and not based on the department they work for. This created homogeneous groups per risk type, and within each focus group, created a different mix of departments.

The participants were contacted before the actual interviews and focus groups to establish their interest in participating. The individual participants were approached using email and organizational Microsoft Teams chat. Upon agreement to participate, meetings were set for an hour, to take place online through Microsoft teams. Jacob & Furgerson (2012) recommend that an hour is a reasonable time for an interview. This is to improve willingness to participate and to ensure engagement so that quality insights can be extracted. In the meeting invite, each of the participants was issued a consent letter including the background and objectives of the research, as well as proof of permission to conduct research within the bank. This ensured they have adequate information about the research and allowed time to decline should they not be comfortable participating. The participants were not issued with the research questions beforehand; this was to avoid them researching the topic and interfering with their true experience of big data in the organization.

Due to the “work from home” model in the organization introduced during the Covid-19 pandemic, it was deemed fit to continue the interviews online to encourage ease of participation, as face-to-face interviews are only considered under extreme conditions, and are unculture-like for the organization.

The interviews were self-administered, and the researcher facilitated the focus group discussions. An interview guide was used to facilitate the individual interviews and focus groups. The interview guide included a script for the interview's opening and closing, as recommended by Jacob and Furgerson (2012). This helped to reinforce the research objective and gave structure to the discussions. The confidentiality clause was read to the participants, and they were also given the opportunity to ask questions before the formal interview questions began and permission to record the session using the Microsoft Teams feature was obtained. The participants were reminded that the discussions are based on their personal experiences and opinions about big data. Thereafter, the discussions began with introductions, which were helpful in gathering demographic data.

Following the completion of the introductions, the interview guide moved on to questions directly related to the research questions. The questions were open-ended, allowing participants to provide additional information beyond what was requested. The order of the questions was followed where possible, but the questions were also asked based on the answers provided to allow for further probing. Some participants answered some of the questions before they were asked, so the questions were either skipped or further probing was done to get them to elaborate on their responses. In focus groups, participants were given the liberty to shape the discussions, with the researcher guiding the discussions with the interview guide to ensure that there is no deviation from the topic. Discussion and engagement amongst participants were encouraged using follow-up probing questions and clarity-seeking questions.

Of the total twenty-three (23) participants, five (5) were individually interviewed and eighteen (18) formed part of various focus groups. Some of the interviews did not use the entire hour, whereas some went overtime and were accommodated as such. The below table shows the duration of each interview session:

Interview Type	Focus Group Name/ Head of Risk Job Title	Duration
Focus group	Operational risk	1hour 25 seconds
Focus group	Tech and resilience	52m 2 seconds
Focus group	Retail credit risk	49m 32 seconds
Focus group	Specialized risk	1hours 3min 45 seconds
Individual	Head Credit risk	26m 17 seconds
Individual	Head Data Management risk	36m 4seconds
Individual	Head Enterprise risk	33m 46s
Individual	Head Risk Measurement	36m 28 seconds
Individual	Head Retail Operational risk	52m 26seconds

Table 3: Interview Duration

The longest individual interview recorded was 52 minutes 26 seconds, whereas the shortest recorded was 26 minutes 17 seconds. The overall average duration of the individual interviews was 37 minutes. The longest focus group discussion was 1 hour and 3 minutes, whereas the shortest was 49 minutes, and 32 seconds, with all focus group sessions totalling an average of 56 minutes.

3.8 Data analysis strategies and interpretation

A thematic analysis approach was followed to analyse data gathered from this study. Thematic analysis is a qualitative analysis used to analyse classifications and present themes (patterns) that relate to the data and is thought to be the best method for any study that seeks to discover meaning through interpretations (Ibrahim, 2012). This method was suitable for this study because it provided the opportunity to understand the responses more deeply, drawing necessary relationships, linking various concepts, and drawing thematic conclusions

(Ibrahim, 2012). Detailed steps on how the data was analysed can be found in chapter 4.

3.9 Possible limitations and challenges of the study

The weaknesses in the study were identified as follows:

I. Lack of full disclosure of bank big data and risk strategies due to policy restrictions
II. Lack of generalisation as the study was a case study and may not represent industry.
III. The methodology could have allowed for participants to provide information that does not benefit the study thereby skewing the result of the study.
IV. The lack of thorough understanding of big data as it is a subject that sits outside of risk management functions might have hindered the absolute quality of responses in some cases.
V. A qualitative study limits the number of participants and therefore creates a more subjective view of the subjects, as the results may not apply to other banks.
VI. The study is limited in geography and excludes the financial institution's branches that are in other countries, thereby limiting generalisation.

3.10 Quality Assurance

The quality assurance of this study was done based on recommendations by O'Brien et al., (2014). To ensure, the participants selection was on their competency to answer the posed research questions as it relates to risk management, therefore only seasoned risk specialists where selected. Secondly, the first interview was also used as trial to test the quality of the questions. This trial informed the correction of the interview process where required. To improve the trustworthiness and validity of the data collected, a semi-structured interview guide was followed during the interviews with no deviations except to further probe on the answers. The audit trail of interview recordings and transcriptions was kept and stored in a computer that only the researcher had access to.

3.11 Ethical considerations

The researcher went through an ethics workshop course and met all requirements by obtaining an ethics certificate of competence from the university. Post ethical clearance, the researcher contacted the financial institution's compliance and people risk departments to get permission to conduct the study as well as complete an assessment to ensure the study had no impact on the employees. Furthermore, the letter of consent and ethics clearance were attached to interview invitation, and invitees advised to decline the invitation should they not consent to be interviewed. During the process of the interviews, strict POPI act (protection of personal information) guidelines were followed as well as the confidentiality guidelines of the financial institution. Personal information was not collected beyond the names, tenure, and role of the individuals. Information that had no bearing on the outcome of the study or that did not help answer the research question was not collected. To ensure there is transparency between the researcher and participants, confidentiality and privacy clauses were reiterated before the interview could commence and participants were given an opportunity to ask for clarity or opt out of the interview. Raw data collected from the study was stored in a secure manner (in an encrypted laptop hard drive that required a password to access) and only used by the researcher for the intended purpose of the study.

CHAPTER 4. PRESENTATION OF FINDINGS

4.1 Introduction

The primary objective of this study was to investigate the extent of the usage of big data technologies in spotting, analysing, and managing risks in a banking environment. According to the research and the conceptual model presented for this study, there is potential for big data to be used in risk management processes where big data techniques can be applied to risk data as part of the risk management capabilities. Thematic analysis was employed when analysing results for this study as it is thought to be the best choice for any study that attempts to uncover phenomenon through interpretations as well as adding a methodical component to the data analysis (Alhojailan, 2012). The study was done inductively to allow for broader generalizations, and to ensure that the themes are properly related to the data (Alhojailan, 2012).

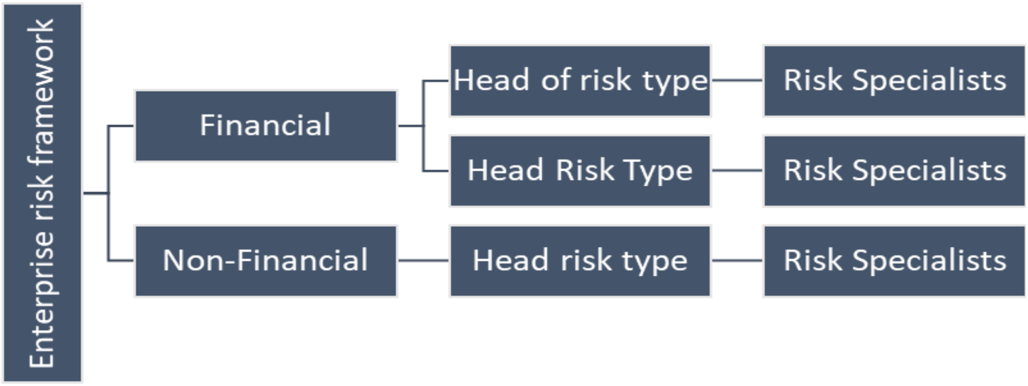
The purpose of this chapter is to summarize research findings and as such, the results presented will be according to the sections presented in the interview guide. The first section of the interview guide was introductions and the collection of demographic data whereas section 2 was used to answer the study research questions. This chapter will at first present demographics followed by the analysis approach and the research question specific results.

4.2 Demographic analysis

The first section of the interview questions dealt with demographics and getting participants' views on their understanding of big data, as well as their experience using it.

4.2.1 *Demographic profile of the sample*

The sample for this study was chosen via purposeful non-probability sampling, with the selection based on the financial institution's enterprise risk framework that is used to identify the important risks, known as principal risks, and related sub risks. According to this framework, principal risk types are risks that are recognised for their significance to the group's business strategy and as such informs the risk operating model. The operating model and team structures was used in identifying participants as depicted below:



As the study was focused on the challenges of using big data within risk management, the target population only included risk management professionals. There was a total of twenty-three (23) participants in the study, sixteen (16) of whom were female, and seven (7) males as depicted below:

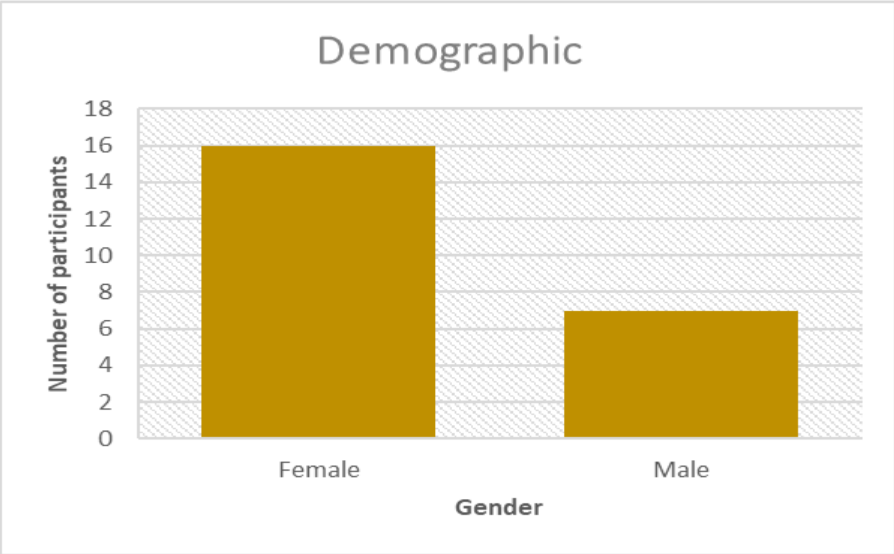


Figure 9: Demographic profile

As previously indicated, the focus group participants had a risk type in common but differed in the departments they work for. Below, it can be seen how many males and females were in each group as well as the different types of departments that were represented in each group. The largest group was the operational risk group, and the smallest was the technology and resilience risk group. The specialised risk group was made up of three risk types as can be seen in the "Composition by Departments". This group came about because of the perceived individuals' advanced knowledge and experience of the subject compared to other groups. The following table shows the composition of each interview by gender and department:

Focus Group Name/ Head of Risk Job Title	Composition by gender	Composition by Departments
Operational risk	1 Male, 6 Female	Africa region, Group, Resilience risk (Physical security and Premises), Product solutions Cluster
Technology risk	2 Males, 1 Female	Cyber security, Tech functions
Retail credit risk	1 Male, 3 Female	Product solutions Cluster, everyday banking Cluster

Specialized risk	2 Male, 2 Female	Security and Forensics, legal risk, Enterprise risk (Risk strategy)
Head Credit risk	Female	Everyday banking Cluster
Head data management risk	Female	Group Information technology
Head enterprise risk reporting	Female	Enterprise risk (Group risk strategy)
Head Risk measurement	Male	Group Operational risk
Head Retail operational risk	Female	Product solutions Cluster

Table 4: Sample Distribution by Gender and Department

The table below shows the composition of the sample by financial institution tenure, whether they have direct report and whether they are in a leadership role. One of the heads of risk (head of data management risk) does not have direct reports and one of the focus groups (Specialised risk) did include someone in the leadership role with direct reports and was added as part of a focus group as their homogeneity with the specialists boosted the flavour of the discussion. It is also to be noted that, in some structures, although someone may be a specialist, they can also be a team leader and will therefore have direct reports, although seniority will differ to the heads of risk. An average of tenure is provided for the focus groups below:

Interview Type	Focus Group Name/ Head of Risk Job Title	Financial institution Tenure	Direct Reports	Leader Role
Focus group	Operational risk	2 months ~ 24 years	No	No
Focus group	Tech and resilience	4 ~ 10 years	No	No
Focus group	Retail credit risk	8~13 years	Yes	No
Focus group	Specialized risk	7 ~ 11 years	Yes	Yes (one of four)
Individual	Head Credit risk	5 years	Yes	Yes

Individual	Head Data Management risk	14 years	No	Yes
Individual	Head Enterprise risk	10 years	Yes	Yes
Individual	Head Risk Measurement	11 years	Yes	Yes
Individual	Head Retail Operational risk	5 months	Yes	Yes

Table 5: Tenure

Below is a graphical depiction of the composition of participants per risk type. This is used for the purpose of providing context for the analysis pertaining to the exposure to big data usage between the two groups. As can be seen in Figure 10 below, five (5) of the participants were from the financial risk types, whereas the rest of the participants came from the non-financial risk type background.

Figure 10: Participants by risk type

4.2.2 *Big data understanding*

Participants were asked what their understanding of big data is. This question was important to create context for the interview. It also allowed the participants to start thinking about the application in their current roles and if at all it is a foreign subject to them. According to D'Ignazio & Bhargava (2017), the common misunderstanding of what big data means has resulted in a situation in which individuals are habitually placed in situations where the outputs of big data initiatives do not serve them. So, this also helped answer the bigger question of whether big data is used at all in risk management i.e., the more people are not familiar with the term, or only identifying from a theoretical point of view gave an indication of the level of application in their departments, as well as guide how subsequent questions will be asked.

The following word bubble shows some of the words used by the participants when describing big data:

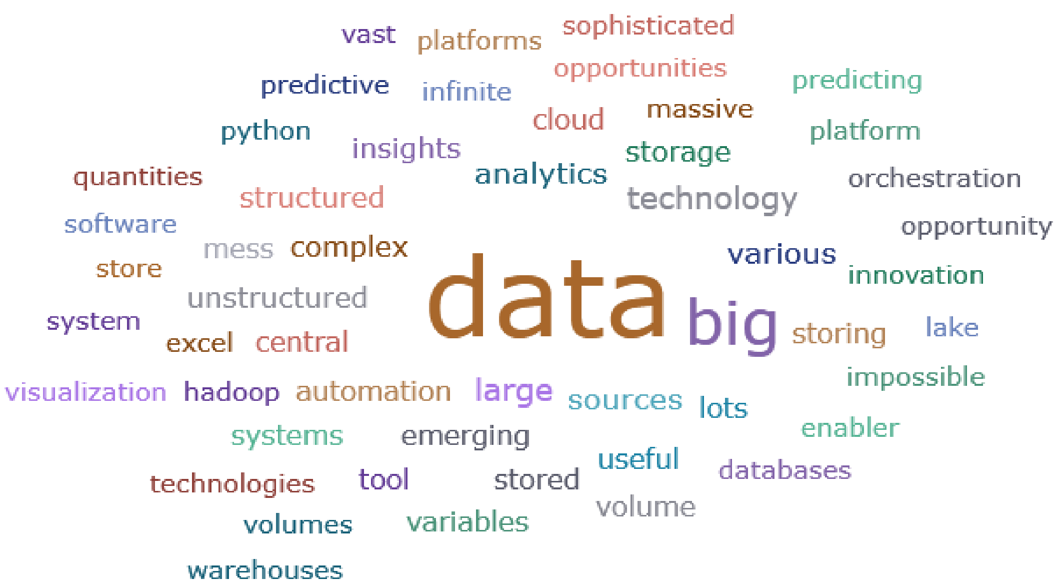
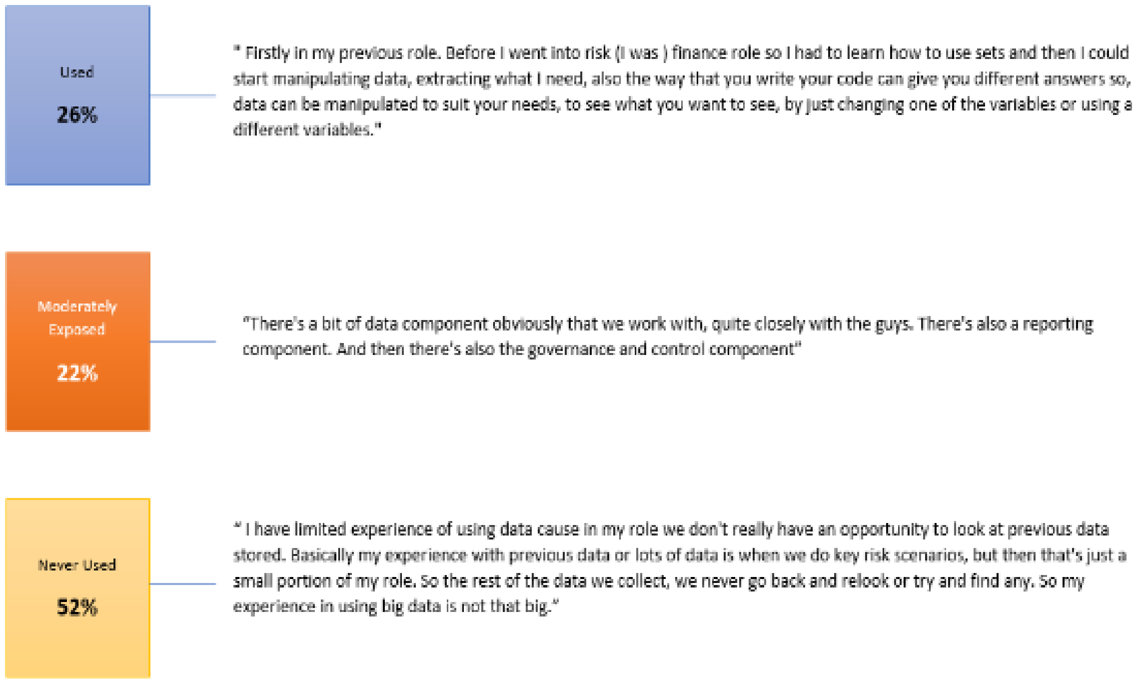


Figure 11: Big data Definition

One of the participants described big data saying "The volume itself means that you cannot study (it), using Excel to analyse it, for example, it requires more sophisticated tools for you to analyse that data. And I think one of the most recent tools that have blasted into the buckets that relate to analysing the data is Python"; while another said, "My understanding of big data, firstly, vast amounts of data and it probably has an element of unstructured mess sitting in it".

The subsequent question was "What is your personal experience with using big data? This was a follow-up question to probe into the extent of big data knowledge. Some of the specialized risk focus group participants did have some big data experience, and one of them gave an example of how they use big data. He mentioned that “We use the data to help other business units make informed decisions. So, you might see some emails flying around, please declare the outside business affiliations that you have. You know that data is driven by the data sets that we have”.



However, most of the participants responded negatively indicating that they have never used or been exposed to big data before, whereas a smaller number of participants had been moderately exposed as can be seen in the diagram above.

4.3 Risk Management and Big Data

The second section of the interview dealt with answering the research questions as it pertains to the research topic and the aim of the study. The first part is the procedure that was taken to analyse the data, followed by presentation of the findings as per the research objectives/aims. Data analysis is the process of making sense of text and visual data. It entails preparing the data, delving deeper into the data's comprehension, depicting the data in a format that makes sense and forming an interpretation of the greater meaning (Creswell, 2018).

4.3.1 *Analysis approach*

The data collected in this study was analysed using thematic analysis, a type of qualitative analysis described by Braun & Clarke (2008) and according to Miles & Huberman's (1994) model for the thematic analysis process. According to Braun & Clarke (2006), thematic analysis is a method used to identify, analyse, and report on patterns or themes within data. Thematic Analysis was adapted for this study because it gives the researcher an opportunity to understand the potential of any issue more widely (Alhojailan, 2012). Thematic analysis allows one to connect the numerous thoughts and opinions from the study and compare them to the facts obtained in different situations at different periods throughout a project (Ibrahim, 2012).

There are two types of approaches within thematic analysis (Inductive and theoretical thematic analysis) (Braun & Clarke, 2006). For the purpose of this study, inductive analysis was selected. This is because of the interest to let data speak for itself as opposed to fitting it into an existing theory. The themes identified by an inductive approach are strongly related to the data itself. According to Thomas (2003), the primary objective for following an inductive approach is to (1) succinctly summarize extensive and varied raw textual data; (2) establish explicit connections between the research objectives and the summary findings obtained from the raw data; and (3) develop a model or hypothesis that elucidates the fundamental structure of the experiences or processes observed in the raw data.

The Miles and Huberman (1994) model for analysis is depicted in the diagram below. It consists of four interconnected stages: data gathering, data reduction, data display, and data conclusion-drawing/verification. This diagram shows that the analysis did not follow a linear process but was an iterative process involving going back and forth between different stages of the thematic analysis.

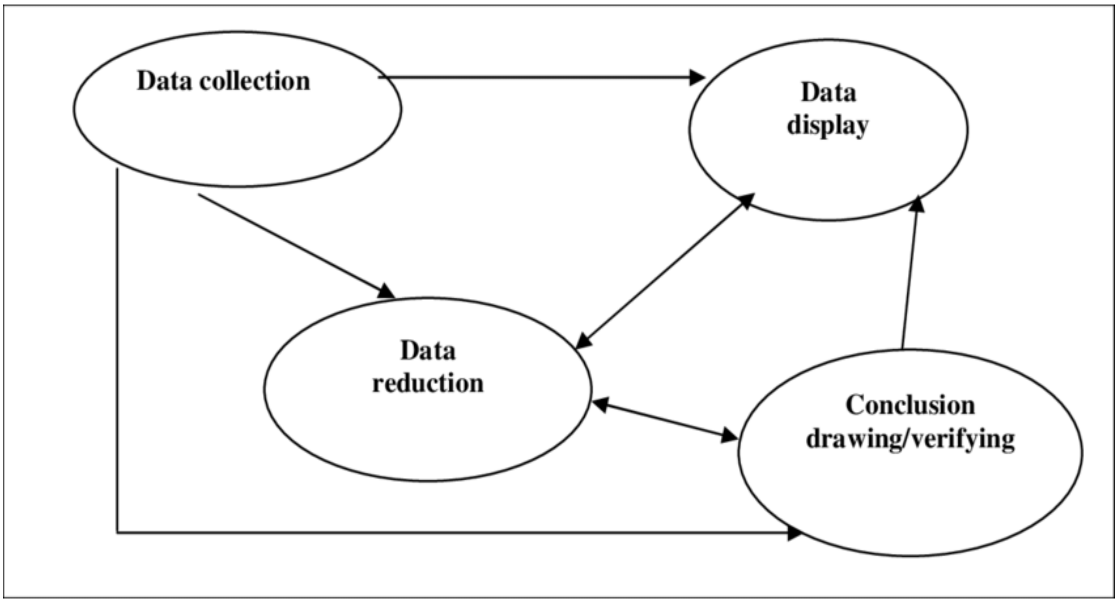


Figure 12: Myles Hubert model for data analysis (Hasan & Crawford, 2006)

The following describes the thematic analysis stages followed during the analysis:

Stage 1 – Familiarising with the data:

After the data was collected through interviews, as well as transcribed and recorded using MS Teams' software, the transcribed data was downloaded and viewed using Microsoft Word. For validity, the researcher relistened to the recordings to verify the accuracy of the transcripts and ensure nothing of importance was left out. Post verifying the contents, the focus groups transcripts and individual interview transcripts were stored in separate folders. Then proceeded to re-read the transcripts, taking them through a process of data reduction, removing all unnecessary words, replacing personal identifiers, and removing remarks and conversations unrelated to the study; this included filler words, replacing real people's names with speaker and participant numbers, as well as ensuring the transcripts read well. Then the next step was to manually highlight sections to 1) separate questions from answers and 2) highlight sections with potential ideas and thoughts that can help answer the research questions. The researcher did not attempt to analyse the data during the interviewing phase, therefore analysis occurred only after all interviews were concluded.

Stage 2: Creation of Codes

In the second stage, the soft copies of the transcripts were uploaded to Atlas.ti software. According to Alhojailan (2012), using software for qualitative data analysis is beneficial in terms of strengthening the rigor of the analytical procedures for validating that which does not match the researcher's impressions of the data (Braun & Clarke, 2006).

The documents were grouped in software folders according to focus groups and individual interviews. Subsequently, the documents were coded per group using the software AI coding feature. The AI feature returned a total of 676 codes. The researcher then went through each code and the quotations to verify each code for accuracy. Each code was then reviewed for relevance and additional manual codes were added through the review. Thematic codes where then created according to research questions and used as code categories. This was to help with grouping related codes and quotations, and ultimately deducing themes. A total of 20 manual codes were created out of reviewing individual quotations and labelled according to what they spoke about in relation to the research question, bringing all codes to a total of 696. The coding process was iterative since transcripts were evaluated multiple times as new codes developed or existing ones evolved.

The next step taken was categorisation where codes were reviewed and categorised in order of importance and relevance (Öhman, 2005) and further grouped sub-categories where required. The code categories that came about were "uses in risk", consisting of a group of pre created codes and labelled quotations that mentioned how big data can be applied in risk management. The second category was "challenges", taking codes and quotations where participants were answering questions relating to challenges. The last category was "barriers to adoption", which was derived from both challenges as well as answers relating to adoption. This was created with the potential to answer research question 3. The diagram below shows a summary of the coding process:

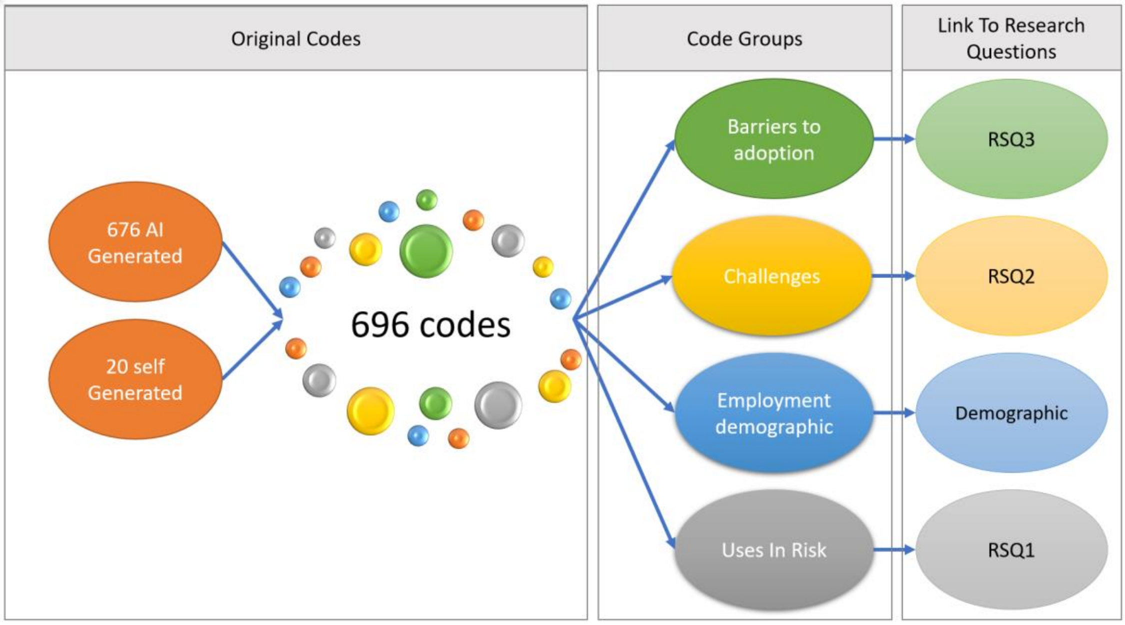


Figure 13: Codes

All data relating to each question was organized and presented in chronological order. By placing the data into conceptual clusters for analysis, the researcher was able to investigate any differences, similarities, and interrelationships to make a basis for plausible explanations. This aided with identifying how the categories related to each other.

Stage 3: Searching for Themes

A theme represents some level of structured response or meaning within the data set and captures something relevant about the data in respect to the research question (Braun & Clarke, 2006). The themes for the study were not predetermined from the theory in chapter 2. This was because the researcher decided to follow an inductive approach of thematic analysis, and to let the data speak for itself. The theory in this case was only used to guide the interview and research questions.

To get to themes, the researcher reviewed codes within each code grouping. Within each grouping was a number of codes as depicted in Figure 14 below.

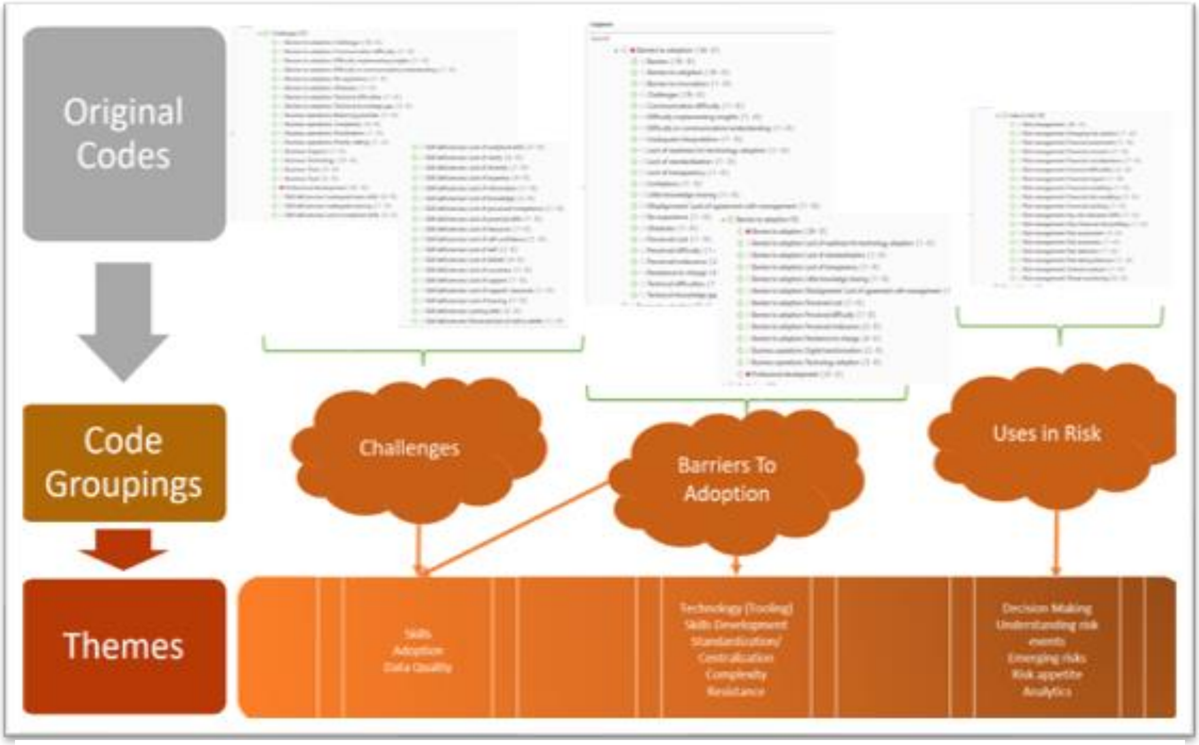


Figure 14: Deriving Themes

The codes within each group were further clustered together according to the similarities as shown in Figure 15 below. Since quotations were coded individually, they allowed for deeper studying of the interview segments, helping with the creation of linkages between common codes, and the identification of codes that came up the most.

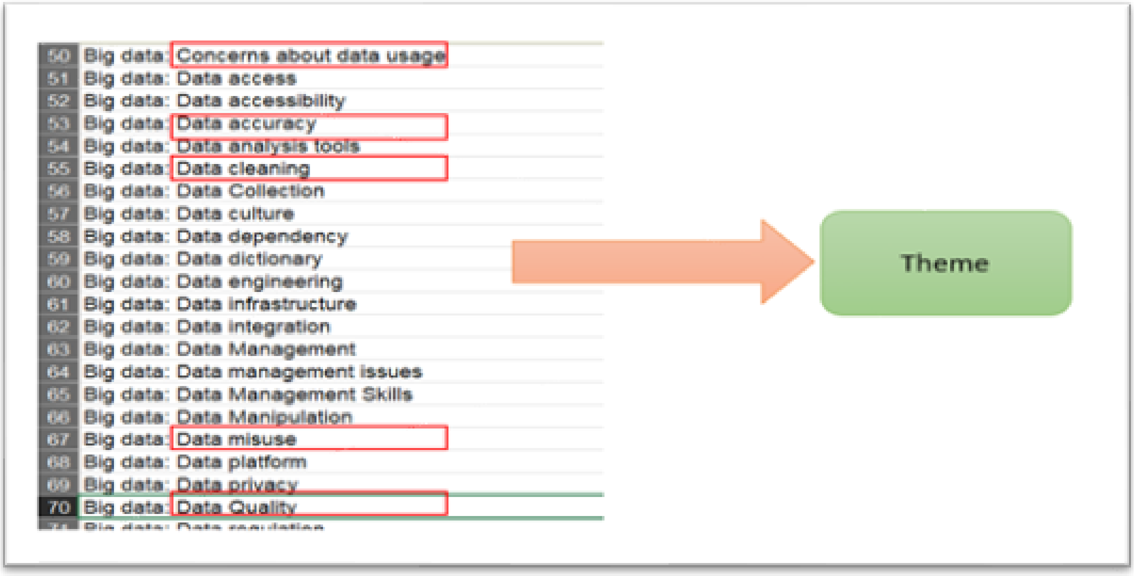


Figure 15: Grouping Themes

Stages 4 – Reviewing Themes

The emergent themes were reviewed, and similar ones were combined once again to come up with a final list of themes as depicted in Figure 15 above. Because of the linkages between research question 2 and 3, both challenges and barriers to adoption categories were used interchangeably to ensure all codes were picked up and themes are thorough and a true representation of the data. Post coding and identifying initial themes, data was transferred to Excel, where codes that shared some characteristics or referred to related concepts were organized into final themes. Major themes emerged because of thorough and systematic reading and categorization of the transcripts, noting quotations that can be useful as evidence when presenting results.

Stages 5 – Defining and naming themes

The themes were labelled and named after they were generated, based in part on insights gleaned from the review of literature and the researcher's interpretation of the data in relation to the linked research and interview questions. Post giving codes their names, new codes were created to help capture and highlight the main themes, as well as highlight sub-themes as depicted in the table below:

Codes	Resultant Theme Names
Big data: Data Quality Business operations: Quality assurance Data analysis: Data quality Data analysis: Data quality concerns Data analysis: Data quality control Data Quality Data quality control Accuracy Big data: Data accuracy	Data Quality
Barriers to adoption: Communication difficulty Barriers to adoption: Difficulty implementing insights Barriers to adoption: Difficulty in communication/understanding Barriers to adoption: Inadequate interpretation Barriers to adoption: Lack of readiness for technology adoption Barriers to adoption: Lack of standardization Barriers to adoption: Lack of transparency Barriers to adoption: Limitations Barriers to adoption: little knowledge sharing Barriers to adoption: Misalignment/ Lack of agreement with management Barriers to adoption: No experience Barriers to adoption: Obstacles Barriers to adoption: Perceived cost Barriers to adoption: Perceived difficulty Barriers to adoption: Perceived irrelevance Barriers to adoption: Resistance to change	Adoption
Organizational effectiveness: Importance of upskilling Organizational effectiveness: Skill acquisition Professional development: Acknowledgement of skillset Professional development: Skill development Professional development: Skill gap Skill deficiencies: Inadequate team skills Skill deficiencies: Inadequate training Skill deficiencies: Lack of analytical skills Skill deficiencies: Lack of perceived competence Skill deficiencies: Lack of practical skills Skill deficiencies: Lack of resources Skills gap: Lack of data analysis resources Skills gap: Lack of digital literacy Skills gap: Lack of experience	Skill

Table 6: Themes

Stage 6: producing reports

From this, data was displayed using different visualization methods which were then exported from Atlast.ti and stored in the local drive. Data visualization organizes data and aids in the organization of concepts and thoughts (Thomas, 2003). The code book and quotation manager reports were also exported to Microsoft Excel Workbook with the responses of all research participants to each question displayed side by side for further analysis and presentation of results. This was done to make it easier to compare participants' responses to specific questions and themes. The researcher also used methods such as Microsoft excel pivot tables and charts to group and analyse codes and themes. The output of the themes and extracts of connected quotations created by the analytical procedures outlined above will be in the following sections.

4.3.2 *Results for Research Questions*

The results are presented according to the order of the research questions i.e. the first part looks at the application of big data in risk management (RSQ1: **What is the perceived role of big data analytics in improving risk identification, assessment processes and mitigation strategies within the South African bank?**), the second set of results are challenges of using big data in risk management as it pertains to research question 2 (RSQ2: **What are the key challenges faced by the bank in effectively implementing and utilizing big data technologies for risk management?**), and lastly, the results for research question 3 (RSQ3: **What factors hinder the adoption and integration of big data technologies and practices in the bank's risk management processes?**) presented as adoption of big data.

4.3.2.1. Application of Big Data in Risk Management

Interview questions 1 to 3 were concerned with addressing the research question as it relates to the role of big data in an organization and subsequently within risk management. The first interview question was used to create context as it allowed the participants to reflect on how big data can be used and if it is even deemed as something to care about. Question 2 prompted thinking around the application in risk management process and the third as a further probe into how it can be applied to the benefit of the participant's team (and role). This question helped the participant to reflect or prompt thinking about their own impediments in so far as the usage of big data is concerned.

All participants did agree that big data application is important in the bank with some of the examples made around usage in client profiling and product provisioning. One of the participants remarked, "I see big data. It's much broader than the risk thing. Just like what is already in the frameworks is the governance and the management of data that we are using".

The results provided a total of 23 individual codes that were grouped into the final five (5) themes as captured in Figure 16 below:

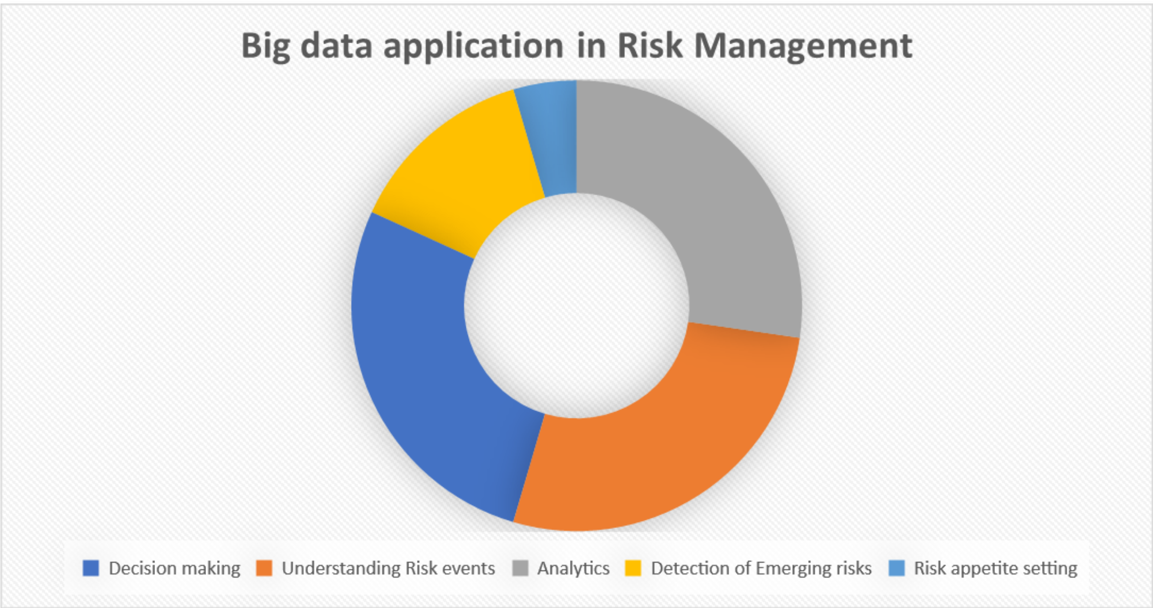


Figure 16: Big Data Application

I. Decision making

When asked how big data can be used in risk management, decision-making and understanding of risk events ranked the most in the themes.

Participant	Excerpt	Interpretation
Participant 14	"I think it's important for us to use big data in risk management. If used correctly I believe data can help you just in decision making. It can also help you in predicting events or emerging risks and so forth. And when used correctly and especially just from a cyber perspective, you always want to or need to try and be a step ahead of your threat in order to effectively protect the organization"	Predictive analytics Cyber risk trends
Participant 1	"You need data to understand what your boundary events are. Essentially what I'm trying to say is you need data to be able to drive any strategy within the bank, irrespective of whatever risk type. That is because you don't know if you're going to be making the right or wrong decisions"	Risk events strategy

Table 7: Decision Making

According to Elgendy et al., (2022), data-driven decision-making is a solution for providing more informed, quality decisions that combine human decision makers' intuition and experience with data analysis, resulting in more rational choices that lead to better results.

II. Understanding Risk events

Another way that big data can be applied to risk management was said to be in the process of understanding risk events that occur in the bank (the identification and analysis thereof) summarised in table 10 below:

Participant	Excerpt	Interpretation
Participant 5	“You could have a look at what is happening from the client side of things and then have a look at what is happening for example the risk event side of things. Is there a correlation? maybe there are more transaction volumes than we are experiencing a lot of process-related risk events. Or now we have the situation of load shedding in the country. What is that? How is that translating?”.	Client/risk relationship Risk classification

Table 8: Understanding risk Events

This demonstrates the claim that using big data for risk management allows banks to manage their risks at an early stage (Dvorski Lacković et al. 2020).

III. Analytics

Mentioned as frequent in the results was the use of big data for analytics. Big data analytics is the process that analyses big data sets using algorithms to extract diagrams, reports, and useful and unknown information (Dicuonzo et al., 2019). Table below captures a participant’s remark:

Participant	Excerpt	Interpretation
Participant 3	“Big data can be used to see how many clients do we have and what transactions are these clients performing and at what frequency? at what value in which locations during which times. “ “being able to analyse all of that can really set us up to be able to serve our clients better and be able to give them products that are going to	Predictive analysis Trend analysis Client journeys

	resonate with their needs, therefore meaning we create a long relationship with the clients and one that is also beneficial to the clients and it's not necessarily just a transactional banking account because as we know, human beings go through so many life stages. There are so many life activities that happen during one's lifetime and at any one of those points, there is a need for some financial interaction. And if we already have that light on board and we are harnessing the detail that we have, then we should be able to come up with some innovative products that can assist our clients through every life stage that they go through."	

Table 9: Analytics

IV. Detection of emerging risks

Another way big data can be used is in the identification of emerging risks. Identifying emerging risks can help banks mitigate risks before they materialize into an actual event. Table 12 below shows how participants remarked in relation to this:

Participant	Excerpt	Interpretation
Participant 3	"Looking for earlier indications of distress".	Early indicators, key risk indicators
Participant 1	"you'll always hear in a bank, they always talk about customers who are new to the bank or customers with spin scoring, meaning that you don't really have a	Application scoring

	Bureau profile. So, there are various ways that we could use big data to be able to enter those markets where we say we don't really know who (person) is on paper”,	

Table 10: Detection of emerging risks

V. Risk Appetite setting

Risk appetite scored the least as a component where big data can be applied, however is one of the significant processes in risk management. Risk appetite is defined as an investor's willingness to bear risk (Shen & Hu, 2007). According to Gontarek (2016), risk appetite is the written expression of the aggregate degree and categories of risk that a financial institution will take or avoid in order to meet its business objectives. To demonstrate the role big data can play in risk appetite, a participant’s remark is captured below:

Participant	Excerpt	Interpretation
Participant 1	“You can't set your risk appetite statement, for example, if you don't have data in terms of what your book looks like”.	Informed risk appetite setting Risk appetite and Balance sheet correlation

Table 11: Risk Appetite Setting

The results showed the following in relation to the risk framework model proposed by Dvorski Lacković et al., (2020) that underpins risk management theory, showing different risk management stages where big data can be applied. The following Pareto chart (Figure 17) highlights the results for each of the risk management activities in their order of frequency where similar codes were grouped together to come up with the frequency. The cumulative frequency line shows the sum of all codes in relation to the risk framework presented in the literature review.

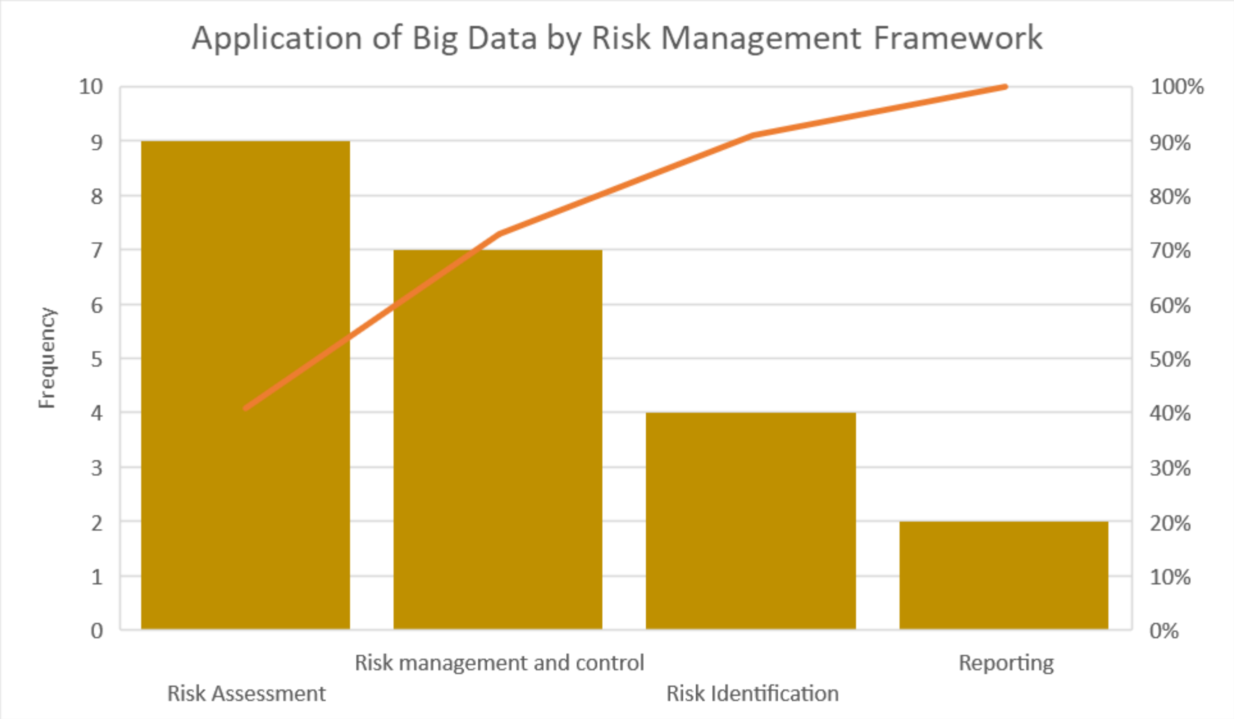


Figure 17: Pareto chart showing the application of big data by Risk Management framework activities

According to the data, risk assessment was identified as the biggest area where big data can be applied, with reporting scoring the least. Risk assessment includes activities such as data analysis through the calculation of various risk indicators, real-time simulation of risk indicators, and predictive analysis (Dvorski Lacković et al., 2020). The code groupings for risk management and control included the management of risk profiles with one participant remarking "(can be used) in understanding non-financial risk profiles, understanding what potential reputational impacts would take place because that's currently very much subjective." In relation to risk identification, codes such as emerging risk analysis and risk detection were grouped which is in line with what theory suggests. To demonstrate how data gets used for risk identification, particularly in more financial risk types, a participant remarked "Yeah, in a model risk area. They use all that historic big data to develop models and to backdate those models to see where the models are stored, appropriate, how to update them, and all that kind of stuff."

Many participants did not mention reporting when asked about the potential application of big data in risk management, however, someone remarked "If it's being used, it's been on a very minimum basis. It's mostly used just for reporting. Maybe they'll be used on some level to report for reporting purposes to gather stats and so forth, but not so much because even if you look at the issues that are there, there are so many issues that are currently open, issues that are overdue".

4.3.2.2. Challenges of using Big Data in Risk management

There were 9 questions created to answer research questions 2 and 3. The questions did not follow chronological order and depended on how participants answered subsequent questions and if the interview question had already been indirectly addressed. Because the interview style was in such a way that encouraged a discussion, some of the questions were created for probing purposes, and the interviewer did not always differentiate between research questions 2 and 3 as the two are interlinked from a discussion point of view.

This section serves as the anchor for the aim of the research as the results extracted here provide insight into why big data is not used extensively in risk management. It sought to establish what impedes risk management professionals from using big data, although, following research question 1, they did show an understanding of the importance of applying it for risk management. The data did indeed indicate that, although big data may be used in some risk types, especially financial risk types, there were challenges that prevented the greater risk community from fully embedding big data usage in their processes. Remarking on challenges faced, one participant summed it up by saying that "challenges are skill set, time for research, and the culture of not opening doors to fresh minds. We need to make an initial investment. There is currently no follow through and there's no commitment, Senior management buy-in, lack of financial support, no commitments, lack of driving innovation".

When coding, the results provided a total of 248 individual codes that were grouped into the final three themes, mainly data quality, adoption, and skill as seen in the pareto chart below:

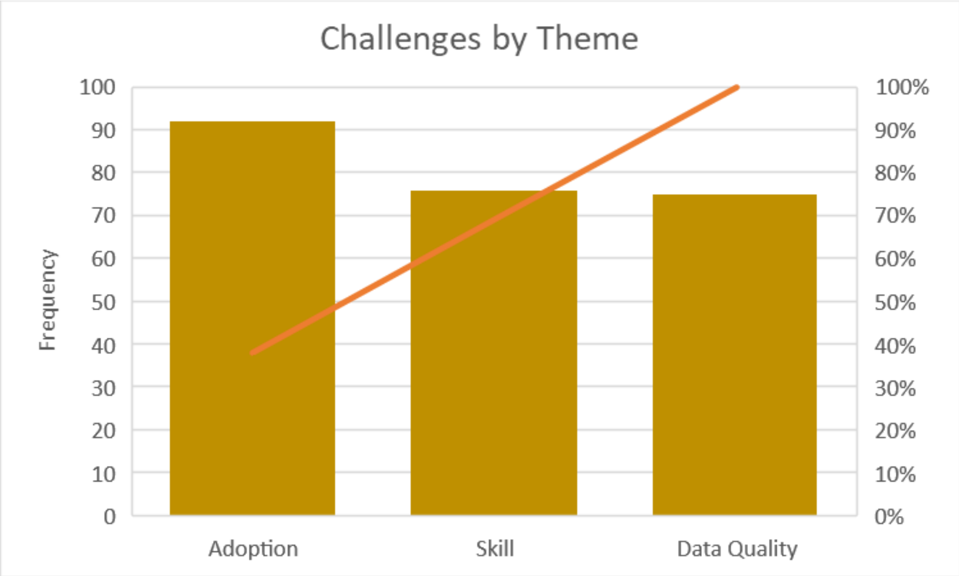


Figure 18: Challenges of using Big Data by theme

Adoption was coded the most, followed by skill and data quality, with the size of each bar indicating the frequency with which each theme was cited by participants. Although the skillset was not as frequent as adoption, it was a significant feature of many participants' experiences when it comes to the use of big data as well as the reason why the use of big data is not being adopted.

Table 14 below shows the themes with their sub-themes as well as the number of times they were mentioned and coded as frequency.

Rank	**Theme**	**Subtheme**	**Frequency**
1	Data Quality	Trust and Integrity	43
		Platform and business process	32
Total			**75**

2	Adoption	Lack of Collaboration	37
		Legacy	18
		Resistance to change	10
		Time pressure	9
		Culture	7
		Bureaucracy	6
		Perceived Complexity	5
Total			**92**
3	Skill	Resource constraints	37
		Skills development	18
		Data Literacy	16
		Challenges in recruitment	5
Total			**76**
			243

Table 12: Themes

It is worth noting that, most of the participants cited trust and integrity of the data as the highest reason why they do not use big data followed by the lack of collaboration amongst risk and technology teams. Each individual theme is explored below:

I. Data Quality

In the section where participants were asked about why they don't use big data and if they have considered using it, one of the main themes was data quality issues, which lack of trust and integrity as well as platform issues being the sub-themes.

Trust and Integrity

According to participants, the lack of trust in the data was mostly the reason why they don't use big data. This is because of the belief that the data may be inaccurate (either from a source they don't trust or because of the data not being governed). Data integrity is the premise that information process, computer equipment and/or software, people, and so on, or any collection of these elements, meet an expectation of sufficient and appropriate quality in some given context (Hoisl B, 2018).

Participant	Excerpt	Interpretation
Participant 2	"In our organization, we are struggling to get the basic data management principles embedded to allow us to fast track and become a digitally LED organization. So, there are a lot of Metadata management components that we still need to address."	Lack of data management foundation
	"I think that's what the struggle has always been, it's almost the trust of the data. So, I think in lots of instances where it's internal data, trust becomes a bit of an issue. And it's primarily because we don't have our full end-to-end data quality management processes in place. So as	Data trust Inadequate data quality processes

	much as you would like to use the data you are not confident enough whether the data is accurate, complete, valid, and timely." "We have tons and tons and tons and tons of data available, but because we can't trust the data it, it's sometimes holding us back from using it, which is a challenge"	

Table 13: Trust and Integrity

One of the participants noted that it showed that the lack of usage of data is not so much that the data is not available, but the willingness to use it with confidence is low.

Platform and business process

The second subtheme was more system/ platform related. Most of the participants felt that the data that is fed into systems is not of good quality. Participants expressed a lack of trust in the platforms that carry the data, believing that not enough care is taken when capturing and moving the data from one source to the other. One of the major contributors of poor data quality mentioned was that data is captured incorrectly from the source, impacting downstream channels, and subsequently resulting in mistrust of the data as seen by participants' comments in table below:

Participant	Excerpt	Interpretation
Participant 18	"Having worked in the first line and full group legal risk, I think is unique in that, a lot of the work that is done is paper based".	Manual processes Systemic issues
Participant 14	"We've got this system in place, but we still using Excel to do like all our reports and so forth. We do quite a lot of things	Data trust

	manually, whereas we have the opportunity, we've got the tools in fact and the teams have spent quite a lot of time to enable us to extract, do some analysis, and so forth. But we still find ourselves sitting and doing analysis and trends and so forth on Excel spreadsheets and I think for me personally, what I found is why I end up doing that is, I would go on to some of our tools and the data. I would question the data accuracy, so then I'll go to my spreadsheet and rather do the calculations myself because at least it's something I can say I trust, and I can back it up rather than spending two weeks engaging with the teams to say guys actually this doesn't look right"	
Participant 16	"If people are lazy, if they have not been trained or informed of downstream applications that we'll be using this data now, it really the issues are systemic there and then we see that later when we start to do the correlation and the aggregation of the data."	Garbage in garbage out, improper use of systems
Participant 7	It's almost as though people who are capturing the data don't understand why they are capturing the data, you know, or what the point of it should be because it's a very long thing and it doesn't get to the root of what the issue is sometimes"	Lack of understanding of downstream impacts

Table 14: Platform and business process

II. **Adoption**

The second theme that emerged was adoption of big data. According to Dvorski Lacković et al., (2020), the adoption of big data technologies can aid banks in avoiding and pre-empting situations like financial crises. The findings suggest that overall, the use of big data has not been adopted into risk management. This data shows that this is as result of bureaucracy within the channels; the lack of collaboration or a centre of excellence; time pressures – choosing between what is important and what is urgent; the perception that big data is a complex subject to be shied away from; as well as legacy and culture issues.

Lack of collaboration was cited the most, followed by legacy amongst the subthemes. Many participants believed that to effectively grasp and utilize big data, they needed to interact with colleagues who possessed complementary skills. This seemed to be attributed to the lack of a centre of excellence as another participant remarked “there must be a centre of excellence there's no awareness, big data use must start from a centre of excellence or a big data partner.” This is because most of the participants did not regard themselves as data experts but rather as risk subject matter experts. In support of this, a participant noted that “It (big data usage) should be more of a group-wide decision. So, it should be pushed down from the top. There should be like a data office, and they should have the skills and when business needs a certain skill, you know they should reach out to the data office and that's skill should be provided you know.” As a follow-up question, participants were asked about where they thought big data initiatives should be driven from. Most of the participants said that they did believe technology should drive the tooling portions and that risk management teams own the data.

Bureaucracy was mentioned as one of the reasons for the lack of adoption. Participants alluded to difficult processes, regulations, and a lot of handoffs as the reason why people tend to not bother tapping into big data and other innovations. They described the process of getting hold of the data tedious and time-consuming with another participant noting that “We have to go and figure this stuff out, and it takes time to do that. Go to this person, go to this place, go to this link. Go to this team. It's a mad run-around. It's the same governance, red tape, and time and effort to just get hands on the data”.

In addition to that, a theme of culture emerged, mainly classified as either risk culture or data culture. One participant emphasized the importance of data culture by saying that “I feel data culture is so important for the organization. If we don't have that how we expected to know what technologies are there, how are we are expected to plug into data analytics, how can we start transforming and driving value out of the data for the organization”. Remarking of data culture, another participant said, “We are more comfortable with the qualitative stuff and not the quantitative stuff”.

Many participants described big data as a complex phenomenon, especially those who are in the nonfinancial risk types. Another said that “I think they're (people) scared of the term digitization because they don't understand what that means”, additionally that “It's that perceived level of complexity that comes with it that I think would deter people from wanting to try it, from wanting to explore it and I suppose lack of understanding”.

Table below captures some of the remarks made in relation to the sub themes:

Participant	Subtheme	Excerpt
Participant 19	Lack of Collaboration	“There’s also no consistency in the way the data is pulled or utilized because each team manages that on their own.”
Participant 2	Legacy	“The other thing is around legacy systems and that comes with the complexity of the bank that we work with that we've built a lot of our architecture around legacy systems etcetera. So, it does become one of a challenge when it comes to understanding whether the basic data management controls are in place”.
Participant 17	Resistance to change	“…so, you know we’ve done it this way for years. Why should we change it? type of attitudes hanging around that digitization doesn't even hit the road maps in those types of teams because we simply just scrambling to stay ahead of requests from executives from managers in terms of providing data.”

Participant 4	Time pressure	"I was given the go-ahead to go and explore this and kick off the project in a non-financial risk etcetera and it was a great idea. I've been asking for these resources. I was told, do it in your own time. So, that's why I didn't pursue it any further".
Participant 16	Culture	"The problem I see and the problem I've been experiencing (is) we don't have that data culture. We don't have a chief data officer. If we had a Chief data officer, you could think of someone that strategically leads the data culture for the organization."
Participant 19	Bureaucracy	"I think a big roadblock is obviously, the CID way of doing things. So, you know first getting data under approval, then going through a privacy impact assessment then going to requesting access. Sometimes you need to explain why you need access, you know that's a whole process on its own. Then the second one would be from a legal point of view, you know, are we allowed to view that data in that way, are we overstepping our bounds, is there a POPI implication with what we're doing".
Participant 5	Perceived Complexity	"It would be that perception of difficulty or complexity in these things and that, it's not for me. So, I'm not even going to try it. So, I would not necessarily be interested as a result of that and think the perception that it's also for a certain type of people, it's for the quantitatively minded, it's for the actuaries."

Table 15: Adoption

III. Skill

The skill theme is made up of four (4) themes mainly resource constraints, skills development, data literacy and challenges in recruitment. Although skill was coded the least, it was always an emergent concern when it comes to the main reason why big data is not used in risk management and featured as a subtheme of other main themes.

Participants acknowledged a lack of resources such as time to build their big data skills. The data showed that teams did not always have analytics skills as well as the capacity to take on a more quantitative role within the team. Participants expressed low confidence when it came to analytical work although they mentioned the financial institution does allow them the freedom to innovate their processes.

Resource constraints

Resource constraints in this context looked at the capacity of employees to learn and implement big data skill, and results are presented below:

Participant	Excerpt	Interpretation
Participant 7	"We don't have people who can extract the data".	Lack of resources
Participant 23	"I'll make a practical example, like currently we don't (even) have the manpower or the resources to actually do what we need to do". "We got like deadlines to meet and have targets to meet, and that upskilling takes a lot of time and it's usually time away from that BAU."	Lack of resources Time constraints
Participant 4	"My team is not analytically inclined. They are SMEs in what they do. But, in terms of looking at data and throwing associations from data together to get to kind of insights etcetera, we even have poor sight that doesn't exist in my team currently."	No analytics capability. Nature of the skill is qualitative

Table 16: Resource Constraint

Skills development

Skills development was the second most cited issue when it comes to the skill theme. Many participants alluded to skill development not being a priority for the business and mentioned that to learn how to use big data they have to self-learn. They reported that the difficulty was a matter of taking time for self-learning and balancing business priorities as depicted in the Figure 19 below:

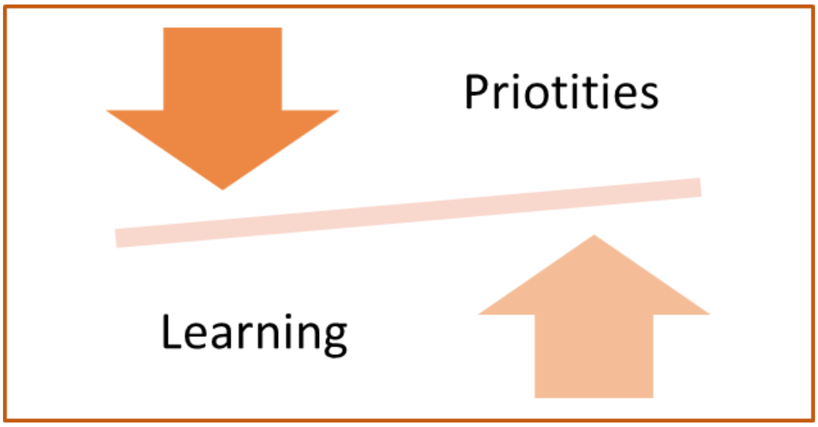


Figure 19: Learning vs. Priorities

The participants who perceived that there was no direct drive to get the risk teams to develop the big data skill. These technical skills have been argued to be critical to comprehend as technologies evolve and become more ingrained in modern organizations (Mikalef & John Krogstie, 2020). Table below shows a remark made in relation to skills development:

Participant	Excerpt	Interpretation
Participant 21	"there's an understanding that we need to be enabled and capacitated and a lot of us sitting within their risk space, I don't think I'm being capacitated. If I didn't have a background in IT, I would not be able to engage with the systems that I currently have in the manner that I'm able to engage with them. But a lot of my colleagues don't have a similar background. A lot of my colleagues don't have even excel skills set, pivot tables as that bare basic introduction to pivot tables, VLOOKUP's, things like that".	Upskilling current resources

Table 17: Skills Development

Data Literacy

Another emerging subtheme was the lack of data literacy. According to D'Ignazio & Bhargava (2017), data literacy is the capacity to understand, work with, analyse, and debate with data. Participants remarks are summarised below:

Participant	Excerpt	Interpretation
Participant 2	"So, data literacy is a general issue across the industry, and when I say data literacy, it is really the day-to-day management of your data through its life cycle because if data literacy was something we really had, I mean if that skill was enriched in our business, we would definitely see the benefits especially when it comes to big data et cetera. But right now, we don't have that skill."	Understanding of data management lifecycle
Participant 17	"When they kind of decentralized all of the CDO functions, I don't think each function went away and each business unit went away to understand the implication of that and that brought data literacy in the bank down in my mind because we lost skills without replacing them."	Data literacy deterioration Effects of federated model

Table 18: Data Literacy

Challenges in recruitment

Difficulties in finding the resources in the market came up as one of the issues that was highlighted by participants.

Participant	Excerpt	Interpretation

Participant 17	"In the world, there's a sort of shortage of certain data skills. There's just not enough to go around, and that's why these consultancy firms charge an arm to assist us. Sourcing those specialized skills is a challenge, they're not easy to find, and these guys come at a premium".	Skills shortage in the market Scarce skill
Participant 19	"You know how many people can pull the data from a Hadoop and use it in a way you know, push it into a dashboard. Not a lot."	Scarce skill

Table 19: Recruitment Challenges

According to Miller (2014) the global scarcity of big data skills is widespread, and every location faces similar issues.

4.3.2.3. Adoption of Big Data

Following the indication from participants about the challenges they face when it comes to big data usage, numerous probing questions were asked to understand the level of adoption and possible hinderances. In this regard, research question 3 was used as an extension to research question 2, to say that if there is big data adoption, what level is it, and if there isn't, then what is stopping the adoption thereof, and garner if there are any advances made to adopt big data usage into risk management processes. The probing questions were framed in such a way as to allow both considerations.

Participants were first asked about their level of readiness in adopting big data and there were mixed responses to this. Readiness to adopt technologies can be described as people's proclivity to embrace and use new technologies for achieving goals in their personal and professional lives (Nugroho et al., 2017). Most of the focus group participants shied away from answering this question, showed some hesitation, or gave a generic answer while iterating challenges. However, individual managers did respond to the question directly.

Some participants alluded to risk functions being ready to adopt big data technologies. This is seen from remarks below:

Participant	Excerpt	Interpretation
Participant 21	"We can see from a risk management perspective that we are now starting to want to leverage the information that we have now been capturing and reporting on"	Appetite to adopt
Participant 1	"I think for me, there is a large gap in how we consume it. I think we're ready. I think just getting it into the accuracy of 100% comfort, I think we still have a long way to go there".	Big data Roadmap

Table 20: Readiness to Adopt Big Data

However, most of the participants disagreed with risk functions being ready to adopt big data into risk functions. The following remarks were made in support of that:

Participant	Excerpt	Interpretation
Participant 5	"The level of business is 0 and currently, I'm looking at the current resources that they have, and I'm looking at their capabilities, I'm looking at their mindsets. I'm also looking at the current scope of work that we do. I don't think there is "	Level of business readiness
Participant 5	"I think the level of readiness is at a different level because if I had to look at for example group operational risk, I've already admitted to the fact that there is a risk analytics function that is sitting there and maybe not just group operational risk, but the different risk types I think",	Level of business readiness Readiness in pockets

Participant 1	“Readiness, I would say we're not 100% there. We pull up bar just based on the large data in legacy systems and all those things, so the size of the bank makes it a bit difficult, and I think mergers and acquisition histories and all those things”	Roadmap
Participant 4	“it's also in its infancy stages but getting us to the adoption of big data. Right now, not great. I'd say maybe 5-6 years’ time. That picture could change”.	Roadmap

Table 21: Level of Readiness

When asked whether they felt empowered to use or implement big data in their risk management processes, the following results depicted in table 24 came about:

Empowerment	**Focus group**	**Management**	**Total**
Yes, there is empowerment	6	3	9
No, I do not feel empowered	3	-	3
There is some empowerment	5	2	7
no. of voices	**14**	**5**	**19**

Table 22: Empowerment

Most of the participants said they felt empowered, with a smaller number showing disagreement. All technology risk participants alluded to being empowered to use big data and in contrast, participants in the operational risk teams felt less empowered. Alluding to this, a participant said “I feel like I'm not empowered by the entire bank. The systems are incorrectly used, and the supporting teams are not visible enough. When you want something, they refer you to a standard and I want simplicity. I don't want a huge document". The remaining participants said that there was some empowerment, however, felt that more needed to be done. A participant relayed this by saying "I think there is some empowerment. But as I say, we're not following it through. And that's where the problem comes in. So, we keep saying we want new ideas, we want innovation. But when it comes forward? Then it just kind of falls off the radar", while another said, "It's a yes or no answer, so I do feel empowered that you know it is something that can be useful, but the limitation is the cohesion between the tools that are there and time because we are always limited in what we can do with the given time, that's the challenge".

Participants were then asked if they think enough is being done to ensure the use of big data gets adopted into risk management practices. Similarly, there were mixed responses. Some participants believed there was a level of commitment, with remarks such as "I think now it's getting the airplay that it deserves. Previously I wouldn't have said so, but I think, it's becoming quite top of mind", while some did not think enough was being done. As observed in the participants' comments, one said "I don't think enough is being done. I mean, we do have certain campaigns and things that kind of pop up every now and then, then there's some digital campaigns (such as) - bring your ideas. People bring ideas forward. Brilliant ideas that come up. But we never actually follow through, that's the problem. And as I say we need that. I mean the commitment and that (of) senior management actually".

To further deduce what contributes to the lack of readiness reported above, all negative answers to the above questions were coded further to see what elements contribute to the lack of readiness which impacts the adoption and usage of big data in risk management. According to the model proposed in Rusly, Corner, & Sun (2012), factors affecting readiness can be psychological and structural. Employee attitudes (beliefs and intentions) are examples of psychological elements, whereas organizational capabilities and resources are examples of structural factors (Rusly et al., 2012). Figure 20 below indicates what contributes more to the lack of readiness:

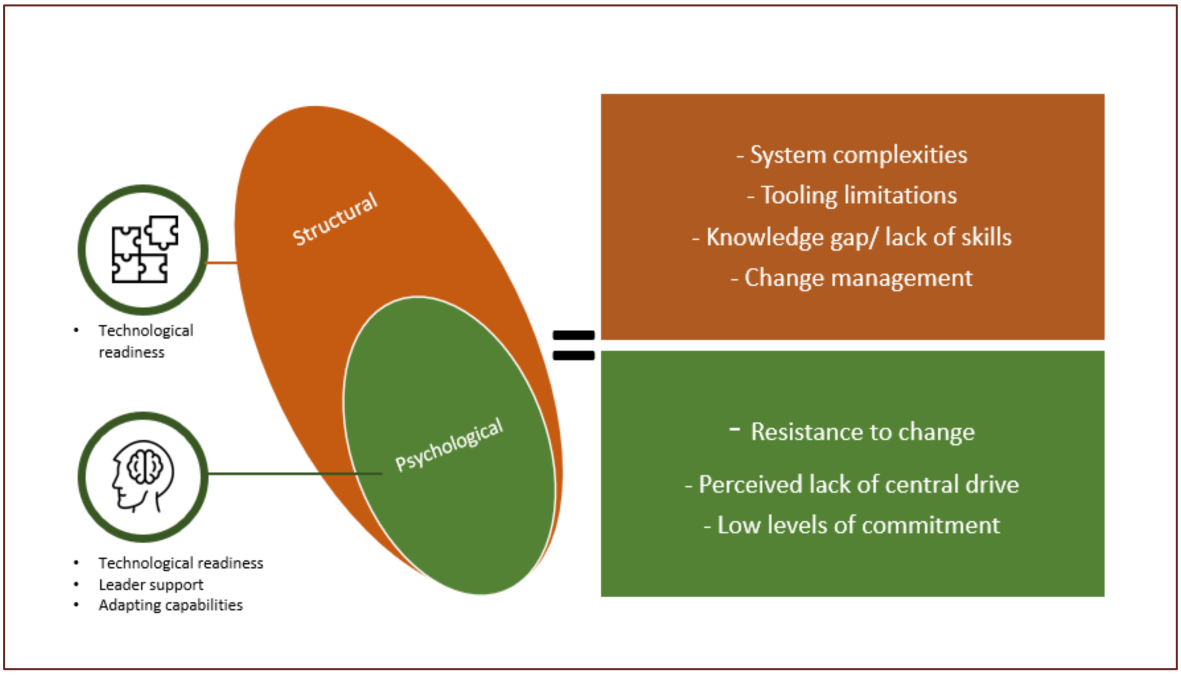


Figure 20: Factors affecting Big Data adoption

One of the probing questions was around maturity levels where big data was reported to be used. The term "maturity" can be defined as the condition of being fully developed, flawless, or prepared, or alternatively, as the state of being ripe (Braun H. , 2015). Most of the participants said that the maturity levels were low. One participant remarked, “It's not a question of the delay we have adopted it. It's a question of maturing”. Participants were asked what can be done to increase maturity levels, and these factors were derived to be skillset and data literacy, data management (and quality), tooling, and central function. These factors were then used to analyse the perceived maturity levels. The mapping was done according to the frequency of codes i.e., the number of times participants reported a factor as a barrier to adoption. Figure 21 below provides a mapping of the maturity levels:

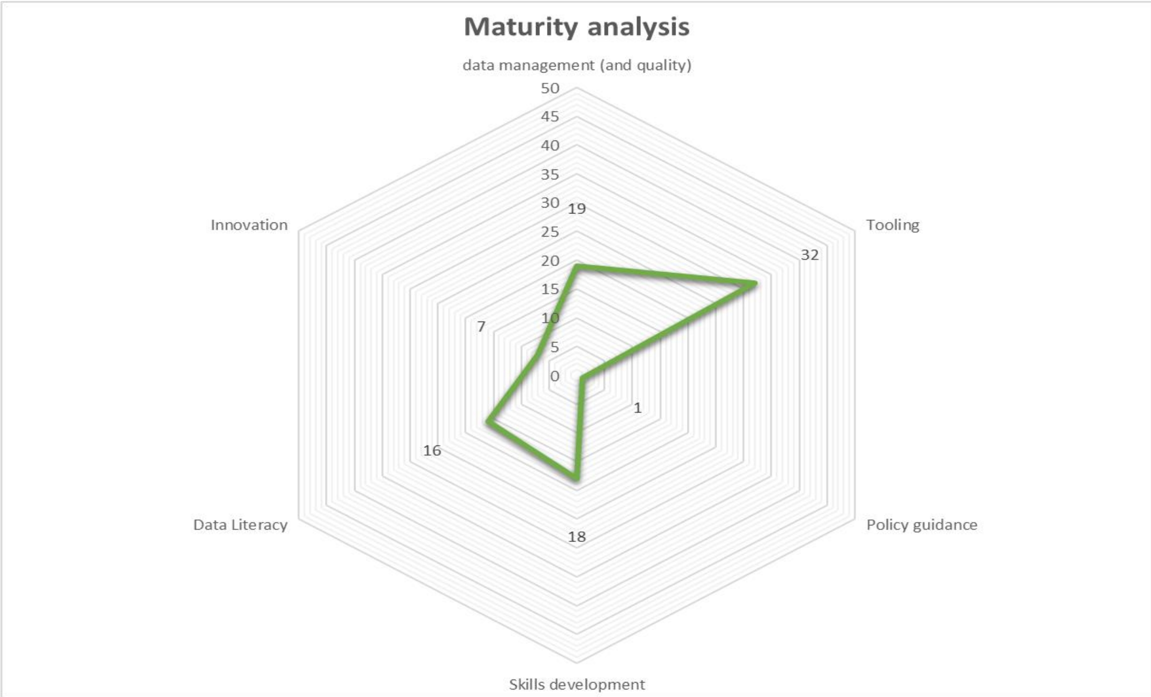


Figure 21: Big Data Maturity Analysis

According to the diagram, policy maturity is low. This is because the current policies don’t recommend the use of big data as a tool or capability to be used in risk management. Innovation also ranked the lowest, as can be seen in the preceding sections, there is almost always no follow-through in innovation initiatives, and if there is, it happens mostly in teams that are more centres of excellence such as the technology functions. Commenting on this, a participant said “I think we're not doing enough, for example, (the technology office) has some initiatives that have been launched to encourage the development of automation and get some cutting-edge technology and solutions that are being put forward to host an event or competition, you know, once a year. It’s meaningless. It's pointless. there's that divide that happens. So, what now? I need to go to a more specialized tech team to be more part of that digital innovation strategy”.

Data literacy scored one of the lowest in maturity levels. To explain this, a participant said “data literacy is a general issue across the industry, and when I say data literacy is really, the day-to-day management of your data through its life cycle. This ties back to the lack of centralised function indicated in the lack of adoption section cited as lack of collaboration as well as skill gaps.

Tooling refers to the technology required to use big data. It is the highest in terms of maturity levels however, the issues raised are around the availability of tools is that the usage is happening in silos and big data tools are not being used correctly. A participant remarking on tooling said “It all comes down to, you know, people using the tool for the right purpose and not getting confused in terms of what is it that you want to do using big data versus a small requirement to say I need to do a management report. It's two different types of users of data and you need to tool appropriately.” To increase maturity levels, skills development is one of the major factors that must be considered as well as the drive to ensure risk communities are data literate.

4.4 Summary of the findings

The chapter set out the results of the study pertaining to the challenges of using big data in risk management. These were results from responses by 23 participants interviewed as part of both individual interviews as well as focus groups. The chapter introduced the analysis method used, which was thematic inductive analysis, used to let data form the themes instead of deriving them from literature review. It was also demonstrated that thematic analysis was appropriate for this qualitative study as it allowed for the researcher to link an analysis of the frequency of a theme to one of the entire texts, increasing the research accuracy as well as its overall meaning.

Results for the three research questions where presented. For the first research question looking at the application of big data in risk management, a number of themes such as decision making, usage in risk event detection, risk appetite and analytics came about, with the themes of decision-making and understanding risk events ranking the highest. Another type of big data application that emerged was the detection emerging risks. Identifying emerging risks can assist banks in mitigating problems before they become actual events. As a component where big data can be applied, risk appetite was ranked the lowest. The findings demonstrated how big data may increase risk identification accuracy, proactive risk mitigation, strategic decision-making, and overall organizational resilience.

The second research question looked at challenges preventing the use of big data in risk management. This section served as the research anchor because the findings thereof provided insight into why big data is not widely used in risk management. The findings addressed challenges such as lack of skillset, data quality, technology infrastructure, talent acquisition, and bureaucracy. The data did show that, while big data was used in specific risk types, particularly financial risk types, there were problems that prevented the larger risk community from completely integrating big data usage into their processes. It is worth mentioning that most interviewees mentioned data trust and integrity as the primary reasons for not using big data, followed by a lack of collaboration between risk and technology teams.

Lastly, results pertaining to the overall adoption of big data technologies into risk management processes were presented. Participants in the more technology inclined risk teams all mentioned feeling empowered to use big data, whereas those in the operational risk teams felt less empowered to use big data. The difficulties raised were around the availability of tools, with utilization occurring in silos and big data technologies not being used effectively.

In this regard, the next section will discuss the findings of these results.

CHAPTER 5. DISCUSSION OF FINDINGS

5.1 Introduction

The purpose of this chapter is to discuss the findings set out in chapter 4. In this discussion, a structure similar to chapter 4 will be followed. The first set of questions looked at the demographic and the participant's understanding of big data. That was followed by research question-specific data. The discussion will thus address the definition of big data provided by the participants, the application of big data in risk management, the challenges as well as the factors affecting the adoption of big data in risk management.

5.2 Big Data Understanding

Participants were asked what their understanding of big data is, rather than providing them with a technical explanation. This was to allow the definition to be shaped by the participants while they reflect on their own engagement or experience with the term. It was found that, although most of the participants had an idea of what big data was, it was mostly theoretical knowledge. In addition, some of the definitions provided by participants differed in how they experienced data. For example, the participants working in business-facing functions saw big data as the use of data analytics capabilities whereas in the more specialized roles it was seen holistically to mean data management lifecycle and data literacy. In more manually intensive risk areas, big data was seen as a complex quantitative capability that only certain people can use.

As seen in Figure 11: Big data Definition, it can be deduced that the definition provided by participants encompasses a mixture of a phenomenon usually coined as VUCA, a term that was used to characterize the new theater of battle as volatile, uncertain, complex, and ambiguous, now used by corporate groups and researchers to define the international setting of 21st-century business (Elkington, 2020). The words that came up such as emerging, variation, and variable show that big data is seen as something that is ever-changing and comes in different forms, and one can never really be sure exactly what it is. We also see words such as “mess”, “Sophisticated”, “Impossible” and “complex” coming up, indicating that big data is seen as something complex and ambiguous, making it not so easy to understand. The concept that big data is a complex phenomenon can also be found in Grander & Silva (2021) definition of big data. They state that that the data is processed using a range of advanced algorithmic techniques and presented using highly technical lingo when discussing complexity. This denies risk specialists comprehension of how any data was processed, how the results were obtained and how they may be criticized. This lack of understanding could also be attributed to time pressures, as many participants spoke about of lack of time to learn or implement big data because of the need to prioritize business-as-usual tasks.

When participants spoke on the concept of big data, they also focused on the nature or forms that it comes in. One of those is that big data extends beyond structured data, and contains elements of unstructured data, such as media. This is synonymous with what was found in literature where authors such as Maja & Letaba (2022), define big data as data classified into three types: structured, unstructured, and semi-structured. All are synonymous with what various authors have described big data to be, including the predominant 5v model (volume, velocity, variety, veracity, and value) by various authors. Other concepts that emerged were technologies, storage, and analytics, capturing the additional characteristics of big data which are described as real-time analytics, new infrastructure, and tools (high-performance computing, storage, networking, data models) (Maja & Letaba, 2022). Looking at the literature and the evolution of big data, it can be concluded that the concept will continue to encompass more technologies and more complexities, as more people use it and get exposed to it, and as such develop more definitions. Although there are common themes that emerge, it is not envisaged that big data will ever have one common definition, and there will be continuous additions to what is currently the 5vs, as more data types and sources emerge.

5.3 Risk Management and Big data

This chapter discusses the findings from the analysis performed on the research results according to the three main research questions proposed in the first chapter of this study. The primary objective of the study was to gather data in so far as the use of big data is concerned and to understand the nuances surrounding the application of big data as a capability in risk management processes, as well as understand how banks can leverage large data sets that sit at their disposal. The purpose of this discussion is to highlight key findings and implications thereof. The first part looks at the application of big data in risk management, the second set of results are challenges of using big data in risk management, and lastly, factors hindering the adoption of big data.

5.3.1. Application of Big Data in Risk Management

The first research question was for the participants to express where and how they think big data can be applied to risk management. The basis of this question was that, if something is not useful for a process then there is no need to adopt it. This was to get a sense of whether participants deemed big data as an important component that should be used in risk management, or if it was irrelevant to the discipline. This question also helped the participant to reflect and prompt thinking about their own impediments in so far as the usage of big data is concerned. The analysis of the findings showed that indeed big data is deemed a beneficial capability that can be used in risk management processes.

The themes that came about were the application of big data in decision-making, the process of understanding risk events, the detection of emerging risks, analytics, and risk appetite setting.

Decision Making

Most of the participants expressed that big data could be used in the process of decision-making. Decision-making is important in risk management as it drives a lot of the strategy and the amount of risk the bank will be willing to take, as the core of organizational effectiveness and competitiveness has always been found in decisions made (Elgendy et al., 2022). This finding is synonymous with a lot of literature written on big data, as it is often highlighted as the main reason why organizations must pursue big data. For example, Adil & Abdelhadi (2022), mentioned that using big data helps with enhanced decision-making by providing valuable insights, generating new business models, products, and services, creating knowledge, implementing management strategies, innovating products, and services, enabling dynamic market analysis, and assisting managers in decision-making through customized dashboards (Adil & Abdelhadi, 2022).

The most basic definition of the term "data-driven decision-making" is the use of data analysis to inform policy and procedure decisions (Picciano, 2012). It has been said that intuition itself has shortcomings when it comes to decision-making as it falls short of recognizing other factors that may not be obvious to the decision-maker. According to Elgendy et al., (2022), a lot of pressure has been placed on decision-makers to make the best judgment possible in a timely manner. It was interesting to see this coming up as a top theme from risk specialists, as it shows that they are fully aware of the contribution their processes make toward executive decisions.

Many of the risk report requirements escalate to board committees, as recommended by guidelines such as King corporate governance codes, that the risk management process and risk evaluation should be handled by a special committee or a board committee that reports to the board (King Report On Corporate Governance For South Africa, 2002). The utilization of big data analytics can enable banks to make decisions that are more informed and data-driven, resulting in enhanced business performance and a competitive advantage. The capacity to derive practical and applicable insights from extensive sets of data can empower risk specialists or managers responsible for making decisions to recognize developing patterns, inclinations of customers, and prospects in the market, thereby enabling them to allocate resources more efficiently and devise well-informed strategies.

Understanding risk events

Participants also expressed that big data can be used for the identification and understanding of risk. Most said that although they have systems for capturing the events, they never go back to see what the data says. The data shows that the adoption of proactively identifying risk events can decrease the probability and severity of risk events, thereby protecting the assets, reputation, and interests of stakeholders within the organization.

According to Dicuonzo et al., (2019), big data can be used for the early identification of risk events. Supported by the theoretical framework by Dvorski Lackoviæ et al., (2016) adopted for this study, the similarities can be drawn in that the framework is centered around using big data to understand and enhance processes to limit risk exposure. The Figure below illustrates how big data may be utilized in banking to improve the process of managing all significant risks:

Figure 22: Risk Management Framework (Dvorski Lacković et al., 2020)

According to this framework, big data usage in the risk management process is expected to reduce the amount of time needed to respond to new risks that banks are faced with through direct risk identification, real-time risk quantification, daily control of risk exposures, and increased reporting transparency (Dvorski Lackoviæ et al., 2016). However, both the literature nor the results are not clear on how this can be done, or if there is a distinction in risk types where this can be applied more effectively than the other, e.g., financial risks vs nonfinancial risk events. Also, risk events are a process where some form of a system is generally used, and although some people see it as technical, the capturing thereof is quite manual and tends to be incomplete, so the translation to how big data can be applied to that process was quite vague and not found in literature.

Analytics

Another theme that emerged for the application of using big data for analytics. Most of the participants mentioned that big data could be used in analytics. The participants were not elaborative on the kinds of analytics that can be applied and mostly focused on giving examples of how analytics can help the business. Most of the benefits expressed were around the use of analytics for client identification, customer journeys and predictions, and trend analysis. Data analytics is a discipline that encompasses the management of the entire data lifecycle, which includes data collection, cleansing, organization, storage, analysis, and governance (Ravinder & Srinivas, 2021), however, in relation to this, participants only mentioned the analysis components, assuming data would have been cleansed already (if data quality- one of the challenges expressed, is managed). On the other hand, Dicuonzo et al., (2019) states that BDA is the process that analyses big data sets using algorithms to extract diagrams, reports, and useful and unknown information. Looking at the overall results we can say that participant's application of big data leaned more into this definition.

Participants also mentioned the use of unstructured data from external sources, that can be fed into the analytics. This agrees with the Shabbir & Gardezi (2020) who says that big data analytics is a strategy of analyzing a large volume of data gathered from various sources in an unstructured, semi-structured, or structured form using various analytic methods. According to the data collected, unstructured data can help detect reputational threats, customer behaviour and perception about the organisation, helping come up with controls and pre-empting events such as protest actions and the effect of loadshedding on consumer spending and actions.

Detection of emerging risks

In the results, it also emerged that big data can be used in the identification of emerging risks and help banks mitigate risk before it materializes into an event. An example of emerging risk detection could be earlier indications of customer distress and instances where customers don't have a credit score (also known as spin scoring), helping banks make more accurate judgments about the creditworthiness of borrowers and potential lenders (Al-Dmour et al., 2022). The idea of spin scoring is a new finding, as a phenomenon that does not come up in literature.

According to McKinsey (2015), Banks operate in an ever-evolving environment that is influenced by stringent regulatory requirements, emerging risks, and intense competition. Emerging risks can be ever-evolving threats such as cybercrime and escalated levels of fraud in banking. According to Shalaginov et al., (2019), cybercrime is growing at a fast pace and will require sophisticated systems to be handled. In addition, fraud is another of the ever-evolving risks as criminals become more and more sophisticated and require the betterment of systems to handle it. The use of big data within banks could aid in the detection of fraud in existing monetary transactions (Nobanee et al., 2021). This result is synonymous with what was found from Nobanee et al., (2021) that big data can be essential for fraud detection, and risk management in commercial banks. Although big data cannot eradicate such risks, it can help discover them early on so that future progress into riskier pathways can be avoided (Eltweri et al., 2021).

Revisiting the theoretical framework here, it can be seen that big data can be applied in the detection of emerging risks, especially with the help of unstructured data.

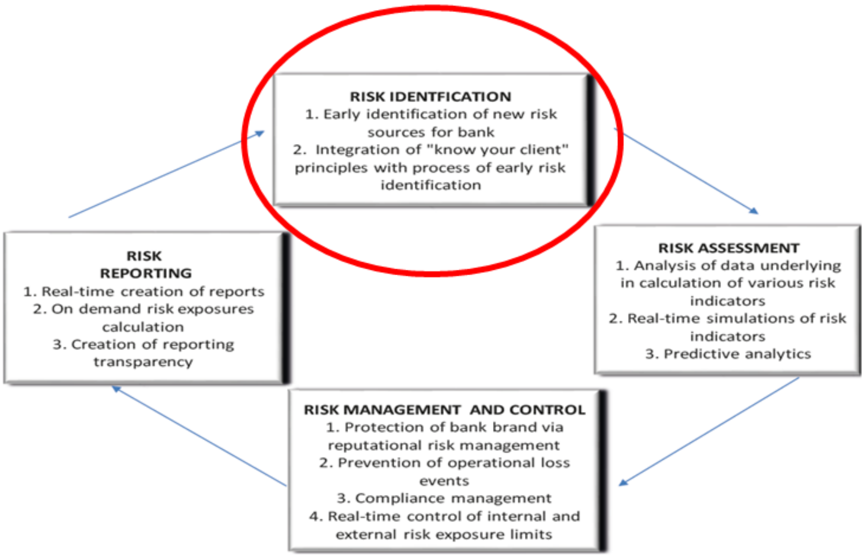


Figure 23: Risk Management Framework (Dvorski Lacković et al., 2020)

In the reviewed literature, many authors agree that quick identification and quantification of emerging risks, as well as transparency when reporting, are critical components of risk management (Khamitkar & Rafi, 2020), and it can be concluded that the results do match what is available in the literature and can be a good fit for generalisation, although not collectively from a single source. This will depend on the focus of an organisation or direction of research taken.

Risk Appetite setting

According to the results, big data can also be applied to risk appetite settings. Participants were of the opinion that big data can be used for understanding business financials, and the accuracy thereof can aid in improved risk appetite setting. Risk appetite is the written expression of the aggregate degree and categories of risk a financial institution is willing to take or avoid in order to meet its business objectives (Shen & Hu, 2007). There was not a lot of literature found that focused on the use of big data for risk appetite. The research studies found were focused more on the calculation of risk appetite using mathematical formulas.

Risk appetite setting forms a threshold for measuring the effectiveness of the control environment and to see how much capital is required to hedge against risks, and this could be the reason why this result emerged. This finding could be unique in that, participants considered qualitative data points in the calculation of risk appetite instead of the traditional methods of using economic variables, for example as done in Shen & Hu (2007) and Gai & Vause (2005). The use of big data for risk appetite setting is significant in that it can enable banks to access timely and detailed information that is essential for establishing adaptable frameworks for risk appetite. This can help them adjust their risk tolerance in response to evolving market circumstances and corporate objectives.

In the literature, it was mentioned that banks will need to beef up their risk management processes in order to keep up with the ever-increasing regulatory requirements. According to McKinsey (2015), big data can help in driving some of the compliance required both internally and to the regulators. However, there was no explicit evidence of this from both the literature as well as in the results, on exactly how big data can aid in regulatory compliance. Regulatory changes are said to have had a substantial impact on corporate practice, including the definition of a general risk management framework, strategic decision-making, stress tests and scenario analysis, and risk reporting, and institutions of all sizes are subject to these requirements (Principles for effective risk data aggregation and risk reporting, 2013). Although there was no finding directly related to how big data can enhance regulatory compliance, it is still worth noting and assuming that applying big data in risk management processes, will result in compliance with regulation as a by-product, especially in regulations such as Basel IV, IFRS 9, and BCBS 239 where compliance is demonstrated through data and reporting (examples in Mawutor (2014).

In their research findings, Soltani Delgosha et al., (2021) also found that the most essential applications of big data in banks are fraud detection and credit risk analysis. They also concluded that the primary motivators for beginning big data initiatives were the enhancement of decision-making and new product/service development, which is found to be synonymous with the above findings. Although we find common themes with literature when looking holistically at the results for research question 1, it is not so for regulation and risk appetite setting.

5.3.2. Challenges of using Big Data in Risk management

This research question sought to understand why big data is not used extensively in risk management. Challenges garnered from the results include skill set, trust and integrity of data, complexity, risk and data culture, lack of senior management buy-in, lack of financial support, commitments, and lack of driving innovation amongst others. These topics were grouped into three main themes mainly data quality, skill, and adoption. Adoption was the most frequent theme, followed by skill and data quality, with trust and integrity of data being the highest reason for not using big data.

Data Quality

Looking at data quality, participants expressed that the reason for not using the data available was that they could not trust it. Different reasons for the lack of trust were issues around the way data is captured into the systems as well as the lack of data management and data control issues. It is perceived that data sitting in systems is not accurate, and therefore people stay away from using it, this is especially because the data would primarily be used in reports that in turn get used by boards. This results in the abstinence from using data sets as users would prefer using their own manual data sources that they can vet and verify to be accurate. As per the findings in Surbaktia et al., (2020), a significant proportion of business leaders, specifically one-third, expressed a lack of trust in the information they rely on to make informed decisions. Although data quality concerns are not a new issue, the enormous volumes of data, the rapid velocity of receiving data, and the great diversity of data from various sources have raised its importance (Soltani Delgosha et al., 2021).

According to Liu (2023), data quality issues have been a significant concern within the information systems field. However, because the data quality is defined as "fitness for use," meaning that: an item that is deemed suitable for a particular purpose may not necessarily exhibit adequate standards for a different application, the notion of data quality can be subjective (Liu, 2023). Despite the recognition of data quality as a significant theme, no subsequent efforts were made to explore the level of sufficiency of data quality for the intended purpose, as can be seen in other studies. Therefore, this finding may not be a full representation of the overall population, although most of the participants expressed this as a common concern.

Adoption

The second theme in discussion from the findings is adoption. Although it has been mentioned by many authors that big data can help banks avoid and pre-empt financial risks, adoption is still a major concern. Big data adoption can often be hindered by management and cultural impediments, such as a shortage of business and technical skills and worries about data privacy and security, lack of business and technical skills, as well as data privacy security concerns (Braun H. , 2015). According to the research results, this is hindered by bureaucracy, lack of collaboration, time pressures, perception of complexity, and legacy and culture issues.

Participants cited a lack of collaboration and a centre of excellence as one of the challenges with the use of big data. They believed that big data use should be driven from a centre of excellence, or a big data partner and that technology should drive the tooling portions while risk management teams own the data. It is envisaged that if there is a centre of excellence then the decision on tooling can be made easy. Furthermore, the ability to collaborate with more technically inclined teams could support the use of big data in risk management. According to Soltani Delgosha et al., (2021), the most significant challenge to achieving successful integration of data and big data, as perceived by 57% of bank executives they interviewed, was the presence of multiple information silos. Additionally, they found that there was a lack of operational/strategic data sharing between different business unit. Although the results in this study found a similar result, theirs was more inclined to the sharing of actual data as opposed to expertise. According to the Kiziltan (2018), expert input and facilitation of collaborative efforts are essential for the functionality of a big data analysis system to enable multifaceted issues to be understood.

Bureaucracy was cited as one of the reasons for the lack of big data usage and consequently, adoption. This was cited to be due to a lack of innovation as a result of difficult processes, regulations, and many process handoffs. There was not much about bureaucracy in the literature, except where studies focused on bureaucracy in education and how bureaucracy can hinder innovation. Bureaucracy can be a challenge in big organizations as there are many departments and employees handling different parts of the process. The centre of excellence as discussed in the preceding section can help minimize this, especially in organizations that have adopted federated models, as is the case for this case study.

Additionally, a theme of culture emerged, mainly classified as either risk culture or data culture. Organizational culture is the collective philosophy, ideology, values, assumptions, beliefs, hopes, behaviours, and norms that unify and define an organization (Senarathna et al., 2014). It is characterized as an informal framework of values and norms that regulate interactions between individuals and groups (Senarathna et al., 2014). Data culture is essential for organizations to be able to transform and drive value out of data. According to Acker & Clement (2019), data culture can be looked at as the process of cultivating data, datafication, and cultures of use. In this case, results show that participants seemed to be more comfortable with the use of qualitative data than quantitative data. Additionally, this issue was attributed to not having a centre of excellence to drive data initiatives.

What was found interesting in the literature particularly was the mention of cultural analytics that examines the emergence of data cultures through data collection, and how they are addressed, investigated, analysed, and interpreted. However, the study was not mature enough to explain this phenomenon in detail. Various studies have documented the impact of organizational culture on innovation, but none was found that specifically focuses on big data culture, or how organizational culture can hinder the adoption of big data. According to the Data and AI Executive Survey, whose participants included 64 c-level technology and business executives representing very large corporations, it was made clear that there is still an eminent need for data-driven cultures, where data is treated as an important business asset that receives more attention, investments, and resources (Elgendy et al., 2022). In this study, it was not determined how organizational culture affects the use of big data or how the perception thereof affects the extent of the usage and adoption, and therefore similarities and differences with literature are inconclusive.

Big data is a complex phenomenon, especially for nonfinancial risk types, and its perceived complexity deters people from exploring it. According to the findings, this is attributed to lack of understanding and time pressures. Participants expressed that, because data is processed using advanced algorithms and presented using technical lingo, it became a difficult concept to grasp.

Literature does point out big data as being complex, but not always as an impediment to using it. According to the Kiziltan (2018), managing vast and quickly growing volumes of data has shown to be difficult, but a fundamental change to make it easier to process is currently taking place. Due to this perceived complexity, participants did not feel the need to explore big data and believed it does not belong in their area of specialty.

Other themes such as lack of time to learn or implement big data due to the need to prioritize business-as-usual tasks. This finding was not represented in the literature and might be specific to the organization. Legacy and resistance to change were also among the themes that came up as a challenge causing a lack of adoption. Although research studies to support legacy as an impediment to adoption were not many, studies relating to resistance to change do exist, even though they don't focus primarily on big data. Bedeley and Iyer (2014) identified one of the challenges of big data usage as legacy, stating that, even though a specific system may have been selected to replace a group of legacy systems in a functional domain, the implementation will not be successful unless the legacy system's unique critical requirements are identified, documented, and addressed by the replacement system.

According to Srivastava & Agrawal (2020), it is observed that employees exhibit resistance towards changes due to a multitude of controllable and uncontrollable factors. Resistance observed in certain situations may be because of the tendency to perceive change as a shock since it is coupled with uncertainty (Srivastava & Agrawal, 2020). This may as well be true in this case looking at the findings where participants prefer doing things the way they have always done them (Acker & Clement, 2019).

Skill

The data revealed skillset as another challenge facing risk professionals when attempting to use big data and associated capabilities. The skill theme was made up of four themes: resource constraints, skills development, data literacy, and challenges in recruitment. Participants acknowledged a lack of resources such as time to build their big data skills, as well as a lack of analytics skills and capacity to take on a more quantitative role. They expressed low confidence in analytical work, although the financial institution does allow them the freedom to innovate their processes. According to Braun (2015), the scarcity of skilled personnel and the absence of professionals specialized in big data will pose a substantial limitation to the generation of value from this resource.

Resource constraints hinder the capacity of employees to learn and implement big data, as they have deadlines and targets to meet, making the acquisition of big data skills an afterthought. Risk teams need to develop big data skills, which are critical as technologies evolve. However, there is no direct drive to get them to develop these skills. D'Ignazio & Bhargava (2017) says that big data learning has to involve both technical skills and emancipation through learner-guided explorations, facilitation over teaching, accessibility to a diverse set of learners, and a focus on real problems. Without this support, it is not clear how risk professionals will be able to prioritize acquiring big data skills, as the quality of the data tools and processes depends on the person utilizing them (Maja & Letaba, 2022).

Data literacy is the capacity to understand, work with, analyse, and debate with data. The process of reading data entails comprehending the nature of data and the specific facets of reality that it conveys. The process of working with data encompasses the tasks of generating, obtaining, refining, and overseeing data operations such as filtering, sorting, aggregating, comparing, and executing various data manipulation techniques (D'Ignazio & Bhargava, 2017). A participant noted that when the chief data office functions were decentralized, skills were lost without replacing them. If that was not the case, data literacy would have been enriched in the business, enhancing, and promoting the use of data. Business leaders need a data-driven mindset to deliver value from big data, and it will be impossible to deliver big data value if the leaders themselves don't have data-driven mindsets (Braun, 2015) or don't make decisions that are in favour of big data usage and stewardship.

Another issue that emerges is that even trying to get the skill outside has become a mission. A seemingly global "pandemic", this shortage of data-proficient professionals in the market is constraining the ability of organizations to extract substantial benefits from big data (Miller, 2014). Technologies are emerging rapidly, and the organization's workforces have to adapt at higher rates. This causes these resources to be constrained in the market, making recruitment processes for these skills difficult. According to Burton, Mastrangelo, & Salvador (2014), this is due to the lack of undergraduate, professional, and executive education programs that is partly to blame for the skills shortage in data strategy and technical management positions.

According to Soltani Delgosha et al., (2021), Lack of technical and skilled workforce, high implementation cost and lack of top management support are major obstacles to using big data in banks. It is argued that organizations require individuals who possess extensive knowledge in statistics and machine learning, as well as managers and analysts who can proficiently employ big data with a rigorous analytical methodology (Braun, 2015). Business analysts and bank managers with critical insights and analytical competence are also needed (Soltani Delgosha et al., 2021). Looking at the data presented, and it can be concluded that this skillset is very limited in the risk management communities, and according to Braun (2015). The production of this kind of talent is challenging and requires several years of training This means that a data-oriented mindset and data-driven decision-making are essential for creating value from BDA (Soltani Delgosha et al., 2021). These expressions are synonymous with what data for this study is showing.

Culture and resistance to change could be drivers of the lack of skillset, i.e., if there is no need for change people will not make much effort to upskill themselves. Another thing is, because this is not a drive within the organization, risk specialists almost must take the initiative to upskill themselves in their own time. Risk management changes tend to be driven by regulation, and if there is no regulation that forces the use of big data or knowledge thereof, that change might not necessarily happen, and the effects will be seen where adoption is concerned. The implementation of big data technologies necessitates a novel skill set that is unfamiliar to most IT departments (Zakir et al., 2015). Moreso, risk management professionals as they are traditionally not technologically and quantitatively inclined.

BDA projects have serious challenges, such as a lack of required skills, technical infrastructure, unsupportive senior executives, and inadequate organizational culture, which is reflected in the results of this study (Soltani Delgosha et al., 2021). Other studies showed challenges of big data in financial organizations, such as massive data storage, various data types, efficient computing systems, and optimizing memory usage (Dvorski Lacković et al., 2020) as well as an expression that many BDA projects are not initiated due to high infrastructure costs (Soltani Delgosha et al., 2021). Both of these arguments contain themes that did not come up in the results of this study.

Looking at the overall results for research question 2, it can be concluded that although there were exceptions such as culture and bureaucracy, most of the challenges expressed as similar to what is available in the literature, albeit the research focused on the technical challenges or aspects of big data such as complexity and storage and not how people perceive the challenges. In order to effectively reconstruct businesses with a focus on data utilization, it is imperative for companies to employ a team of proficient and knowledgeable professionals such as information strategists, data architects, data governance specialists, visualization experts, and chief analytics officers (Burton et al., 2014). Maja & Letaba (2022) suggests that to overcome these issues of skills, banks should engage in collaborative efforts and form partnerships with their competitors, including Fintechs and start-ups to enable them to leverage complementary skills and resources and safeguard their revenue.

5.3.3. Adoption of Big Data

Research question 3 was used to understand the level of adoption of big data usage in risk management processes. According to Deutsche Bank research (2015), the following factors are driving big data technology adoption in the financial industry: (1) explosive data growth (increased number of banking instruments and transactions), (2) regulation (requirements regarding real-time risk management and related financial transactions; enhanced risk reporting; risk simulations), and (3) fraud detection and security (operational risk management issues, such as detecting fraud risk among employees) (Dvorski Lackoviæ et al., 2016).

According to the findings, some participants did not believe there is a level of commitment to increasing the levels of adoption, while others did not think enough was being done to drive adoption. Technology readiness is not a measure of one's aptitude or capacity, but rather of perception or thinking about technology (Nugroho et al., 2017). In the literature, five (5) categories of technology adopters can be found, namely explorers, pioneers, doubters, fear (paranoids), and laggards are the five user categories. The data shows that risk professionals did not see themselves as pioneers of big data usage, but rather as explorers, where they have time outside of their normal business as usual tasks, hence the need to want to rely on a centre of excellence to drive the usage. This can also be attributed to not seeing themselves as data experts and hence not thinking about driving or pioneering the use of big data sits in their space. The participants did not express an unwillingness to learn and therefore cannot be classified as laggards. In this case, it is justified why the level of adoption is low.

The data showed that the readiness to adopt big data was in pockets and silos and that some business units were more ready than others. That could be because of the perception that some risk types (financial and non-financial) are more quantitative than others, for example, model risk vs operational risk. The level of readiness for model risk would be more than that of operational risk, as they already use some form of data and history to develop credit models. Another reason could be that individuals' readiness to adopt technology can differ in that some can be positive and optimistic with a tendency to be a pioneer in the use of new technology, or negative, that is, feeling uncomfortable and apprehensive about the technology (Nugroho et al., 2017).

A study by Abankwa (2023) measured the propensity to adopt, comparing small and medium-sized banks, and they found that small and midsize banks differed significantly in their readiness to use BDA, where the midsize banks were more prepared to adopt big data. These findings are not surprising looking at the amount of data mid and big banks already have access to, as well as the maturity of the technologies they use. Compared to the data from this study, this finding applies to the bank in discussion, however, it was expected that the level of readiness would have been a bit more than midsized banks and not similar in context. Although this study did not go further to apply methodologies to measure actual levels of readiness, the data synthesized is enough to make this comparison, looking at factors presented in the study by Abankwa (2023). It is to be noted however, that, although the bank overall may be ready to adopt, it may not be so for risk management functions specifically, and therefore a blanket conclusion that risk teams would be ready to adopt big data usage may be misleading, and may, in fact, correlate more with the levels of readiness associated with small banks, especially looking at results where participants said they don't feel empowered to use big data.

Participants reported low maturity levels where some form of big data was being used. According to the findings presented, this can be attributed lack of policy guidance where recommendations for the use of big data as a tool or capability to be used in risk management can be driven. Although this is true for Bank in question, the generalization cannot be applied as this finding is too specific. A drive for innovation was also reported to be low, meaning that risk professionals spent their time only on their required tasks and not on bringing new ideas, this translating to things being done how they have always been done. also ranked the lowest, with no follow-through in initiatives. Data literacy scored one of the lowest in maturity levels, while tooling was the highest in terms of maturity levels. To increase maturity levels, skills development, and the drive to ensure risk communities are data literate must be considered.

Capgemini Consulting (2014) defines three levels of analytics maturity: Beginner, Proficient, and Expert. Beginner involves no defined data structure, poor data governance, basic reporting, and scattered talent. Proficiency involves data available for existing and potential customers, statistical and forecasting tools, and a well-defined recruitment process (Dvorski Lackoviæ et al., 2016). According to the data, the bank is still very much at beginner levels of maturity. Participants were asked about their level of readiness in adopting big data and there were mixed responses. Some participants alluded to risk functions being ready to adopt big data technologies but admitted that there was still a long way to go. According to Dvorski Lackoviæ et al., (2016), organizations with a lesser level of analytics maturity will have a more difficult time extracting value. Participants identified four factors that can be used to increase maturity levels: skillset, data literacy, data management, tooling, and central function.

Most of the studies in the literature looked at the drivers and challenges of adoption, mainly in SMEs, and not dominantly in big commercial banks. The results of one particular study focused on how factors such as data privacy hinder the adoption of big data whereas others looked at factors that encourage technology readiness, studying attitudes and behaviour towards technology change. For example, Nugroho et al., (2017) says that organizational change may be impacted by an awareness of individuals' general mental models. This study did not study behaviours towards change, however, attitudes towards big data usage were noted to be positive, with participants willing to learn if the right support is given. Furthermore, research studies on readiness focused on building frameworks and models to measure reading, which for the purpose of this study, was not done. All studies stated above, as well as others available in the literature review, illustrate the immense potential of using big data in risk management if the adoption gap is closed.

5.4 Conclusion

In this chapter we recapped the results from the research as well as look for similarities and differences that appear between the findings and literature. We deduced from both the finding and literature that, big data is a beneficial capability that can be used for analytics for capabilities such as client identification, customer journeys and predictions, and trend analysis. However, it was concluded that, although direct benefits or relationship between using big data and compliance is not clear, compliance with regulation will be a by-product of using big data in risk management processes.

The data also showed that risk professionals face four challenges when attempting to use big data and associated capabilities: resource constraints, skills development, data literacy, and recruitment. Resource constraints hinder the capacity of employees to learn and implement big data. It was also conclusive from both literature and the findings from this study that adoption of big data is hindered by bureaucracy, lack of collaboration, time pressures, perception of complexity, and legacy and culture issues. Even though literature seem to focus more on the technical challenges such as data processing and difficulties collecting data. The main take away was that big data is a complex phenomenon, and its perceived complexity deters people from exploring it due to lack of understanding and time pressures that compete with business priorities.

Lastly, we looked at the level of adoption of big data in risk, to ascertain the extent to which big data utilization has been embraced in risk management processes. The adoption of big data technology in the financial industry is being driven by various factors, including the exponential growth of data, regulation, and fraud detection and security, as per the findings of Deutsche Bank research conducted in 2015. Risk management professionals perceive themselves as explorers rather than pioneers in the realm of big data utilization. The data showed a favourable disposition towards the use of big data, with participants expressing a willingness to acquire knowledge, provided adequate support is provided.

CHAPTER 6. CONCLUSIONS & RECOMMENDATIONS

6.1 Introduction

The purpose of this study was to discover the challenges faced when using big data in risk management as well as understand why big data is not widely used in risk management. The study further went to explore how big data can be used in risk as well as the level of readiness to adopt big data into risk management. Following the discussion of results in chapter 5, a consolidation of the key findings for the research question will be presented in this chapter. This will be followed by recommendations and suggestions for future studies.

6.2 Conclusions for Research Question 1

The first research question was used to determine where and how big data can be applied to risk management. In consolidating the findings, it can be concluded that big data is indeed deemed a beneficial capability that can be used in risk management processes. The key findings were that big data can be used for decision-making which is important in risk management as it drives the strategy and size of risk the bank will be willing to take. Secondly, that big data can be used in analytics to enable banks to make decisions that are more informed and data-driven, resulting in enhanced business performance and an increased competitive advantage. It can also be concluded that adopting proactive identification of risk events can decrease the probability and severity of risk events, thereby protecting the assets, reputation, and interests of stakeholders. This answered first research question and was not found to be different from what literature initially suggested, but instead provided more ways to apply big data in risk management such as risk appetite setting.

As presented in earlier chapters, literature indicated that banks are highly regulated and risk management is at the centre of the regulation however, the results did not provide evidence of how big data can improve compliance to regulation. Therefore, the limitation found in the findings was that it could not be directly established how big data can address the regulatory aspects of risk management.

6.3 Conclusions for Research Question 2

The second research question looked at the challenges experienced when using big data with the aim to understand why big data is not widely used in risk management. Three main concerns can be noted as hinderances when using big data, namely, Data quality, lack of skills and the lack of adoption. Data quality is defined as "fitness for use," meaning that an item that is deemed suitable for a particular purpose may not necessarily exhibit adequate standards for application, and in this regard, a significant proportion of business leaders expressed a lack of trust in the information they rely on to make decisions. This was evident in the literature especially concerning the complexity of harmonising data coming from multiple sources, as well as ensuring the integrity of the data through its lifecycle.

The second challenge noted in the research findings was adoption of big data, and that the contributors to this are bureaucracy, lack of collaboration, time pressures, perception of complexity, legacy, and culture issues. The most significant challenge to achieving successful integration of data and big data was the presence of multiple information silos and a lack of operational/strategic data sharing between different business units.

This study also examined the impact of organizational culture on the adoption of big data. The data showed that risk professionals were more comfortable with the use of qualitative data than quantitative data. From this it can be concluded that big data is not used because it is perceived as complex quantitative capability that would be a step out of a comfort zone for risk professionals, and its perceived complexity deters people from using it. These findings also highlight the amount of change management that would be required to get risk professionals up to speed with big data and to adopt it in their processes.

The last challenge was skillset and lack of resources to perform tasks related to big data application. Resource constraints hinder the capacity of employees to learn and implement big data, as they don’t have enough time and people to add on additional big data capabilities. A shortage of data-proficient professionals in the market is constraining the ability of the bank to extract substantial benefits from big data. Culture and resistance to change could be drivers of the lack of skillset, as well as the fact that risk specialists must take the initiative to upskill themselves in their own time. Although skillset is fully covered in literature, culture was not too evident as a challenge preventing the use of big data in risk management. Other studies showed challenges of big data such as the complexity and enormity of data storage, various data types, the lack of efficient computing systems, and optimizing memory usage.

This answered the research question in that, it helped highlight the reasons why big data is not being used despite banks collecting a myriad of data daily.

6.4 Conclusions for Research Question 3

Research question 3 was used to understand the level of adoption of big data usage in risk management processes. According to Deutsche Bank research (2015), the following factors are driving big data technology adoption in the financial industry: explosive data growth, regulation, and fraud detection and security. The data showed that risk professionals did not see themselves as pioneers of big data usage, but rather as explorers. The readiness to adopt big data was in pockets and silos and some business units were more ready than others due to the perception that some risk types are more quantitative than others. Individuals' readiness to adopt technology can differ in that some can be positive and optimistic with a tendency to be a pioneer in the use of new technology, or negative with very less desire to change the way things have always been done.

The data showed that there is low maturity level, attributed to lack of policy guidance and no drive for innovation. The data also suggested that the bank is still at beginner levels of maturity in adopting big data technologies. Factors identified to help increase maturity levels are concluded to be skillset, data literacy, data management, tooling, and formation of a central data function. Most of the studies in the literature looked at the drivers and challenges of adoption, mainly in SMEs, and not dominantly in big commercial banks, as general comparisons across literature were made to draw these conclusions.

6.5 Recommendations

According to Dvorski Lackoviæ et al., (2016), the use of big data can be hindered by various factors such as silos, the considerable amount of time required to analyse large data sets, a shortage of skilled personnel, unstructured content, the exorbitant cost of storing and analysing voluminous data sets, and the complexity involved in collecting and storing such data. To tackle these concerns, it is suggested that institutions devise a systematic approach to draw in analytics experts, merge data from internal, external, and social media sources, establish a framework for governing data, and encourage the adoption of optimal methodologies. Therefore, the following recommendations are made:

It is recommended that the bank create a comprehensive talent acquisition strategy that can effectively source and recruit proficient analytics experts. This includes working on the retention strategy to ensure that they don’t lose their data resources as it becomes more difficult to replace them.

The bank must foster a work environment that promotes continuous learning and professional growth more especially for big data topics. Employee skill development and training should be prioritized in order to give staff members the information and abilities necessary to handle big data for risk management. This can be done through making data certifications and learning programs more accessible, while also providing time for people to learn vs. letting the work priorities compete with their big data upskilling. Employees can be encouraged to enrol for certifications such as Certified Analytics Professional (CAP) and Certified Data Management Professional from the Data Management Association (DAMA), that can provide a solid base for understanding of data and how to manage and use it.

Risk executives can run campaigns and scorecards that encourage knowledge sharing. The risk specialists must also become open to learning and sharing within the organization as to establish a culture of knowledge sharing. This will provide an opportunity for collaborative learning and skill development. This will lessen the discouragement towards the use of big data as well as reduce the silos.

Big data quality proved to be a major concern for risk specialists. Therefore, risk management executives must take on initiatives to intensify data capturing and acquisition processes, improving data governance processes to guarantee data accuracy and increase data trust. This can be done through technologies such as Robotic process automation, generally known as RPA, where the collection of data and the storage thereof is done through automating repetitive tasks and the implementation of efficient data capturing mechanisms, resulting in reduced data quality issues.

Data quality can also be improved by employing methods such as data validation and data cleaning that can aid in filling in variables such as missing values, removal of duplicates, and standardization. Data standardization can ensure uniformity across many data sources, standardized data formats, units, and naming conventions.

In addressing the lack of collaboration and bureaucracy, it is recommended that efforts be made to break down information silos that exist within the company to promote a collaborative and data-sharing culture. This can be done by creating unambiguous procedures for the exchange of data and strategies between various business divisions, encourage information sharing and cross-functional teamwork. In this regard, faster and streamlined approval processes can be implemented and pointless administrative obstacles removed.

6.6 Suggestions for further research

This study provided valuable insights into the application of big data in risk management as well as the challenges of usage thereof. However, there were limitations regarding the study in that it did not cover extensively the “how” but rather the why. These limitations offer an opportunity for prospective future research and improvements. Although the study revealed valuable insights, it could benefit from a larger sample to get more coverage and generalisation.

This study can be conducted quantitatively to measure aspects such as big data maturity levels and experiments on how big data can be applied, through building models that demonstrate how big data analytics can be applied on risk data. This includes the capacity to conduct real-time analytics and predictive modelling.

Literature underscores the importance of regulatory obligations that are associated with risk management. A study can be launched to find the relationship between big data and improvements in regulatory compliance, to see if using big data does improve regulatory compliance or not. Researchers can further examine how big data might be used to automate reporting, simplify compliance procedures, and guarantee compliance with intricate and changing financial requirements.

Furthermore, the data showed that culture plays a role in the lack of adoption of big data in risk management. A study can be launched to find out the extent of big data usage in different organisational cultures and environments.

Lastly, to test the variability of big data, a study can be done to see how external data can enhance the ability to detect emerging risks, e.g., through social media feeds and socio-economic factors external to the bank.

REFERENCES

Abankwa, K. (2023). Big Data Analytics: Adoption Readiness of Small and Midsize Banks.

Acker, A., & Clement, T. (2019). Data Cultures, Culture as Data – Special Issue of Cultural Analytics. *Journal Of Cutural Analytics*.

Adil, B., & Abdelhadi, F. (2022). A BIG DATA ANALYTICS FRAMEWORK FOR COMPETITIVE INTELLIGENCE SYSTEMS. *Journal of Theoretical and Applied Information Technology*.

Al-Dmour, R., Amin, E. A., Saad, N., & Zaidan, H. (2022). Interrelated Factors Influencing the Adoption of Big Data Applications: Empirical Study in Jordan. *Jordan Journal of Business Administration*.

Alhojailan, M. I. (2012). THEMATIC ANALYSIS: A CRITICAL REVIEW OF ITS PROCESS AND EVALUATION . *West East Journal of Social Sciences*.

Amanullah, M. A., Habeeb, R. A., Nasaruddin, F. H., Gani, A., Ahmed, E., Nainar, A. S., . . . Imran, M. (2020). Deep learning and big data technologies for IoT security. *Computer communications*, 495-517.

Arena, F., & Pau, G. (2020). An overview of big data analysis. *Bulletin of Electrical Engineering and Informatics*, 1646-1653.

Balios, D., Kotsilaras, P., Eriotis, N., & Vasiliou, D. (2020). Big Data, Data Analytics and External Auditing . *Journal of Modern Accounting and Auditing*, 211-219.

Bedeley, R., & Iyer, L. (2014). BIG DATA OPPORTUNITIES AND CHALLENGES.

Braun, H. (2015).

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology. 77-101.

Burton, R. M., Mastrangelo, D., & Salvador, F. (2014). Big Data and Organization Design. *Journal of Organisation design*.

Businesstech. (2023, 03 13). *Businesstech*. Retrieved from https://businesstech.co.za/news/business/672039/absa-hikes-dividend-as-load-shedding-batters-south-africa/

Cheng, X., Liu, S., Sun, X., Wang, Z., Zhou, H., Shao, Y., & Shen, H. (2021). Combating emerging financial risks in the big data era: A perspective review. *Fundamental Research*, 595-606.

Creswell, J. W. (2018). *Research design: qualitative, quantitative, and mixed methods approache.* Thousand Oaks, California, SAGE Publications, Inc.

D.Ravinder, & M.Srinivas. (2021). ROLE OF DATA ANALYTICS IN BANKING FINANCIAL SERVICES AND INSURANCE BFSI) - AN EXPLORATORY STUDY. *THE JOURNAL OF ORIENTAL RESEARCH MADRAS*, 155.

D'Ignazio, C., & Bhargava, R. (2017). Approaches to Building Big Data Literacy.

Dicuonzo, G., Galeone, G., Zappimbulso, E., & Dell'Atti, V. (2019). Risk Management 4.0: The Role of Big Data Analytics in the Bank Sector. *International Journal of Economics and Financial Issues*, 40-47.

Dr. K. Chitra, B. S. (2013). Data Mining Techniques and its Applications in Banking Sector. *International Journal of Emerging Technology and Advanced Engineering*.

Dvorski Lackoviæ, I., Kovšca, V., & Vincek, Z. (2016). Framework for big data usage in risk management process in banking institutions. *Central European Conference on Information and Intelligent Systems*. Retrieved from https://repozitorij.foi.unizg.hr/en/islandora/object/foi%3A3470

Dvorski Lacković, I., Kovšca, V., & Lacković Vincek, Z. (2020). A Review of Selected Aspects of Big Data Usage in Banks' Risk Management. *Journal of information and organizational sciences*, 317-330. Retrieved from https://www.researchgate.net/publication/347651966_A_Review_of_Selected_Aspects_of_Big_Data_Usage_in_Banks'_Risk_Management

Dworkin, S. (2012). Sample Size Policy for Qualitative Studies Using In-Depth Interviews. *Archives of Sexual Behavior*, 1319-1320.

Elgendy, N., & Elragal, A. (2014). Big Data Analytics: A Literature Review Paper.

Elgendy, N., Elragal, A., & Päivärinta, T. (2022). DECAS: a modern data-driven decision theory for big data and analytics. *Journal of Decision Systems*.

Elkington, R. (2020). LEADERSHIP DECISION MAKING- MAKING LEVERAGING BIG DATA IN VUCA CONTEXTS.

Eltweri, A., Alessio, F., & Osama, K. (2021). Applications of Big Data within Finance: Fraud Detection and Risk Management within the Real Estate Industry. New York, NY, USA.

Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 1.

Gai, P., & Vause, N. (2005). Measuring Investors' Risk Appetite. *Munich Personal RePEc Archive*.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 147-144.

Gontarek, W. (2016). Risk governance of financial institutions: The growing importance of risk appetite and culture.

Grander, G., & Silva, L. F. (2021). Big data as a value generator in decision support systems: a literature review.

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough? *Field Methods*, 59-82.

Guptal, A., Verma, A., Kalra, P., & Kumar, L. (2014). Big Data: A Security Compliance Model.

Hammarberg, K., Kirkman, M., & de Lacey, S. (2016). Qualitative research methods: when to use them and how to judge them. *Human Reproduction*, 498-501.

Haru, A. H. (2021). Challenges of Cybercrime on Online Banking in Nigeria a Review. *IDOSR JOURNAL OF BANKING, ECONOMICS AND SOCIAL SCIENCES* , 17-23.

Hessami, A. (2010). Perspectives on Risk, Assessment and Management Paradigms.

Hoisl B, S. S. (2018). Modeling and enforcing secure object flows in process-driven SOAs. *Software and Systems Modeling (SoSyM)*.

Ibrahim, A. (2012). THEMATIC ANALYSIS: A CRITICAL REVIEW OF ITS PROCESS AND EVALUATION. *West East Journal of Social Sciences*, 1.

Indriasari, E., Gaol, F. L., & Matsuo, T. (2019). Digital Banking Transformation: Application of Artificial Intelligence and Big Data Analytics for Leveraging Customer Experience in the Indonesia Banking Secto. *8th International Congress on Advanced Applied Informatics*.

ISO. (2018). *Risk management — Guidelines*. Retrieved from ISO: https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en:term:3.1

Jabareen, Y. (2009). Building a Conceptual Framework: Philosophy, Definitions, and Procedure. *International Journal of Qualitative Methods*, 49-62.

Jacob, S. A., & Furgerson, S. P. (2012). Writing Inter riting Interview Pr view Protocols and Conducting Inter ocols and Conducting Interviews: Tips for views: Tips for Students New to the Field of Qualitative Research. *The Qualitative Report*, 1-10. Retrieved from https://doi.org/

Jahja, A., Razimi, M., & Ramalu, S. (2021). 2021. GENERIC QUALITATIVE RESEARCH IN MANAGEMENT STUDIES. *JURNAL RISET AKUNTANSI DAN BISNIS*.

Kanchu, T., & Kumar, M. M. (2013). RISK MANAGEMENT IN BANKING SECTOR -AN EMPIRICAL STUDY. *International Journal of Marketing, Financial Services & Management Research*.

Kazi Imran Moin, D. Q. (2012). Use of Data Mining in Banking. *International Journal of Engineering Research and Applications*.

Khamitkar, S., & Rafi, M. (2020). BIG DATA: A Survey Paper on Credit Risk Management. *JETIR*.

(2002). *King Report On Corporate Governance For South Africa.*

Kiziltan, A. (2018). Challenges of Big Data Adoption in Turkish SMEs: A Case Study.

Kushwahaa, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights*.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*.

Liu, C.-T. &.-L. (2023). Creating and Validating an Information Quality Scale for E-Commerce Platforms. *Journal of Electronic Commerce in Organizations.*, 21. 1-28.

Maja, M. M., & Letaba, P. (2022). Towards a data-driven technology roadmap for the bank of the future: Exploring big data analytics to support technology roadmapping. *sciencedirect*. Retrieved from www.sciencedirect.com/journal/social-sciences-and-humanities-open

Masindi, N., & Singh, P. (2022). A Panel Data Analysis into the Impact of Regulations and . *Journal of Economics and Business*, 16-30.

Mawutor, J. K. (2014). Analysis of Basel III and Risk Management in Banking. *European Journal of Business and Management*.

McKinsey. (2015). *The future of bank risk management.* Retrieved from <https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/risk/pdfs/the_future_of_bank_risk_management.pdf

Mikalef, P., & John Krogstie. (2020). Investigating the Data Science Skill Gap An Empirical Analysis.

Miller, S. (2014). COllabOratIve approaches needed to close the Big Data skIlls gap. *Journal of Big Data and Organization Design*.

Misut, M., & Jurik, P. (2021). DATAFICATION AS A NECESSARY STEP IN THE PROCESSING OF BIG DATA IN DECISION-MAKING TASKS OF BUSINESS. *INTERNATIONAL CONFERENCE ON INNOVATIONS IN SCIENCE AND EDUCATION (NATURAL SCIENCES AND ICT).* PRAGUE, CZECH REPUBLIC.

More, R., & Moily, Y. (2021). Big Data Analysis in Banking Sector . *International Journal of Engineering Research and Applications*, 1-5.

Njie, B., & Asimiran, S. (2014). Case Study as a Choice in Qualitative Methodology. *IOSR Journal of Research & Method in Education (IOSRJRME)*, 35-40.

Nobanee, H., Dilshad, M. N., Dhanhani, M. A., Neyadi, M. A., Qubaisi, S. A., & Shamsi, S. A. (2021). Big Data Applications the Banking Sector: A Bibliometric Analysis Approach. *SAGE*, 1-18.

Nugroho, M. A., Susilo, A. Z., Fajar, M. A., & Rahmawati, D. (2017). Exploratory Study of SMEs Technology Adoption Readiness Factors. *Procedia Computer Science*, 329-336.

O'Brien, B., Harris, I., Beckman, T., Reed, D., & Cook, D. (2014). Standards for Reporting Qualitative Research. *Academic Medicine*, 1245-1251.

Öhman, A. (2005). Qualitative methodology for rehabilitation research1. *Journal of Rehabilitation Medicine*, 273-280.

Philipp Härle, Andras Havas, Andreas Kremer, Daniel Rona, Hamid Samandar. (2015). *The future of bank risk management.*

Picciano, A. G. (2012). THE EVOLUTION OF BIG DATA AND LEARNING ANALYTICS IN AMERICAN HIGHER EDUCATION. *Journal of Asynchronous Learning Networks*.

Powell, R. A., & Single, H. (1996). Methodology Matters. *International Journal for Quality in Health Care*, 499-504.

(2013). *Principles for effective risk data aggregation and risk reporting.* Basel Committee on banking supervision.

Ragavan, R. (2003). Risk Management In Banks. *CHARTERED ACCOUNTANT*.

Regulated-institutions. (n.d.). Retrieved from South African Reserve Bank Prudential Authority: https://www.resbank.co.za/en/home/what-we-do/Prudentialregulation/regulated-institutions

Rosman, R. (2009). Risk Management Practices and Risk Management Processes of Islamic Banks: A Proposed Framework . *International Review of Business Research Papers* , 242-254.

Rusly, F. H., Corner, J. L., & Sun, P. (2012). Positioning change readiness in knowledge management research.

Sagiroglu, S. a. (2013). Big Data: A Review.

Saunders, A. M. (2006). Financial Institutions Management: A Risk Management Approach.

Seale, C., Gobo, G., Gubrium, J., & Silverman, D. (2004). *Introduction: Inside Qualitative Research.* SAGE.

Senarathna, I., Warren, M., Yeoh, W., & Salzman, S. (2014). The influence of organisation culture on e-commerce adoption. *Industrial Management &*

data Systems, 1007-1021. Retrieved from http://hdl.handle.net/10536/DRO/DU:30066870

Shabbir, M. Q., & Gardezi, S. B. (2020). Application of big data analytics and organizational performance: the mediating role of knowledge management practices. *Journal of Big Data* .

Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research.*, 366-378.

Shakya, S., & Smys, S. (2021). Big Data Analytics for Improved Risk Management and Customer segregation in Banking applications. *Journal of ISMAC*, 235-249.

Shakya, S., & Smys, S. (2021). Big Data Analytics for Improved Risk Management and Customer Segregation in Banking Applications. *Journal of ISMAC*, 235-249.

Shalaginov, A., Johnsen, J. W., & Franke, K. (2019). Cyber Crime Investigations in the Era of Big Data.

Shen, D.-B., & Hu, K.-H. (2007). Bank Risk Appetite Measurement and the Relationship with Macroeconomic Factors: Case of Taiwan's Banks. *International Journal of Information Systems for Logistics and Management*, 25-39.

Soltani Delgosha, M., Hajiheydari, N., & Fahimi, S. (2021). Elucidation of big data analytics in banking : a four-stage. *Journal of Enterprise Information Management*, 1577-1596.

Srivastava, S., & Agrawal, S. (2020). Resistance to change and turnoverintention: a moderated mediation model of burnout and perceived organizational support. *Journal of Organizational Change Management*.

Srivastava, U., & Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. *Procedia Computer Science*, 643-652.

Statista. (2022). *Statista*. Retrieved from Statista: https://www.statista.com/statistics/1346961/leading-banks-in-south-africa-by-assets/

Statista. (2023). *number-of-banks-south-africa-by-type/*. Retrieved from Statista: https://www.statista.com/statistics/1348690/number-of-banks-south-africa-by-type/

Statista. (2023). *Total assets of Absa Group from 2017 to 2021*. Retrieved from Statista: https://www.statista.com/statistics/1121210/total-assets-of-absa-group/

Surbaktia, F. P., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*.

Tang, J., Ma, T., & Luo, Q. (2020). Trends Prediction of Big Data: A Case Study based on Fusion Data. *Procedia Computer Science,*, 181-190.

Tekiner, F., & Keane, J. A. (2013). Big Data Framework. *IEEE International Conference on Systems, Man, and Cybernetics*.

Thomas, D. R. (2003). A general inductive approach for qualitative data analysis.

Turgut, T. (2019). Risk management process in banking industry. *Munich Personal RePEc Archive*. Retrieved from https://mpra.ub.uni-muenchen.de/86427/

Wahab, S. N., Hamzah, M. I., Sayuti, N. M., & Tan, W. C. (2021). Big data analytics adoption: an empirical study in the Malaysian warehousing sector . *Int. J. Logistics Systems and Management*.

Wang, H. (2021). Credit Risk Management of Consumer Finance Based on Big Data. *Mobile Information Systems*.

Westhuizen, G. v. (2013). Bank Productivity And Sources Of . *International Business & Economics Research Journal*.

Zakir, J., Seymour, T., & Berg, K. (2015). BIG DATA ANALYTICS. *Issues in Information Systems*, 81-90.

APPENDICES

Appendix A- Mapping of Interview Questions to Research Questions

Consistency table: research questions, propositions, data collection and data analysis

RQ #	State Research Question or Objective	State Proposition or Hypothesis	Data collection detail	Data analysis method
1	**What is the perceived role of big data analytics in improving risk identification, assessment processes and mitigation strategies within the South African bank?**	To establish the perceived importance of big data in the organisation and risk functions	Interview guide questions 1 - 3	Thematic analysis

RQ #	State Research Question or Objective	State Proposition or Hypothesis	Data collection detail	Data analysis method
2	**What are the key challenges faced by the bank in effectively implementing and utilizing big data technologies for risk management?**	To establish the reasons why big data is not being used in risk management, this is a basis for answering the research problem	Interview guide questions 4- 9	Thematic analysis
3	**What factors hinder the adoption and integration of big data technologies and practices in the bank's risk management processes?**	To establish the level of big data adoption and garner if there are any advances made to formally adopt big data usage into risk management processes.	Interview guide questions 4- 9	Thematic analysis

Appendix B- Permission to Conduct Research

You forwarded this message on 2023/01/30 7:14 PM.

Subject: RE: Request for Permission to Conduct Academic Research at [redacted]

Good day [redacted]

Thank you.

I do attest and confirm that I will comply with the requirements as stipulated below.

Regards

Boipelo Mosiane
Manager: Secured Lending Data Management
Retail Risk Modelling, Measurement and Data Centre (RMD)

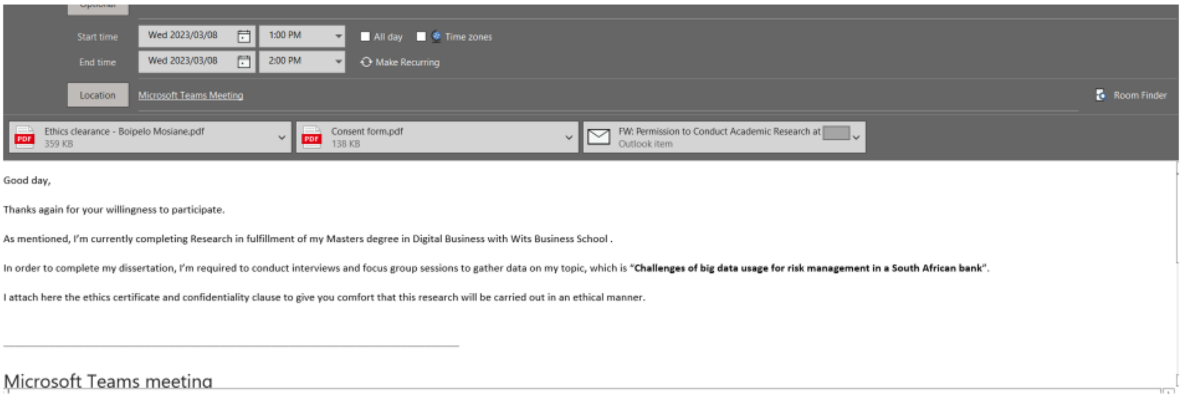
Morning Boipelo.

I noted that you have accepted to comply with requirements as laid down by [redacted], therefore comfortable. Should you at any time during the research find out that the study may have an impact on the brand or reputation of the group, please escalate to myself.

Regards.

[redacted]y
Head: Reputation Risk and Governance

Appendix C- Invitation to Participate in Research



Start time Wed 2023/03/08 1:00 PM All day Time zones

End time Wed 2023/03/08 2:00 PM Make Recurring

Location Microsoft Teams Meeting Room Finder

Ethics clearance - Boipelo Mosiane.pdf 359 KB

Consent form.pdf 138 KB

FW: Permission to Conduct Academic Research at Outlook item

Good day,

Thanks again for your willingness to participate.

As mentioned, I'm currently completing Research in fulfillment of my Masters degree in Digital Business with Wits Business School .

In order to complete my dissertation, I'm required to conduct interviews and focus group sessions to gather data on my topic, which is "**Challenges of big data usage for risk management in a South African bank**".

I attach here the ethics certificate and confidentiality clause to give you comfort that this research will be carried out in an ethical manner.

Microsoft Teams meeting

Appendix D- Information Sheet

STUDY INFORMATION DOCUMENT

Challenges of big data usage for risk management in a South African bank

Greeting

Introduction:
I am currently a student at the University of Witwatersrand Business School completing my research for a Master of Management in Digital Business. I am conducting research on the challenges of big data usage for risk management, with the aim of understanding the reasons why big data is not yet widely used in risk management practices

Invitation to Participate: I am inviting you to take part in a research study. Kindly confirm if you can spare an hour for a Microsoft Teams call by accepting the meeting invite.

This study involves:
The research design that will be used in this exploratory research is a case study, done through interviews and focus groups. The date will be selected according to your availability and will be communicated and confirmed with you prior agreement to participate and to the interview taking place.

I will then schedule a Microsoft Teams call for us, according to your chosen date and time. The questions asked will be about the usage of big data in risk management as well as the perceived value of big data.

Risks of being involved in the study: Should you not be able to participate at the agreed times, a new time will be rescheduled. If it may reasonably be anticipated that you are not comfortable proceeding, a reschedule will be done, or withdrawal allowed.

Benefits of being in the study: No direct benefit to the Participant is expected, and no monetary benefit of any sort is also expected. However, it is envisioned that participants will benefit from introspecting and analysis when providing answers, and possibly benefit from the changed perspective in so far as using big data in risk management is concerned.

The Participant will be given pertinent information on the study while involved in the project and after the results are made available.

Participation is voluntary: Refusal to participate is allowed and the participant may discontinue participation at any time without penalty as stipulated in the consent form; there is no requirement to provide a reason for withdrawing and any data collected on such a person will in default be destroyed, unless the Participant specifically consents to its retention.

Appendix E- Consent Letter

INFORMED CONSENT LETTER **Date: January 2023**

Challenges of big data usage for risk management in a South African bank:

Greetings

I am currently a student at the University of Witwatersrand Business School completing my research for a Master of Management in Digital Business. I am conducting research on the topic stated above.

The interview is expected to last about an hour and will help us understand why big data technologies are not widely used in risk management processes and as part of the enterprise risk management framework. Participation in this regard is voluntary, and you can withdraw at any time without penalty. "Please also note that there is no mandatory requirement from Absa compelling you to undertake this research as it is purely on a voluntary basis". All raw data will be treated with confidentiality and will be reported anonymously without identifiers or traceability to either Absa or the individual.

If you have any concerns, you may contact my supervisor or me. Our details are provided below.

Contact details of the researcher:

Boipelo Mosiane
E : 326318@students.wits.ac.za
T : +27 79 823 0567

Supervisor

Ayanda Magida
E : ayanda.magida@wits.ac.za
T : +27 11 717 3953

Outputs

The results of the study will be perceived challenges experienced with big data usage. The results of the study will be shared with participants to see how the overall results came out.

Contact details of HREC administrator and chair – for reporting complaints/problems:

This study has been approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg ("Committee"). A principal function of this Committee is to safeguard the rights and dignity of all human subjects who agree to participate in a research project and the integrity of the research.
If you have any concern over the way the study is being conducted, please contact the Chairperson of this Committee who is Professor Clement Penny, who may be contacted on telephone number 011 717 2301, or by e-mail on Clement.Penny@wits.ac.za. The telephone numbers for the Committee secretariat are 011 717 2700/1234 and the e-mail addresses are Zanele.Ndlovu@wits.ac.za and Rhulani.Mukansi@wits.ac.za

Appendix F: Interview Guide

INTERVIEW GUIDE

Greetings – Welcome and thank you for agreeing to participate. Most of you don't know me.

- Introduction to the participant (My name, what I do, what I am studying)

Introduction

Just a bit of background:

This is an interview to get data for a research project which focuses on challenges preventing the use of big data in risk management. As we may know, risk management is the core of the bank, but seemly the only place that does not harness the power of big data. So, I want to use this interview to find out why that is in your opinion.

- Ask the participant if they have questions.
- Answer the question raised if any and proceed to the confidentiality agreement.

Explanation of the confidentiality clause

- The raw data collected during the interview and any personal information (Including the name of the bank) will be kept private and confidential and will not be shared in the final report.
- The data collected is not affiliated with, nor does it represent the views of Absa and is solely for educational purposes in fulfilment of my Masters' degree with Wits Business School.

Obtain consent to proceed and to record the session.

- Before we start, are you comfortable with me recording this session? Explain that this will help with data analysis and will be discarded when no longer in use.

About the interview –

The questions are split into two sections. The first section will be a demographic-related question, and the last section will be related to the research questions. All the research questions will focus on work being done around using big data in risk management and exploring the challenges within the organization.

Interview Questions

Section 1- Demographics

1. Tell me about the role you play in the bank.
2. How long have you been in the role and how long with the financial institution?
3. What is your understanding of big data?
4. What is your personal experience with using big data?

Section 2- Big data and risk Management

RQ1- What is the perceived role of big data in risk management?

1. In your opinion, how important is big data in the organization.
2. What/ where in risk management can big data be used if at all
3. How do you think Big data can benefit you (help your team/ department/ function to succeed) in your current role as a risk lead?

RQ2 - What challenges are experienced when using big data and what is being done to address these challenges?

1. Should the use of big data be included in the risk management frameworks?
2. Is big data being used in your team/ function, if so, is it adequately used and to what extent?
3. What is the level of readiness for adopting big data usage in your team/ risk functions? What hinders the usage /adoption of big data technologies in risk management practices? In your opinion, are risk management functions mature enough to use big data technologies/ What factors drive the lack of adoption of big data technologies?
4. if not, what can be done to increase the maturity levels.
5. do you feel empowered and enabled to use big data?
6. What are your thoughts on big data being a technology initiative? Do you believe that initiative must be driven from technology or within risk teams?

7. What type of skill does your current team have. Do you think they are empowered to use big data in risk management functions?
8. What legacy plays in the use of big data?

That brings us to the end of the interview questions, thank you for your valuable insights.

- **Ask the participant if they will mind being contacted for further questions.**

Closing

Thank you for your time and have a great day further!

Appendix G: Ethics Clearance

Graduate School of Business Administration
University of the Witwatersrand, Johannesburg

Wits Business School Ethics Committee
Constituted under the University Human Research Ethics Committee (Non-Medical)

Ethics Clearance Certificate

Ethics protocol number: WBS/DB326318/511

This certificate is only valid with a legitimate ethics protocol number and signed by the Researcher (below).

This certificate is only valid if accompanied by formal permission from the relevant stakeholder(s).

Project title	Challenges of big data usage for risk management in a South African bank
Investigator / Researcher	Ms Boipelo Mosiane
Nature of Project	MM (Digital Business)
Decision of the Committee	Approved, provided stakeholders and participants are guaranteed anonymity and confidentiality.
Issue Date of Certificate	2022-12-14
Expiry date	Date of submission of the project / research report
Chairperson	Prof Anthony Stacey ☎ +27 11 717 3587 +27 82 880 4531 anthony.stacey@wits.ac.za

Declaration by Researcher

One copy must be signed by the Researcher and returned to the Chairperson of the Wits Business School Ethics Committee.

I fully understand the conditions under which I am authorized to carry out the abovementioned research and I guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I undertake to resubmit the protocol to the Committee.

Signature

21/12/2022

Date: