

https://dspace.org/

The Case of a South African Bank's Management of Big Data

Kubheka, Phumelele

2024

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/49206

Kubheka, Phumelele . (2024). The Case of a South African Bank's Management of Big Data [Masters dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/49206

Downloaded from a DSpace repository. https://dspace.org/



The Case of a South African Bank's Management of Big Data

A research report 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 by:

Phumelele Kubheka

Supervised by:

Dr Tebogo Sethibe

Johannesburg, South Africa

2024

ABSTRACT

The banking industry generates vast amounts of data, driven by customer transactions, global payment volumes, and increasing mobile and IoT adoption, yet it struggles with siloed structures that hinder effective utilization. Big Data Analytics, when effectively managed, offers the potential for smarter decisions, more efficient operations, and improved customer satisfaction, which is crucial for the banking sector's role in the economy.

The purpose of this qualitative study was to investigate a South African bank's management of Big Data, with a focus on the tools used, the impact of Big Data, and the realised benefits or challenges. Using semi-structured interviews, a sample of 12 participants from Bank X with an interest in Big Data took part. The participants included executives, senior managers, and middle managers. A thematic analysis technique was used to analyse the data collected.

The study showed that Bank X is using industry supported tools, as supported by literature. The study further showed that Bank X faces challenges around data quality and technology skills, although technologies are assessed for suitability. While a single organisational “version of truth” is still a work in progress, hampered by legacy systems, progress is being made at business unit level. Nevertheless, Big Data has boosted revenue at Bank X, evidenced by double-digited product uptake from data-driven recommendations and identification of new sales opportunities.

KEYWORDS

Big Data, Big Data Analytics, Management, Big Data Tools

DECLARATION

I, Phumelele Kubheka, declare that this research report is my own work except as indicated in the references and acknowledgements. 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: Signature:

Signed at ……

On the …….. day of …… 20….

ACKNOWLEDGEMENTS

If someone had told me a few months ago that I would be writing this section of my research report, I would not have believed them. Not because it was an impossible achievement, but simply because I was not in the best space in my academic journey.

Now that I am writing it, it gives me great joy to express my heartfelt appreciation to the people who have shaped this part of my life.

I must begin by thanking uKhathide, uMasobode, uMagonondo, and my grandfather, uBabomkhulu wami, Elias Kubheka. Were it not for him, the love of education and perseverance would not have been instilled in our family or in me personally. He shaped the way we, as a family, view education, and for that I will always be grateful. The sacrifices he made also shaped the man my father would become.

To my father, Joachim Derick Kubheka: no words can truly capture the depth of my gratitude for your unwavering support throughout my life and academic journey. Thank you for being my constant inspiration and voice of reason, for always being there whenever I needed to cry or share my struggles about school. The esteemed educator in you continues to inspire me to strive for excellence both as a professional and a scholar. Thank you for your continued care and for treating me like a princess. I love you more than words can describe.

. To my mother, Mirriam Senkepeng Kubheka, my sweetheart, thank you, Mama, for the comfort food you prepared when times were tough, and I needed a lift. Thank you for always pushing me to achieve my highest potential. More than anything, thank you for your prayers. You are deeply loved and appreciated.

To my sisters, Nqobile Kubheka and Xolile MaKubheka KaMthimunye, thank you for your constant love and support. Nqobi, I am especially grateful for the hot meals you ensured were ready for me when I was burning the midnight oil. Thank you both for reminding me of the importance of reaching this milestone in my

career. Ngiyanithanda boDade. To Sbari Msebenzi Mthimunye, ngiyabonga NoMaziyane.

To my nieces, Bongokuhle and Thingo Mthimunye: I hope that this body of work you have watched me put together inspires you to dream big and reach even greater heights. Aunt Phumi loves you dearly.

To my employer of choice, thank you for always believing in me and providing me with the resources I need to continue growing in my career and within the organisation. Your support has never gone unnoticed. To my research participants, thank you for availing yourselves for the interviews and for your patience throughout the process.

Finally, I would like to express my deepest gratitude to my supervisor, Dr. Tebogo Sethibe. You have been supportive, patient, and empathetic, while also being strict when necessary. I know I would not have come this far without your encouragement and belief in me, even when I doubted myself. You constantly reminded me of the bigger picture, and for that I will always be grateful. Throughout this journey, you have been one of my greatest inspirations.

TABLE OF CONTENTS

TABLE OF CONTENTS 7

LIST OF TABLES 11

LIST OF FIGURES 12

LIST OF ACRONYMS 13

1. INTRODUCTION 14

1.1. Purpose of the Study 14
1.2. Background of the Study 14
1.3. Research Problem 16
1.4. Research objectives and questions 17
1.4.1. Research objectives 17
1.4.2. Research questions 18
1.5. Significance of the Study 18
1.6. Delimitations of the Study 18
1.7. Definition of Terms 19
1.8. Assumptions 19
1.9. Chapter Outline 20

2. LITERATURE REVIEW 21

2.1. Introduction 21
2.2. The South African Banking industry and Bank X 21
2.3. Theoretical Framework 22
2.3.1 Information asymmetry 22
2.3.2 Resource Based View (RBV) and the Learning organisation view 23
2.4. Artificial Intelligence (AI) 25
2.5. Definition of Topic and Background Discussion 25
2.5.1. Big Data Definition 25
2.5.2. Big Data Background Discussion 26
2.5.3. Big Data Analytics Definition 27
2.5.4. Big Data Analytics Background Discussion 27
2.6. Prior Studies 28
2.6.1. Big Data tools 28
2.6.2. Big Data challenges 30
2.6.3. Business Benefits of Big Data 33
2.7. Conclusion of Literature Review 34

3. RESEARCH METHODOLOGY 36

3.1. Research Approach 36
3.2. Research Design 37
3.3. Data Collection Methods 37
3.4. Population and Sample 38
3.4.1. Population (Case Site) 38
3.4.2. Sample and Sampling Method 39
3.5. Research Instrument 40
3.6. Procedure for Data Collection 40
3.7. Data Analysis and Interpretation 40
3.8. Limitations of the Study 41
3.9. Trustworthiness 41
3.9.1. Transferability 41
3.9.2. Credibility 42
3.9.3. Dependability 42
3.9.4. Confirmability 42
3.10. Ethical Considerations 43

4. PRESENTATION OF FINDINGS 44

4.1. Demographics of Participants 45
4.2. Propositions and Final Themes 47
4.3. Findings pertaining to Proposition 1 47
4.3.1. Data Tools 47
4.3.2. Data Privacy 48

a. Security and Access Management 48

b. Data Governance 50

4.4. Findings Pertaining to Proposition 2 51
4.4.1. Addressing Challenges Associated with the Adoption of Big Data. 51

a. Quality Assurance 51

b. Maturity Levels 53

4.5. Findings Pertaining to Proposition 3 54
4.5.1. Customer 54
4.5.2. Revenue and Sales 55
4.5.3. Operational Excellence / Cost savings 56

a. Fraud Detection 56

b. Improved Decision Making 56

4.6. SUMMARY OF THE FINDINGS 57

5. DISCUSSION OF THE FINDINGS 58

5.1. DISCUSSIONS PERTAINING TO DEMOGRAPHICS 58
5.2. DISCUSSION PERTAINING TO PROPOSITION 1 58
5.2.1. DATA TOOLS 59
5.2.2. DATA PRIVACY 59

a. Security and Access Management 59

b. Data Governance and Quality Assurance 60

5.3. DISCUSSION PERTAINING TO PROPOSITION 2 61
5.3.1. ADDRESSING CHALLENGES ASSOCIATED WITH THE ADOPTION OF BIG DATA. 61

a. Quality Assurance 61

b. Maturity Levels 62

5.4. DISCUSSION PERTAINING TO PROPOSITION 3 62
5.4.1. CUSTOMER 62
5.4.2. REVENUE AND SALES 63
5.4.3. OPERATIONAL EXCELLENCE 63
5.5. CONCLUSION 63

6. CONCLUSIONS & RECOMMENDATIONS 65

6.1. INTRODUCTION 65
6.2. CONCLUSIONS REGARDING WHICH TOOLS ARE USED FOR BIG DATA IN THE BANKING INDUSTRY 65
6.3. CONCLUSIONS REGARDING THE CHALLENGES FACED IN THE ADOPTION OF BIG DATA IN THE BANKING INDUSTRY 66
6.4. CONCLUSIONS REGARDING THE REALISED BENEFITS OF BIG DATA IN THE BANKING INDUSTRY 67
6.5. RECOMMENDATIONS 67
6.6. SUGGESTIONS FOR FURTHER RESEARCH 68

APPENDIX A - Participant information sheet 76

APPENDIX B - Participant agreement form 77

APPENDIX C – Consistency Matrix 78

RESEARCH QUESTIONS: WHAT IS THE INFLUENCE OF BIG DATA MANAGEMENT ON A SOUTH AFRICAN BANK? 78

APPENDIX D – Initial Codes and Themes 79

APPENDIX E - Ethics approval notification.....83

APPENDIX F - Permission letter from organization.....84

LIST OF TABLES

Table 1: Definition of terms 18

Table 2: Chapter outline 20

Table 3: Respondent profile 43

Table 4: Participants Summary 48

Table 5: Summary of Propositions and themes 49

Table 6: Consistency table: research questions, propositions, data collection and data analysis 74

LIST OF FIGURES

Figure 1: A resource-based view of the impact of Big Data on competitive advantage (Erevelles et al., 2016) 26

Figure 2: Mind map representation of Big Data tools 34

Figure 3: Thematic analysis; Source MMDB slides cited (Braun & Clarke, 2006) 46

LIST OF ACRONYMS

ACRONYM	FULL NAME
AI	Artificial Intelligence
ANP	Analytical network process
BD	Big Data Analytics
BDA	Big Data Analytics
BDS	Big Data Solutions
CDO	Chief Data officer/ Chief Digital Officer
CIO	Chief Information officer
CRM	Customer relationship management
DM	Data Mining
ISM	Interpretive structural model
IT	Information Technology
RBF	Resource Based Theory
RBV	Resource Based View

1. INTRODUCTION

1.1. Purpose of the Study

This qualitative study investigated a South African Bank's management of Big Data. With a focus on Big Data Tools used, the influence of Big Data and the realised benefits or challenges of Big Data in the bank.

The introductory chapter of this study begins with the research context, which provides the background of the study. This is followed by the research problem, drawn from the context, along with the research objectives and research questions. Subsequently, the significance and delimitations of the study are discussed, followed by the definitions of terms and the assumptions of the study. The chapter concludes with an outline of the chapters that follow.

1.2. Background of the Study

According to IBM (2020), the world generates 2.5 quintillion bytes of data daily, and 90% of the data available today has been created within the past two years. These statistics highlight the exponential growth of Big Data, which will continue to expand, particularly with the adoption of mobile technologies and the Internet of Things (Lochy, 2019). The area of Big Data tools and platforms is also evolving rapidly (Badshah et al., 2024; Mazumder, 2016), outpacing the availability of skilled professionals and creating a talent shortage for effective implementation and management of Big Data systems (Adewusi et al., 2024).

The South African Reserve Bank Working Paper Series (2025) reported about 90% of the adult population in South Africa is banked. The banking industry plays a critical role in a country's economy (Patil & Dharwadkar, 2017). According to the 2023 banking association of South Africa integrated report, the banking industry employs over 180 000 employees. The industry's contribution to corporate income tax is significant, in 2024 they contributed over 40 billion rands (AFRICA, 2017; BASA,

2023). Maunder (2021) further noted that South African banks hold around R6 trillion in deposits.

The banking sector forms part of the economy responsible for safeguarding financial assets (Hall, 2021). One can already anticipate the vast volumes of information that banks hold. This is supported by Lochy (2019), who argues that the banking industry is among the most data-intensive industries globally. Banks hold extensive customer data (e.g., deposits, withdrawals, point-of-sale purchases, online payments, and customer profiles). However, one of the industry's weaknesses is its siloed, product-oriented organisational culture, which limits its ability to effectively leverage these rich datasets (Lochy, 2019). Ku (2020) reinforces this point, noting the significant volumes of structured and unstructured data generated by the global financial services sector.

A key driver of this data eruption is due the rise in global payment volumes. The South African banking industry is rapidly transitioning towards a "marketplace without boundaries", molded by the impending entry of new digital players defying the status quo and pushing exceptional levels of innovation (Maritz, 2018).

Banks inhabit a privileged and influential position. Unlike social media companies, which primarily capture consumer interests, banks have access to customers' actual spending data. Using this data enables banks to provide tailored products and services (Daly, 2020). Deighton (2024) reported in *The Wall Street Journal* that Chase Bank monetises customer spending data (i.e., transaction histories) by selling it to advertisers. The bank is so confident in the value of this data that advertisers only pay when a customer makes a purchase.

Big Data and analytics allow organisations to manage their information more effectively and extract insights that drive growth. These capabilities are increasingly prevalent across industries, leading to smarter decision-making, greater operational efficiency, higher profitability, and improved customer satisfaction (Ochuba et al., 2024; SAS, 2021).

Financial institutions in South Africa have acknowledged the transformative potential of Big Data analytics, particularly in risk management, growing their business and customer service enhancement (Dlamini, 2024).

The context of this study seeks to provide a view of the volume of data generated by Bank X and to explore how analysing and managing this data can generate value for the institution.

1.3. Research Problem

In a study assessing the factors affecting IT professionals' decision to adopt Big Data Analytics, it was found that the perceived usefulness of Big Data Analytics is a significant influence for IT professionals' behavioural intentions to adopt Big Data Analytics (Parson, 2021). However, Cabrera-Sánchez and Villarejo-Ramos (2019), in their study on organisational adoption of Big Data Analytics, revealed that many organisations currently using these tools remain sceptical about their performance. This raises the question of whether such skepticism explains why financial service organisations continue to lag in implementing Big Data Analytics tools (Amakobe, 2015). Six years later, More and Moily (2021) reinforced this concern, noting that financial services are still not adopting Big Data quickly enough and, as a result, are falling behind.

A significant barrier to adoption is that countless financial service firms continue to rely on decades-old software to manage their daily operations (MuleSoft, 2021). As a result, their capacity to integrate new technology is constrained and burdensome (Sternkopf, 2019).

A study on the influence of Big Data Analytics on decision-making processes in South African financial firms found that decision-making in the country has been evolving to become more inclusive and transparent, particularly within the corporate and public sectors. However, the adoption and effective use of Big Data Analytics presents several challenges, including data privacy concerns, high implementation costs, the need for skilled personnel, and integration complexities. Addressing these challenges requires significant investment in Big Data technologies, robust data governance

frameworks, continuous staff training, and supportive regulatory environments (Dlamini, 2024).

While these findings provide valuable preliminary insights, further qualitative research to more thoroughly examine particular challenges, especially those relating to data governance and the implementation of technology, across a broader range of banking organisations. Conducting comprehensive case studies in various bank sizes, organisational structures, and geographic contexts would provide a deeper contextual understanding and more finely grained insights. (Mazumder, 2016). Although Mazumder (2016) study is significant and highlights the Big Data Tools organisations can adopt, the study is generic and not specific to a bank, furthermore the study was conducted in a developed country. There seems to be a gap in research that is specific to the South African banking industry which is characterised legacy systems.

Accordingly, this study investigates the South African banking context, with particular attention to the management of Big Data Analytics, to assess whether the bank is deriving the value it anticipated when adopting Big Data technologies.

1.4. Research objectives and questions

1.4.1. Research objectives

The objective of this study is to investigate a South African bank's management of Big Data, with specific focus on the tools used for data collection and analysis, as well as the realised benefits and challenges.

The study objective is in threefold:

- To evaluate Big Data Tools at Bank X.
- To examine Big Data management Challenges for Bank X.
- To assess the benefits of adopting Big Data at Bank X.

1.4.2. Research questions

The main research question for the study is: **What is the influence of Big Data management on a South African bank?**

The sub questions are below.

- Which tools are used for Big Data at Bank X?
- What challenges are faced in the adoption of Big Data at Bank X?
- What are the realised benefits of Big Data at Bank X?

1.5. Significance of the Study

This study contributes to the broader body of knowledge on Big Data by adding a South African perspective to the largely international discourse. Specifically, it incorporates local findings into the global conversation on Big Data adoption. The South African banking industry stands to benefit from the insights provided, particularly regarding challenges and benefits identified in international literature. Moreover, the study critically reviews the available Big Data tools. As well as shares the experience of a South African bank.

1.6. Delimitations of the Study

This study is limited to a South African bank's perspective on management of Big Data. The focus is on a commercial bank. The delimitations are:

a. Other types of banks, such as mutual banks, co-operative banks, and small finance banks.

b. Alternative approaches to data collection processes.

1.7. Definition of Terms

Table 1: Definition of terms

Term	Description
Big Data	"A form of data with increased volume, difficult to analyse, process, and store using traditional database technologies" (Abdullahi et al., 2020)
Big Data Analytics	"Analysis of large and varied sets of data to uncover hidden patterns, markets demands and trends, unknown correlations, customer preferences and other useful information which helps in reaching major marketing goals" (Parashar, 2020)
Big Data Tools	Big Data tools provide a high-speed engine for Big Data Analysis, combining various processing techniques, such as machine learning and graph processing (Mazumder, 2016).

1.8. Assumptions

The following assumptions have been made regarding this study:

- The participants of the study mirrored standard perspectives and experiences relevant to the research topic.
- The participants of the study are familiar with Big Data concepts.
- The participants responded honestly during the interview and did not demonstrate bias.
- Big Data strategies are adopted by organisations or businesses and the contributors to Big Data are customers or users.

1.9. Chapter Outline

Table 2: Chapter Outline

Chapter	Content covered
Chapter 1: Introduction	Outlines the study's purpose, significance, background, delimitations, assumptions, and research problem.
Chapter 2: Literature Reviews	Reviews the literature, focusing on definitions of Big Data, adoption trends, and the influence of Big Data Analytics.
Chapter 3: Research Methodology	Details the research methodology, including the research approach, design, instrument, and ethical considerations.
Chapter 4: Findings	Presents the findings of the study.
Chapter 5: Analysis of Findings	Provides interpretations and an in-depth analysis of findings.
Chapter 6: Conclusions and Recommendations	Concludes the study and provides recommendations for practice and further research.

2. LITERATURE REVIEW

2.1. Introduction

This chapter critically reviews prior studies in the South African banking industry, Big Data, adoption of Big Data Analytics, and influence of Big Data Analytics in the banking industry. The chapter is divided into five sections. The first section provides context, of the South African banking industry and Bank X. The second section covers the theoretical base of the study, followed by the definitions and background of the key variables. The fourth section covers prior studies in the topic, and the final section concludes the chapter.

2.2. The South African Banking industry and Bank X

The South African Banking industry is regarded by some as the most advanced in the continent. South Africa's banks are integral not only to domestic financial operations but also to regional and international financial networks (Mncube & Viegi, 2024). The sector is characterised by a highly concentrated market structure, robust regulatory oversight, and a growing emphasis on digital transformation and financial inclusion (Cele & Mlitwa, 2024).

The South African Reserve Bank (SARB) is responsible for providing oversight and regulatory frameworks for banks in the Republic, acting as the central bank and resolution authority (SARB, 2025).

There a four major banks in South Africa: Absa, FirstRand, Nedbank and Standard Bank. By the end of 2024, the South African operations of the major banks accounted for 83% of total banking sector assets (PWC, 2025).

This study was focused on one of these major banks. The Bank has a Global presence, with operations in various regions across the continent. It is one of the country's largest banks by assets and operates across several business segments

including retail banking, business banking, corporate and investment banking, and wealth management.

Regionally, the bank has a strong footprint Africa, operating through subsidiaries and representative offices in multiple countries. This presence has allowed it to position itself as a key player in pan-African trade, infrastructure finance, and corporate banking.

The organisation is publicly traded on the Johannesburg Stock Exchange and maintains a healthy capital position. It is widely regarded as systemically important within the South African financial system, playing a critical role in both domestic and regional financial markets. Despite operating in a challenging economic climate, the bank continues to demonstrate resilience through diversified income streams, technological innovation, and a clear strategic vision.

The organisation has a federated operating model, giving each business unit end-to-end accountability for managing and overseeing its operations, while still maintaining some level of group-wide coordination. This study focused on individuals across the bank with an interest in data.

2.3. Theoretical Framework

Three theoretical frameworks are discussed in this section, however only two will form part of the theoretical base for the paper. The frameworks discussed are; the Information Asymmetry Theory, followed by Resource Based View (RBV) / Resource Based Theory (RBT) then the Learning organisation view.

2.3.1 Information asymmetry

A symmetrical business relationship exists when both participants in a transaction have access to the same important information. However, in many transactions, one side has access to additional or superior information than the other, resulting in what is known as information asymmetry (MasterClass, 2020).

Big data strategies, by their nature, increase levels of information asymmetry, which can disempower the very users who contribute to data-driven insights (Mandlik & Kadirov, 2018). Barakat and Sayegh (2021), in their qualitative study on information asymmetry in the age of Big Data Analytics, found that Big Data often exacerbates the problem by enabling a powerful few to access and use knowledge that the majority of consumers cannot. Another factor contributing to asymmetry is the limited information-processing skills of customers (Kshetri, 2014).

The Internet has, to some extent, reduced traditional asymmetries between consumers and organisations (Kshetri, 2014). Several studies indicate that consumers increasingly rely on social media reviews, comparison platforms, and reputation mechanisms to make rational purchasing decisions and to counteract information imbalances(Barakat & Sayegh, 2021). It remains worthwhile to explore how the banking industry is ensuring that Big Data benefits the industry and the customers.

2.3.2 Resource Based View (RBV) and the Learning organisation view

The theoretical foundation of this study is based on the Resource Based View (RBV) / Resource Based Theory (RBT) and the Learning organisation view. RBV is widely used when exploring the influence of technology investments at an organisational level. It argues that valuable, rare, unique, and non-substitutable resources form the basis of competitive advantage (Bharadwaj, 2000). Shan et al. (2019) support this, noting that resources are the foundation of competitive advantage for organisations to guarantee long-term and continued advancement.

Organisational data forms part of these resources. Big Data Analytics is one of the latest technologies to enhance competitive advantage in modern markets (Shan et al., 2019). Insights from Big Data has the potential to enhance business processes and boost business opportunities. It further plays a critical role in improving customer service. Insight from Big Data are critical for helping organisations outperform competitors. Banks in particular host vast amounts of customer data, including spending patterns, which can provide strategic insights.

Big Data is part of the bank's strategic response, as reported in their annual integrated report. The report highlighted that Big Data and advanced analytics have enabled the bank to better understand customer needs and enhance value to clients. In this way, Big Data extends business intelligence and analytics, serving as a tool to execute corporate strategy (Mazzei & Noble, 2019). Figure 1, adapted from Erevelles et al. (2016) a RBV of the impact of Big Data on competitive advantage.

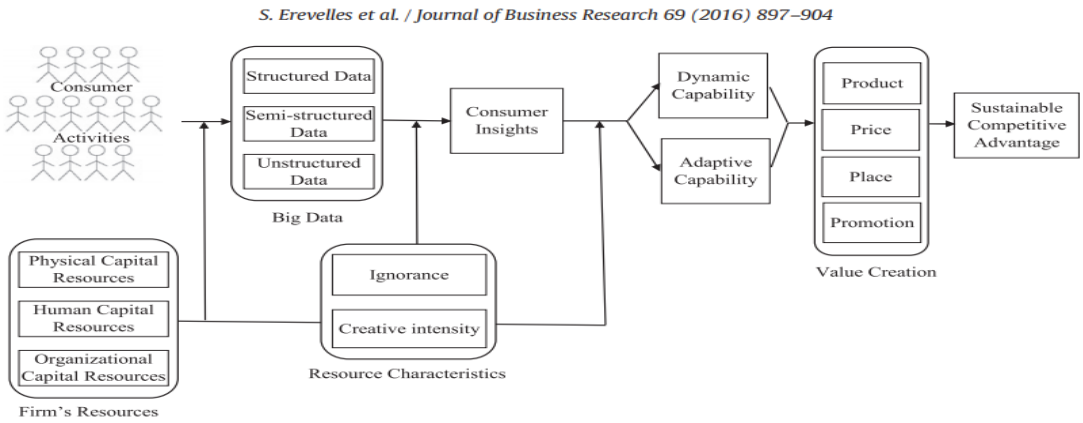


Figure 1: A Resource-Based view of the Impact of Big Data on Competitive Advantage (Erevelles et al., 2016).

Organisations are consistently striving to learn from new data flows. It is therefore relevant to consider the influence of organisational learning as a theoretical framework (Mazzei & Noble, 2019). A learning organisation, as defined by Armstrong(2019) is one that is "skilled at: (i) creating, acquiring, interpreting, transferring, and retaining information and knowledge; and (ii) modifying behaviour to respond to this new knowledge and insight." From the earlier example, it is clear that Bank X has adopted a learning organisation approach, as it uses prescriptive analytics.

Another theoretical framework considered was information theory. As defined by Tutorialspoint (2021), "Information theory is a mathematical approach to the study of coding of information along with the quantification, storage, and communication of information." This framework was not adopted in the study due to its characteristics.

2.4. Artificial Intelligence (AI)

Artificial Intelligence (AI) has emerged as a new way of reframing Big Data (Kumar et al., 2024). AI has existed for more than 60 years and has experienced cycles of growth and decline, often referred to as “AI winters and springs” (Duan et al., 2019). AI is often conceptualised through four goals: systems that think and act like humans, and systems that think and act rationally (Elish & Boyd, 2018).

Burgess (2017) AI as “computers or machines that seek to act rationally, think rationally, act like a human, or think like a human.” Similarly, Burke (2020) defines defines AI as “technology that can perceive, learn, and reason to extend the capabilities of people and organisations.”

Ertel (2018) defines AI as “the ability of digital computers or computer-controlled robots to solve problems that are normally associated with the higher intellectual processing capabilities of humans”.

The world was engulfed in Big Data before it even recognised the concept (Maryville, 2021). IT professionals quickly realised that human minds could not handle the task of examining huge volumes of data and converting information into usable formats to improve business decisions. To handle this, AI algorithms were developed to generate insight from complex datasets (Maryville, 2021).

2.5. Definition of Topic and Background Discussion

2.5.1. Big Data Definition

Big Data has been defined in several ways. Constantiou and Kallinikos (2015) define it as “large volumes of data generated and made available online and in digital media ecosystems.” Baig et al. (2019) similarly describe it as “large volumes of information.” Abdullahi et al. (2020) expand the definition, stating that “Big Data is information with augmented volume, hard to analyse, process, and store using traditional database technologies. ”Big Data refers to “datasets whose size is beyond the capability of

typical database software tools to capture, store, manage, and analyse" (Venkatesh et al., 2019).

(Tole, 2013) defines Big Data through its characteristics which are volume, velocity, variety, and veracity. Volume refers to "the quantity of data that is being manipulated and analysed in order to obtain the desired results." Velocity "is the speed data travels from 1 point, to point another." Variety "speaks to the type of data that is warehoused, analysed and utilised." Value "is the quality of data warehoused, analysed and utilised." Veracity "is the fifth characteristic of Big Data and came from the idea that the possible consistency of data is good enough for Big Data" (Tole, 2013).

2.5.2. Big Data Background Discussion

In recent years, the concept of Big Data and its role organisational decision-making has gained significant attention (Mikalef et al., 2016). While the concept is relatively new, data collection and evaluation have long existed. Before computers, information was stored in paper-based registers (Tole, 2013). By the time the term "Big Data" was coined, large volumes of recorded information already existed, which, if carefully analysed, could provide valuable insights (Maryville, 2021).

Big Data offers substantial benefits but also presents challenges, as many fields now face overwhelming volumes of data generated at high speeds (Hassani et al., 2018).

Factors associated with Big Data include diversity, fluctuating quality, frequency of updates, and rapid expansion (Krogstie & Gao, 2015).

Many organisations recognise that the data they own or can differentiate them from competitors and provide a competitive edge if used effectively (Wu & Wang, 2024) As a result, companies today aim to collect and process as much data as possible (Mikalef et al., 2016).

Big Data adoption offers industries prospect to outperform competitors. Although adoption is lengthy and costly, the benefits justify the investment (Baig et al., 2019). Successful adoption, execution, and management of Big Data Analytics requires organisations to upskill employees or recruit individuals with the required expertise

(Cabrera-Sánchez & Villarejo-Ramos, 2019). Research by Khajouei et al. (2017) further indicates that technology adoption is strongly influenced by organisational culture.

2.5.3. Big Data Analytics Definition

Erevelles et al. (2016) define analytics as “tools that help find hidden patterns in data”. Gandomi & Haider (2015) describe analytics as techniques used to analyse and acquire intelligence from big data. Thus, big data analytics can be viewed as a sub-process in the overall process of ‘insight extraction’ from Big Data.

Behl et al. (2019) defines Big Data Analytics as a“mixture of techniques and technologies that businesses can adopt to analyse large volume of data which could be complex and diverse”. Abdullahi et al. (2020) describes it as“the process that involves using algorithms and software tools to extract useful business information from the dataset”. Parashar (2020) defines Big Data Analytics as “the analysis of large and varied sets of data to uncover hidden patterns, markets demands and trends, unknown correlations, customer preferences and other useful information which helps in reaching major marketing goals.”

2.5.4. Big Data Analytics Background Discussion

Gandomi & Haide, (2015) further indicates that technology adoption is strongly influenced by organisational culture.

Srivastav (2021) conducted a theoretical study highlighting the meaning of Big Data analysis in the banking industry. The study categorised analytics into four types: Descriptive Analytics, Diagnostic Analytics, Predictive Analytics and Prescriptive Analytics.

- Descriptive analytics is the first stage of analytics as it is known as the most fundamental type of analytics. In this stage, the large sum of data is broken down into an understandable form. Descriptive analytics encompasses the use of descriptive statistics such as arithmetic operations, mean, median,

percentage etc. Business intelligence (BI) is the usage of descriptive analytics to make raw data comprehensible to stakeholders. The primary goal is to comprehend the underlying behavior or trend of the data as opposed to reaching any definitive predictions or assumptions. The banking sector relies significantly on this type of analytical method because it allows them to easily analyse prospective customers' purchasing patterns, investing activity and savings tendency.

- Descriptive Analytics breaks down large datasets into an understandable form, often through descriptive statistics (e.g., mean, median, percentages). Business Intelligence (BI) is a form of descriptive analytics that makes raw data comprehensible to stakeholders.
- Diagnostic Analytics uses historical data to answer "Why has something happened?" While useful for identifying causes of past issues, its application is limited.
- Predictive Analytics seeks to forecast future events by analysing patterns in historical data. Its accuracy depends on data quality.
- Prescriptive Analytics addresses "What action should be taken?" It combines methods from the other three types and is considered the most advanced form, often relying on machine learning, R programming, or Python. It is customer-centric, integrating internal and external data with feedback to provide optimal solutions.

2.6. Prior Studies

2.6.1. Big Data tools

Nisar et al. (2021) in their study on Big Data and decision-making, found that Big Data tools are essential for improving operational efficiency, enhancing decision-making, and driving innovation in large organisations.

Duan et al. (2019) conducted a theoretical study reviewing 58 documents to identify challenges and opportunities in AI-based systems for decision-making. The study

found that advances in computing power and Big Data technologies have significantly enhanced the power of AI.

Gupta et al. (2018), in a review of 16 papers, examined historic, current, and future directions of Big Data and cognitive computing. They found that AI makes Big Data meaningful through cognitive computing, as analysing Big Data manually is often too time-consuming, necessitating the use of AI techniques.

Surya (2015) conducted an exploratory study on AI in the Big Data era, concluding that AI enables users to automate and improve predictive and descriptive analysis, tasks that would otherwise be monotonous and resource intensive. The study also found that organisations, particularly banks adopting AI and machine learning, are realising significant advantages, including smarter decision-making and better risk management.

Mazumder (2016) , in his book on Big Data Tools, provided an overview to guide organisations in selecting suitable tools. He highlighted that the dynamic nature of the industry has created a wide variety of technologies, adding to the confusion faced by specialists. Computing tools, such as machine learning and graph processing, are essential because they provide high-speed engines for Big Data Analysis.

Venkatesh et al. (2019) conducted a theoretical study on challenges and tools in Big Data Analytics, identifying Apache Hadoop as one of the most influential and established tools. They noted that Big Data platforms are particularly useful for uncovering complex trends and patterns not easily detected using traditional methods. By implementing such systems, organisations can gain powerful insights from existing data. They further explained that combining multiple tools allows for specialised processing across data types, leading to improved management and more accurate insights.

During the data collection phase, Big Data governance is considered. By putting governance in place at this stage, data is collected responsibly while preserving security, quality, and regulatory compliance (Ladley, 2019) Using governance processes such as data quality management provides guidelines and practices to

guarantee the accuracy, consistency, and completeness of the data gathered(Janssen et al., 2020) However, growing privacy concerns about handling Big Data have created new obstacles, highlighting the difficulty of balancing privacy protection with the need to use data efficiently (Bertino & Ferrari, 2017)

Findings from the systematic review on Big Data benefits, challenges and tools by (Almeida, 2017) showed that Big Data Tools can be categorised into three parts using a mind map: (a) computing tools, (b) storage tools, and (c) support technologies (summarised in Figure 2). Furthermore, it was discovered that Hadoop MapReduce is a popular tool, because not only does Hadoop MapReduce process high volumes of data but also has the functionality of handling variety. Apache is an extension of the Hadoop's MapReduce framework, it is often used with data visualisation tools for dashboarding (Almeida, 2017).

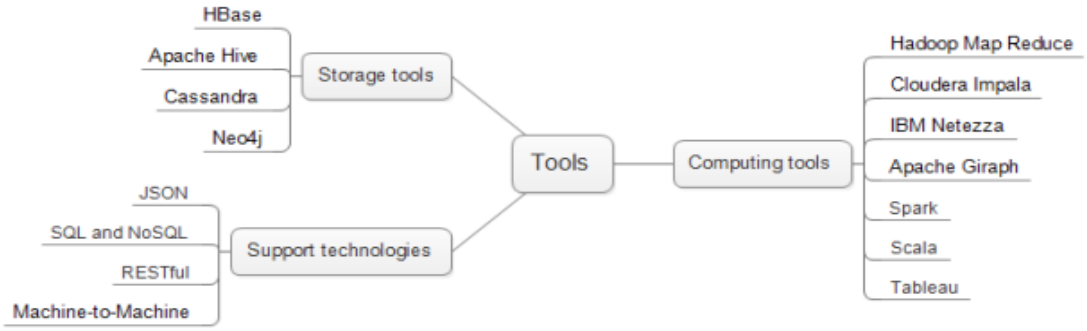


Figure 2: Mind Map Representation of Big Data Tools

Source: Almeida (2017)

Proposition 1

Data tools and Data privacy are essential elements for Big Data Management.

2.6.2. Big Data challenges

Big Data's 4Vs (volume, velocity, variety, and veracity) present multiple difficulties, including efficiently managing and organising data, developing new business models, creating innovative solutions to conventional financial problems, integrating complex structured and unstructured data, and ensuring security and privacy. Fundamental

research in finance and data analytics is needed to address these challenges(Sun et al., 2019).

Chaoui et al. (2022) noted that key aspect of data lifecycle management is having clear processes to govern data from its inception and storage to its eventual archiving or removal. Data retention policies, driven by legal and commercial requirements, are crucial for ensuring data is handled correctly at every point in its lifecycle.

Raguseo (2018) conducted a quantitative study with a purpose of investigating the adoption levels of Big Data technologies in companies. The study was conducted using a sample of 200 medium and large sized French firms. The most frequently identified risk associated with the deployment of Big Data technology was privacy.

Salleh & Janczewski (2019) conducted a qualitative study identifying security-related considerations made by organisations adopting Big Data Solutions (BDS). The foremost security concerns included challenges in acquiring data, the adequacy of legacy security mechanisms, managerial security awareness, top management support, security education, training and awareness (SETA), security personnel skills and experience, employees' perceptions of information sensitivity, regulatory compliance, vendor reputation, and environmental uncertainties.

Behl et al. (2019) developed an interpretive structural model (ISM) as a framework for efficient implementation of Big Data Analytics. From a sample of 23 respondents, the strongest enablers identified were access to relevant data and employees' technical skill sets. The study also found an encouraging link between the ranking of enablers emerging from the ISM and the analytical network process.

Cabrera-Sánchez & Villarejo-Ramos (2019) conducted a quantitative study explaining the adoption and use of Bid Data Analytics by companies, with a focus on understanding implementation challenges using a sample of 199 respondents. The findings showed that robust infrastructure outweighs many implementation challenges. Interestingly, organisations already using Big Data Analytics were still sceptical about its performance.Indriasari et al. (2019) conducted a qualitative study on applying AI and Big Data Analytics in banking to enhance customer experience.

The study showed that banks need to move away from legacy systems and embrace digital transformation. Banks' responses to frontier technology adoption depend on (a) current state, (b) strategy, (c) customer focus, (d) organisational capabilities, (e) brand promise, (f) regulatory requirements, and (g) capital constraints.

Amakobe (2015) qualitative study states that financial service organisations are lagging in their implementation of Big Data Analytic tools, which indicates an unexploited possibility for create value. This was supported by More and Moily (2021), who reported that financial services are not adopting Big Data quickly enough and are therefore falling behind.

In a quantitative study with 405 respondents, aimed at demonstrating the importance of Big Data analytics in supporting world-class sustainable manufacturing (WCSM). Dubey et al. (2016) found that developed countries have greater experience with Big Data tools, while developing countries are still progressing or at the infancy stage of adoption. The study also produced a framework to investigate the importance of Big Data Analytics for WCSM.

Often described as "data about data," metadata provides descriptive information about datasets, making them straightforward to find, understand and manage. This typically contains details such as where the data comes from, its format, originator, creation date, modification history, and access rights. Metadata is crucial for data management and quality, and for supporting data discovery and usability. It helps users and systems understand the context, structure, and content of data, enabling its efficient application across fields like databases, data governance, and digital preservation (Riley, 2019)

Côté et al. (2024) highlight that to achieve accurate, complete, and consistent datasets for data analysis and machine learning, data cleaning is essential. Implementing best practices in data cleaning elevates data quality, producing more reliable and impactful analytical findings.

Proposition 2

Addressing challenges that hinder adoption is essential for successful Big Data Management.

2.6.3. Business Benefits of Big Data

Abdullahi et al. (2020) conducted a qualitative study that adopts Big Data Analytics processes to examine the disruption in the banking industry. It was discovered that the emergence of Big Data in banking industry has led to exponential technical improvements to the industry.

Raguseo (2018) conducted a quantitative study on the adoption levels of Big Data technologies in organisations and the tools used, with a sample of 200 medium and large French companies. The study highlighted frequently recognised strategic, transactional, transformational, and informational advantages linked to Big Data technologies. Key findings included: the most frequently recognised transactional benefit was enhanced productivity growth; strategic benefits centred on delivering superior products and services; transformative benefits related to capability development through new skills and hiring; and informational benefits focused on data management, accuracy, and accessibility. The most commonly cited risk was loss of privacy.Géczy, (2014) conducted a theoretical study exploring interdisciplinary characteristics of Big Data at the intersection of technological and operational enablers. The study defined the Big Data problem as a gap between the required and available organisational capabilities to manage collected data effectively. Big Data Analytics also helps organisations confront legacy issues and opens opportunities to explore newly emerging scientific domains.

Mandlik & Kadirov (2018) conducted a qualitative study advocating a readjustment of current management approaches given organisations' growing reliance on Big Data strategies. The findings showed that Big Data strategies have improved services in the financial sector, including reducing fraudulent transactions and protecting vulnerable consumers from identity theft.

Mikalef et al. (2016) presented a theoretical discussion leading to a research framework explaining how Big Data can drive competitive performance. The findings suggest that, when properly implemented, Big Data initiatives allow firms to create new commercial value and improve productivity and innovation.

Amakobe (2015) conducted a qualitative study on the impact of Big Data in banking, emphasising the need for continued investment. The findings indicated that Big Data can unlock marketing potential and that advanced analytics help banks better manage rising compliance costs and the risk of non-compliance.

Mungai and Bayat (2018) conducted a mixed-methods study evaluating digital disruption from Big Data in South Africa's financial services industry, with a focus on financial inclusion. Using secondary data from the World Bank Findex database and semi-structured interviews, the study found that Big Data has changed the banking industry by contributing to employment growth, improved labour productivity and risk management practices, greater economic contribution, innovation, and enhanced financial inclusion.

Parashar (2020) conducted a qualitative study on the features, significance, and uses of Big Data Analytics in banking, analysing transactional and sentiment data. The study found that insights from Big Data inform market observations and help banks deliver better services to internal and external customers, while also strengthening both active and passive security systems.

Proposition 3

Effective Big Data management has a positive influence on business outcomes.

2.7. Conclusion of Literature Review

The theoretical foundation of this study follows the Resource-Based View (RBV)/Resource-Based Theory (RBT) and the Learning Organisation View. The literature highlights the synergy between AI and Big Data, making AI one of the most significant enablers within Big Data (Kumar et al., 2024). Big Data and AI are inextricably linked: the former is useful when implemented correctly, and the latter relies on the former to succeed (Surya, 2015). From the literature reviewed, the benefits and challenges of Big Data are evident; however, the issue of management persists. Rossi (2022) characterises Big Data management through five key elements:

data policies and governance, specific technologies, organisational change, data access and storage, and infrastructure.

The primary objective of this research is to investigate a South African bank's adoption of Big Data Analytics. The secondary objective is to determine the influence of Big Data and the realised benefits and challenges within the bank.

3. RESEARCH METHODOLOGY

Chapter 1 introduced the study, while Chapter 2 covered the literature review. This section of the study now focuses on the research methodology that was followed in the study. The topics discussed in depth include the research approach, research design, data collection method, population and sample research Instrument, data analysis and interpretation, study limitations as well as the transferability and dependability and the study's ethical considerations. The study is qualitative in nature and the research design is semi structured interviews conducted virtually. The sample of the study are bank employees at different levels.

3.1. Research Approach

This study explored the adoption of Big Data Analytics by a South African bank, with the purpose of understanding its influence, benefits, and challenges. A qualitative methodology was used, as it is designed to provide a rich, contextualised picture of the management of a social phenomenon or business problem (Creswell, 2018). Wahyuni (2012) further supports this, noting that qualitative data provides rich descriptions of social constructs.

Qualitative research typically takes place in natural settings, where human behaviour and events occur (Creswell, 2018). The focus is on participants' insights and experiences, allowing the researcher to understand multiple realities (Creswell, 2018). This approach places emphasis on both the process and the outcome, with researchers particularly concerned with identifying how things occur. (Creswell, 2018).

In contrast, a quantitative methodology would not be suited for this study, as it is generally used when one is interested in testing or confirming a theory (Creswell, 2018). Quantitative research is primarily concerned with the quantity, frequency, or magnitude of a phenomenon, forming the basis of statistical analysis (Schindler, 2019). As this study is not testing a theory, a qualitative approach was deemed appropriate.

3.2. Research Design

Research designed are “types of inquiry within qualitative, quantitative, and mixed methods approaches that provide specific direction for procedures in a research study” (Creswell, 2018). The study design adopted for this paper is a case study, since the purpose of the study is to attain multiple viewpoints of a single organisation at a point in time (Schindler, 2019). The study was conducted in a natural setting for ease of soliciting information from the interviewees. As the key instrument in the study, the researcher examined multiple sources of data (Creswell, 2018) to answer the research question.

3.3. Data Collection Methods

Data was collected through virtual semi-structured interviews, in line with COVID-19 protocols. Schindler (2019) describes semi-structured interviews as "individual depth interview that starts with a few specific questions and then follows the individual’s tangents of thought with interviewer probes; the questions generally use an open-ended response strategy Wahyuni (2012), referencing Bradford and Cullen (2012), notes that qualitative semi-structured interviews are among the most dominant and widely used data collection methods in the social sciences.

Virtual interviews via video conferencing offered similar benefits to face to face interviews, as the researcher could observe and record nonverbal as well as verbal behaviour (Schindler, 2019).

The data collection process followed these steps:

- The researcher contacted identified participants telephonically or via email to summarise the purpose of the study and invite them to participate.
- Participants who agreed to take part were sent meeting invites along with the questions in advance, allowing them to prepare.
- The interviews were conducted virtually using platforms such as Microsoft Teams or Google Hangouts.

- Sessions were video recorded, and the interviewer also used a notepad to capture discussion points.

3.4. Population and Sample

3.4.1. Population (Case Site)

Schindler (2019) defines a population as “people, events, or records that possess the desired information to answer the research question.” Similarly, Gray (2019, p. 227) described population as “total number of possible units or elements that are part of the study.”

The population for this study was the bank itself, with the population consisting of executives and managers with an interest in Big Data. The bank has a global presence, with operations across the African continent, and is one of the largest South African banks by assets. It operates across multiple segments, including retail banking, business banking, corporate and investment banking, and wealth management.

Its cross-border footprint has positioned it as a key player in pan-African trade, infrastructure finance, and corporate banking. The bank’s strategy is underpinned by a purpose-led ethos and a customer-centric business model. Operating under a federated model, each business unit has end-to-end accountability, while still maintaining group-wide coordination. Big Data plays a critical role in this strategic response.

Therefore, this study focused on individuals with an interest in data across the bank. Insights from bank leaders were valuable additions to the body of work, and the findings will benefit the organisation as it continues to explore ways to create customer value.

3.4.2. Sample and Sampling Method

A sample is defined as “a group of cases (participants, events, or records) consisting of a portion of the target population, carefully selected to represent that population” (Schindler, 2019).

Over 20 participants were initially contacted. However, due to availability and suitability, 12 participants were interviewed. This aligns with findings from Bernard (2024) and Squire et al. (2024), which show thematic saturation is often achieved within 9 to 17 interviews. Similarly, Kubat and Öztürk (2025) successfully employed a sample of 12 in their study on perceptions, challenges, and opportunities in qualitative research. Creswell (2018) also recommends between 5 and 25 participants, further supporting the adequacy of the sample size.

The participants were executives, senior managers, and middle managers with an interest in Big Data. They reviewed the research instruments beforehand and were sufficiently experienced to provide meaningful insights. The respondent profile is shown in Table 3.

Table 3: Respondent Profile

Role Level	Number to be sampled
Executive (Chief Data and Analytics Officer, Chief Architecture, Product Owner, Head of Data Science etc.)	5
Senior Managers (Data architecture, Data science & engineering, Product engineer etc.)	4
Middle Managers (Data Analysts, Product Engineer etc.)	3

3.5. Research Instrument

The research instrument was an interview schedule divided into three sections, covering the tools used for Big Data in the bank and the realised benefits and challenges.

3.6. Procedure for Data Collection

The data collection process followed these steps:

- The researcher contacted identified participants telephonically or via email to explain the purpose of the study and invite them to participate.
- Participants who consented were sent meeting invites with the interview questions in advance.
- Interviews were conducted virtually via Microsoft Teams or Google Hangouts.
- Sessions were video recorded, and notes were taken to capture key points.

3.7. Data Analysis and Interpretation

Thematic analysis was used for the collected data. Braun & Clarke (2006) define thematic analysis as “a method for identifying, analysing and reporting patterns (themes) within data”. Tracy (2010) describes that transferability “is achieved when readers feel as though the story of the research overlaps with their own situation, and they intuitively transfer the research to their own action”. The data analysis process in this study followed the below stages summarised by (Braun & Clarke, 2006)).

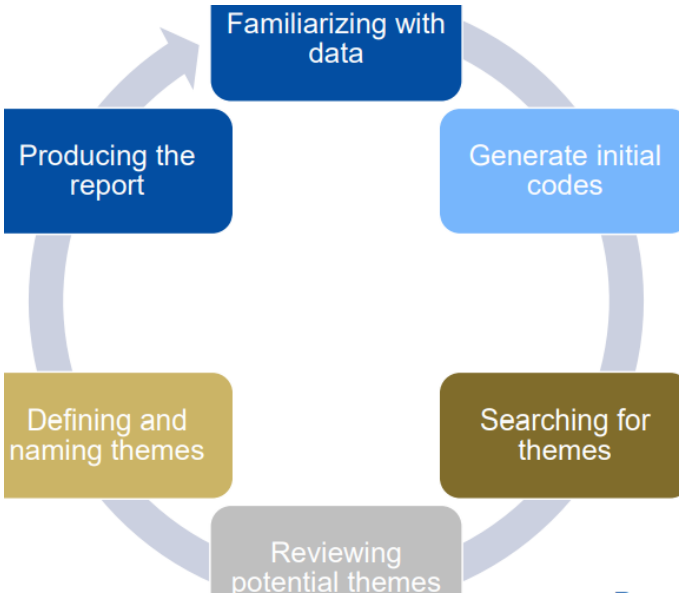


Figure 3: Thematic Analysis: Source MMDB slides cited (Braun & Clarke, 2006).

3.8. Limitations of the Study

- Due to participant availability, the study was conducted virtually using Microsoft Teams.
- Time restrictions were present as the study was conducted for assessment purposes.

3.9. Trustworthiness

3.9.1. Transferability

Bitsch (2005) describes transferability as “the degree to which the results of qualitative research can be transferred to other contexts with other participants, it is the interpretive equivalent of generalizability.” Tracy (2010) further explains that transferability “is achieved when readers feel as though the story of the research overlaps with their own situation and they intuitively transfer the research to their own action”. To ensure transferability, the study’s data collection process was thoroughly documented so that the findings can be applied in other contexts, such as in different banks or organisations outside the financial sector.

3.9.2. Credibility

Anney (2014) defines credibility as “the confidence that can be placed in the truth of the research findings.” The researcher adhered to the university’s ethics framework when conducting the study. The research was reviewed by the assigned supervisor and panel, with regular engagements between the supervisor and researcher. These checkpoints ensured validation of the adequacy of the sample and the appropriateness of the research process. They served as key supporting pillars for establishing credibility.

3.9.3. Dependability

A systematic approach was followed throughout the study to ensure dependability and auditability. The research process was supported by an external audit conducted by the supervisor, who evaluated the data collection process. This evaluation also examined whether the findings, interpretations, and recommendations were supported by the collected data.

3.9.4. Confirmability

The final of Lincoln and Guba’s (1986) four trustworthiness criteria is confirmability, which examines the degree to which findings can be confirmed by others. Morse (2015) defines confirmability as “using strategies of triangulation and the audit trail.” Ahmed (2024), in *A Practical Guide to Improve Trustworthiness of Qualitative Research for Novices*, suggests that confirmability is present when findings are not influenced by researcher bias but are drawn exclusively from participants’ responses.

In this study, triangulation was achieved by interviewing participants across management levels with varied years of service, from different departments, and by reviewing the bank’s annual integrated report. This approach ensured that findings were not based on a single group and that interview insights could be corroborated with official organisational data. This process grounded the study’s findings in participants’ experiences rather than the researcher’s assumptions.

3.10. Ethical Considerations

- The researcher adhered to the University of the Witwatersrand's ethical standards.
- Informed consent was obtained, with participants required to complete consent forms.
- The privacy of participants data was been respected.
- The researcher signed the university's plagiarism declaration form.

4. PRESENTATION OF FINDINGS

This chapter provides a summary and context of the organisation, presents participant demographics, and outlines the findings collected through the research methodology described in Section 3.

A thematic analysis technique was used to analyse the data collected. Braun & Clarke (2006) define thematic analysis as “a method for identifying, analysing and reporting patterns (themes) within data”.

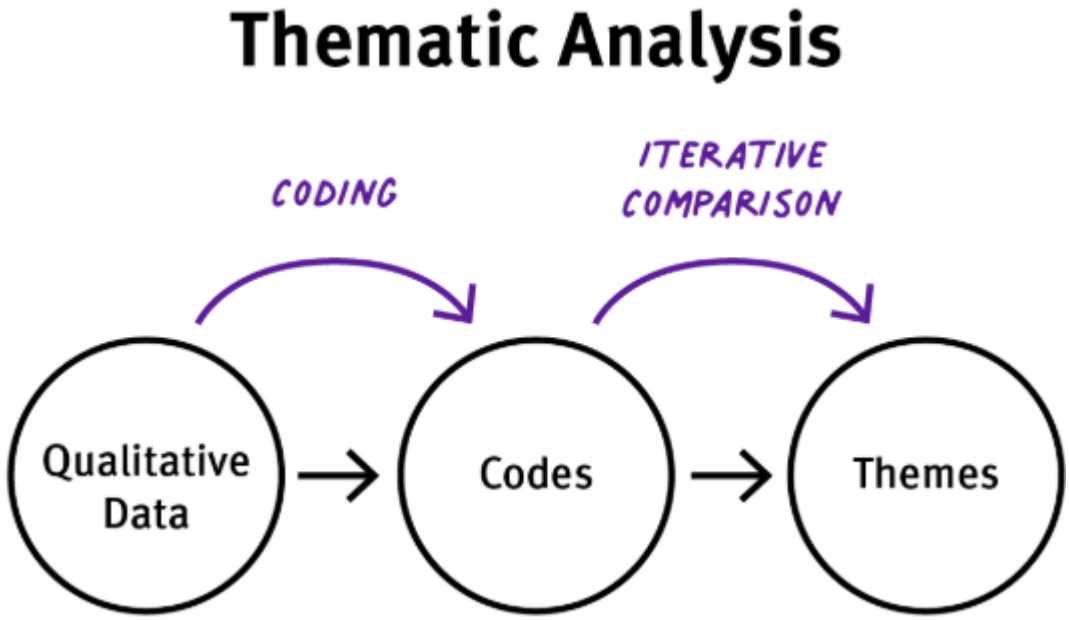


Figure 4: Diagram representation of a Thematic Analysis: Source Nielsen Norman Group

The researcher first familiarised themselves with the data by listening to the recording of and reading through the transcripts. Appendix D of the document provides a view of the initial codes generated from the data. When this process was concluded, the next step included, creating initial themes based on the codes. These themes were then reviewed and the final step being summarising these themes to create final themes and subthemes.

4.1. Demographics of Participants

The study was conducted in Bank X with 12 participants who had an interest in Big Data and took part in semi-structured interviews. The demographic profile includes role titles, corporate levels, tenure at Bank X, business areas, and experience in the field of study.

The participants represented different management levels and had varied years of service at Bank X. Of the 12 participants, five were executives with 4–20 years of service, four were senior managers with 3–11 years, and three were middle managers with 2–12 years. The role titles included Chief Data and Analytics Officer, Chief Architect, Product Engineer, and Lead Data Engineer.

Table 4: Participants Summary

Code	Role titles	Corporate levels	Tenure (Years)	Business Area
P1	Head of Technology: Data Management & Governance	Executive	7	Group Chief Analytics Office
P2	Product Owner	Executive	15	Chief Data and Analytics Office - Retail Function
P3	Data Specialist	Senior Manager	8	Chief Risk Office
P4	Head of Data Science	Executive	7	Chief Analytics Office – Retail Function
P5	Chief Data and Analytics Officer	Executive	20	Retail function
P6	Product Engineer	Middle Managers	2,5	Service Engineering
P7	Lead Data Engineer	Middle Managers	12	Service Engineering – Data Science and Engineering
P8	Head of Data Architecture	Senior Manager	5	Regional Operations
P9	Chief Architecture	Executive	4	Retail and Insurance
P10	Senior Product Engineer	Senior Manager	3	Chief Operations Office
P11	Data Engineer	Senior Manager	11	Regional Operations
P12	Data Analyst	Middle Manager	2	Corporate Functions

4.2. Propositions and Final Themes

Table 5: Summary of Propositions and themes

Proposition	Themes and sub themes

Proposition 1: Data tools and data privacy are essential elements for Big Data Management.	• Data Tools
	• Data Privacy
	  ○ Security and Access Management
	  ○ Data Governance

Proposition 2: Addressing challenges that hinder adoption is essential for successful Big Data Management.	• Addressing challenges associated with the adoption of big data.
	  ○ Quality Assurance
	  ○ Maturity Levels

Proposition 3: Effective Big Data management has a positive influence on business outcomes.	• Customer
	• Revenue / Sales
	• Operational Excellence / Cost saving
	  ○ Fraud detection
	  ○ Improved decision making

4.3. Findings pertaining to Proposition 1

This sub-section presents the findings on the essential elements for Big Data Management identified by the participants. The categories and themes are presented in this section, as well as the views expressed on the proposition below:

Proposition 1: *Data tools and data privacy are essential elements for Big Data Management.*

4.3.1. Data Tools

Participants demonstrated strong awareness of the data tools used at Bank X. They explained that while the bank uses a range of data collection and storage tools, the

enterprise-wide tool of choice is Hadoop. Other tools include both in-house solutions and off-the-shelf platforms such as SAS, SQL, Databricks, Amazon Web Services (AWS) Lambda, and MongoDB. Tool selection depends on factors such as legacy systems, specific business unit requirements, platform capabilities, and cost.

P5: *Big data collection actually starts in the line of business systems*, if *you look at it from an ingestion perspective, we have a number of our own sort of inbuilt proprietary tools that we've built for collection. Over the years, we've used lots of let's call them off-the-shelf tools, but I think we sit with a unique challenge at Bank X that we've got quite a lot of different systems."*

P7: *Our team specifically, we use AWS Lambda, which is a serverless AWS native service. So that's to pull the data and we use S3 to store the data."*

P12: *Of course, other business areas would have other tools they use in addition, but Hadoop is enterprise wide."*

P11: *We use Hadoop because it's an open-source framework tool which assist us to be flexible on what we need to do and it's easy, it's got enough outsourcing community who assist us for free sometimes, making it cost effective."*

4.3.2. Data Privacy

a. Security and Access Management

All participants unanimously indicated that the bank has sufficient precautions in place to prevent unauthorised access to data during the collection and curation process. This is achieved through measures such as:

- Logical access management
- Data encryption
- Periodic security reviews

- Segregation between production and user acceptance testing (UAT) environments

P3: *The bank has its own in-house built access tool which is used to manage access. It integrated with any new tool that is being acquired, because you cannot always change every time there's a new tool. For a user to get access to a certain data set which is already sitting on Hadoop, the access management process guards who has access, because if I'm a data owner of something that is already sitting on Hadoop, the data management process puts me in the workflow to approve or not to approve someone because as a data owner, I'm accountable for that data and I'm accountable to see who is has access and they must justify why they need access and I have to approve as a data owner."*

P7: *In terms of, securing the data. All our data is encrypted using AWS native services. We encrypt our data using KMS and then we've also got IAM roles and policies to make sure that only people that have the correct permissions can access the data."*

P11: *Data is transmitted securely between landing zones and data touch points."*

Majority of participants indicated that the tools and architecture chosen to assist with data security and access management are generally decided by the Security Office of the bank, however, there are also additional governance committees that approve new technology in the Bank. Although the participants do not have conflicting views on security, one respondent pointed out that the process is unpleasant and time consuming, which affects their productivity and the speed at which they deliver solutions. This suggests that while the security strategy may be well-implemented, and secures data at the collection and processing level, the centralized security approach might be causing delays.

P5: *I've got no choice on security tools chosen. We've got an entire Security Office in the group that looks after that. We follow their guidance to the letter."*

P1: *There are a few Governance committees one has to go through before a solution is adopted in the organisation, such as the cloud governance forum cloud-based*

solutions. And then you must go through group architecture council for any new tool. Then there are different specialists like the security and risk specialists and so on, who also need to sign off on the tool.

P10: *There are several forums one must seek approval if you want to change something and it's actually very painful. If you need to change something, let's say you need to add a new component to Data Brick. You must go to an architecture forum and then you have to go to a cloud forum and then you go to a security forum, further delaying delivery."*

b. Data Governance

In the context of this study, data governance refers to the processes related to access management during data collection, the tools used for data collection and processing, and the challenges the organisation faces.

Most participants emphasised the importance of data governance and quality assurance, particularly at the collection stage. They highlighted the use of data management principles to ensure data quality and compliance. To maintain consistency and standardisation by the time data is ingested, quality assurance tests are applied across the value chain through the integration of metadata.

P1: *Using the structured data that metadata provides to evaluation and track data quality, which makes it easier to say are you working with the right thing and you should use that"*

P2: *We make use of data management principles So we ensure that data is monitored for quality and compliance at source as far as possible, we apply quality assurance checks as far up the value chain as possible so that by the time we ingest, it is in a relatively consistent format, relatively standardized, and makes it easier to work with. Although it's inevitable that there will be some level of manipulation of data to finesse it into a an easily work workable form."*

P3: *Governance is quite big in a sense that, for each and every data set which is*

sitting on Hadoop as the as the tool of choice. There there's couple of governance activities like access and data lineage. As an example, if there's something that is being reported, where is the data coming from? How is that data being guided? How is the processing? Who owns that data? Who's the data steward? Who's the data owner for that database? Governance is considered for Hadoop from both a lineage and from a security perspective"

4.4. Findings Pertaining to Proposition 2

This sub-section presents the findings on the influence Big Data Management has on business outcomes. The categories and themes are presented in this section, as well as the views expressed on the proposition below:

Proposition 2: *Addressing challenges that hinder adoption is essential for successful Big Data Management.*

4.4.1. Addressing Challenges Associated with the Adoption of Big Data.

a. Quality Assurance

The participants had contrasting views regarding data cleaning. Some respondents noted that cleaning data at the source is costly; therefore, the organisation often employs workarounds to clean data in the data lake rather than at the source. ***P9:*** *Obviously it will be costly to fix data at source since we extract from multiple data sources, because when you deal with that you need to deal with the reverse engineering to see where the data came from."*

Whilst other participants cited that … ***P5:*** *We can't have a situation where a big data system gives a different answer to a source system. But what we do is we track the quality of the data on the big data system and that feeds into initiatives at source to fix it at source."*

The participants reported a range of challenges encountered during data collection and processing at the bank. These included issues of data discoverability,

understanding data elements, accurately transcribing voice to text due to insufficient natural language processing capabilities, lack of a unified data taxonomy, and persistent data silos. Collectively, these challenges negatively affect the overall quality of organisational data. ***P4:*** *Data discoverability is one of the challenges, you don't know where data is. Then once you identify data, you don't know what is in there, so it's almost inexistent documentation. Then very slow on prem source systems."*

P8: *Extracting data, taking data from one system and then putting it into a big data platform, there's no challenge there. That's the easiest thing that can be done. The challenge that we are encountering is once the data lands on the data lake, some of those fields are difficult to read. Especially when you've got a record from these off the shelf systems. They are written in a form that makes sense in that particular system and if you don't understand what those data elements refers. Only few individuals will understand what that data means, and it ends up not being usable."*

P2: *Standardization of data and a unified data taxonomy remains a challenge. I think other issues are things like making sense of unstructured data. Getting a handle on things like verbatim commentary, for customer complaints. The other problem we also face is accurately transcribing voice to text, especially in South Africa, where there's different native languages that don't have as much data for natural language processing. So, getting those transcribed is a bit of a challenge, however we are working on improving this in collaboration with local universities language departments. Also on the converse side, if you think about using things like large language models, English is very well defined, using something like a Chat GPT becomes easy because of it because it's had so much data to train on."*

P9: *Data silos are a challenge. Each department does its own thing making it difficult to integrate and analyze comprehensively. I mean, you might find that I've got data about you because you've got a car with this bank. But the same information can't be used if you still have a credit card in this bank because we are two different departments. We are actively working on sorting out the silos."*

b. Maturity Levels

The majority of participants agreed that technologies within the organisation are assessed for their adequacy in servicing each use case to process Big Data effectively. However, one respondent raised concerns about the lack of training and the challenge of committing to a strategy long enough to realise value, particularly given the ever-evolving technology landscape.

Respondent 9 highlighted two major concerns: the constantly changing environment, which may lead to a lack of sustained commitment to achieving outcomes, and disparities in skill levels across the organisation. These challenges are being addressed through various skills development programmes.

Overall, the findings suggest that a balance must be struck between rolling out new technologies and ensuring that sufficient training is provided upfront to enable their effective use.

P8: *Assessments are done because if you look at the Redshift for an example, it's got large-scale data analytics. So, in terms of capabilities, the tools do support different use cases that needs to drive business requirements."*

P9:" *I do think we can use better technologies, but, for now, it's adequate. Just that we, we jump into the hype of using new technologies everywhere. I've worked for three different big financial institutions. I haven't seen any company which experiments with new technology like Bank X. However, due to the ever-evolving landscape we sometimes do not commit long enough to see the results. We also have the challenge of varying skills levels, which are being addressed through various skills development programmes. I think that's where we need to find a balance of making sure that when we roll out technology, we get people trained properly beforehand to use that technology effectively."*

Most of the participants advised that enterprise wide a single version of the truth does not exist at Bank X, however the organisation has been actively working on and it and will continue to as business changes and new products are on boarded. The participants are suggesting that the organisation's current operating model and the

issue of legacy systems is one of the contributing factors in the delay. However, at the business unit and function level, there are single source of the truth.

P4: *"Yes, to some extent, a single version of the truth exists."*

P10: *The single version of the truth component has been in discussing dating back to mid-90s or more so late 90s even today, it's still a discussion point. In the context of Bank X, there are so many products that are available in the bank. There is no single place where you can go and say this is my single authoritative source that houses all the products. It's something that we're still building, and the challenge is we build, when we are about to conclude, either the technology change or the business structures change and you have to start all over again. However, for divisions, each and every division will build their own authoritative source to address their needs. Like I said, financial crime will have their own authoritative source, and another risk area will have their own authoritative source."*

P12: *Yes, there is, but single source of truth it's still a mystery for a lot of. In the corporate space there is a single source of truth for the whole entire corporate division. In the retail space, there's a single source in that whole division. In the regional, same applies. But group wide, no. because of our current operating model."*

4.5. Findings Pertaining to Proposition 3

This sub-section presents the findings on the influence Big Data management has on business outcomes. The categories and themes are outlined below, along with participants' views on the following proposition:

Proposition 3: *Effective Big Data management has a positive influence on business outcomes.*

4.5.1. Customer

The majority of participants highlighted that Big Data has improved the customer experience by enabling the bank to be more responsive to customer needs. Big Data

has allowed the bank to create personalised offerings and ensure that customers receive what they need, when they need it.

***P*1:**" *Big data has enabled the banks to produce personalized services to customers. It has also improved the risk assessment, in terms of a credit risk where historical financial data and transaction patterns are analyzed."*

P11: *In the context of the bank, big data has made it possible to fit the specific needs of a customer, which increases the product penetration. For an example, we're not allowed to automatically increase your credit card limit as a customer because of the regulation. But, through the digital channels, the data analytics can inform customers to say, hey, you qualify for a 5,000 increase on your credit card, which in the past took longer to process."*

4.5.2. Revenue and Sales

According to participants, Big Data has contributed to increased revenue. The bank has experienced double-digit product uptake from recommendations generated through its "next best product" solution, compared with the traditional approach. Participants also indicated that Big Data enables the identification of new opportunities for sales, customer retention, upgrades, and cross-selling.

***P*2:** *The next best product I mentioned has also been very successful in product take up. We are seeing 13 to 14% product take up on the recommendations for next best product versus traditional spray and pray approach, which generally translates to about between 1 and 2% depending on the product line, 3% conversion."*

P4: "*We have been able to operationalise our data through dashboards, that generate leads, we can also track what happened to that lead. For example, so the lead would be contact this customer and offer her a product x, and then you track over time if there is a take up. So right now, if I want to know how many of our customers, how many of our products a given customer has, I can find that quite easily."*

P5: *In our area we have a commercialization of data mindset. We are able to identify opportunities for sales, retention, upgrades, cross sell. Our inhouse build Next Best*

Product model is live in all of our branches. Every time a customer walks into a branch into a service queue, after the service interaction is done, the system automatically comes to my analytics world, picks up the three next best products, the service agent has a conversation with the customer, suggests this product, customer then says whether they're interested or not, generates a lead, we win."

4.5.3. Operational Excellence / Cost savings

a. Fraud Detection

Most of the participants expressed how Big Data models have minimised losses due to operational inefficiencies such as real time fraud detection and monitoring.

P9: *In our world, we're using Big Data for fraud detection, risk management and providing personalized financial services to our customers. Even though we're not yet mature enough, but these are the main three things which I see the benefits being shown thus far."*

P7: *In the bank the benefits we have seen with Big Data include operational efficiency, conduct anomaly detections, fraud detections, and mostly customer insights."*

b. Improved Decision Making

P6: *So with that pattern you can recognize how your operational efficiency is going, how your risk management is performing, how your comprehensive or like how your analyst analysis of the data is going and how you are managing your threshold of the targets and limitations of the data.*

We have operationalized the output from data through automated triggers, for example how we get data from the incident management system into our data lake, this is done via an API or a replicator that just checks if someone changes something or there's a new incident, we receive that trigger via an API into our data source."

4.6. Summary of the Findings

This chapter began by providing the organisational context and describing participant demographics. The findings on the essential elements of Big Data management, as identified by the participants, and the influence of Big Data management on business outcomes were then presented. These findings were categorised and discussed in relation to the propositions outlined below.

Proposition 1: *Data tools and data privacy are essential elements for Big Data Management.*

Organisations take different approaches to Big Data management. The majority of participants indicated that the tools used at Bank X were chosen to accommodate the bank's specific use cases. The enterprise-wide tool of choice is Hadoop, supported by multiple collection tools. This combination enables the organisation to efficiently extract, store, and process large, complex volumes of data to meet business and customer needs more quickly.

Proportion 2: *Addressing challenges that hinder adoption is essential for successful Big Data Management.*

Participants unanimously agreed that robust precautions are in place to prevent unauthorised data access during collection and curation. These include logical access management, encryption, periodic reviews, and segregation between production and UAT environments. However, some participants noted that while the security strategy is well-implemented and effective at protecting data during collection and processing, the centralised security approach may cause delays that impact delivery.

Proposition 3: *Effective Big Data management has a positive influence on business outcomes.*

Participants affirmed that effective Big Data management has a positive impact on business outcomes. For example, revenue and sales have increased through the "next best product" offering, and Big Data has also enhanced the customer experience.

5. DISCUSSION OF THE FINDINGS

This chapter presents the findings from the study outlined in Chapter 4 and compares them with the literature reviewed in Chapter 2 to identify similarities and discrepancies. The discussion focuses on addressing the study's objectives:

- To evaluate Big Data tools in the banking sector,
- To evaluate Big Data management challenges in the banking sector, and
- To evaluate the benefits of adopting Big Data in the banking sector.

5.1. Discussions Pertaining to Demographics

The participants represented different management levels and had varied years of service at Bank X. The researcher believed, and this was supported by the literature, that such diversity would provide depth in understanding organisational behaviour and culture from multiple vantage points.

It is standard in recent qualitative research, especially when involving leadership and management, to include role titles and corporate levels to highlight hierarchy and variation in perspectives (Easton & Steyn, 2025). Mamburu et al. (2024), in their study *Exploring Middle Managers' Identity as Strategists within a Public Sector Organisation: A Multilevel Perspective*, support this approach, noting that interviewing across management levels adds depth and triangulation, as senior and middle managers often offer different perspectives on how strategy is enacted and experienced. True to this, participants' perspectives in the current study varied according to their business areas, and the researcher was able to deduce varying levels of organisational maturity based on their feedback.

5.2. Discussion pertaining to Proposition 1

Proposition 1: *Data tools and Data privacy are essential elements for Big Data Management.*

5.2.1. Data Tools

Nisar et al. (2021) state that the quality of decision-making is influenced by the technology, tools, and managerial experience employed in an organisation. This was confirmed in this study, as most participants indicated that the Big Data tools used at Bank X were selected to accommodate the bank's specific use cases. Given the organisation's size and legacy systems, Bank X has employed a range of collection and processing methods to achieve business goals. Venkatesh et al. (2019) conducted a theoretical study on the challenges, disputes, and tools in Big Data Analytics, finding Apache Hadoop to be the most influential and widely established tool. Bank X has adopted Hadoop as its enterprise-wide tool of choice. Supported by multiple collection tools, Hadoop enables the organisation to efficiently extract, store, and process large, complex volumes of data, meeting business and customer needs more effectively. Venkatesh et al. (2019) also note that combining multiple tools offers the advantage of specialised processing for different data types, leading to better data management and more accurate insights, a finding that aligns with Bank X's approach.

5.2.2. Data Privacy

a. Security and Access Management

Participants at Bank X unanimously agreed that the bank employs robust precautions against unauthorised access during data collection and curation, including logical access management, encryption, periodic reviews, and segregation between production and UAT environments. This finding supports Ladley (2019), who argues that implementing governance during the data collection stage ensures responsible practices while maintaining security, quality, and regulatory compliance.

Most participants noted that tools and architectures supporting data security and access management are determined by the bank's Security Office. In addition, governance committees oversee the approval of new technologies. This approach aligns with Da Silva and Jensen (2022), who highlight that a security office plays a specific role in responding to and mitigating threats by prescribing appropriate security solutions. However, some participants cautioned that, while the centralised security

strategy is well-implemented and effective, it can also create delays, thereby impacting delivery timelines. Jablonowski (2019) supports this concern, noting that bottlenecks are common in centrally managed security processes, particularly in large organisations handling high data volumes and frequent queries. Such procedures can slow decision-making and solution delivery. In the case of Bank X, a hybrid model may be a more balanced approach.

b. Data Governance and Quality Assurance

In this study, data governance refers to the processes related to access management during collection, the tools used for data collection and processing, and the challenges the organisation faces. Effective data governance is crucial for managing access, ensuring data quality, maintaining security, and complying with regulations. Alhassan et al. (2019) emphasise that access management is a critical component of governance frameworks.

Most participants underscored the importance of data governance and quality assurance, particularly at the collection stage. The application of data management principles, such as standardisation, accuracy, and validation, is crucial for maintaining data quality (Nisar et al., 2021). Quality assurance tests are implemented across the data value chain, from the point of capture to ingestion, to identify and correct inconsistencies at the earliest stage possible.

Participants also stressed that ensuring consistency and standardisation requires integrating metadata into the governance framework. Metadata enhances transparency by providing detailed information about data context and lineage, including its source, transformations, and quality checks. Riley (2019) supports this view, noting that metadata strengthens governance by improving data transparency and reducing errors.

Bank X's approach also aligns with Ladley (2019), who argues that governance must be embedded at the collection stage to ensure data is managed responsibly while preserving quality, security, and compliance. Rossi and Hirama (2022) further emphasise that high-quality datasets, characterised by accuracy, completeness,

timeliness, and relevance, are achieved through mechanisms such as validation, cleansing, and enrichment procedures.

Similarly, Côté et al. (2024) highlight that data cleaning is essential for producing accurate, complete, and consistent datasets for analysis and machine learning. Implementing best practices in data cleaning elevates data quality, resulting in more reliable and impactful analytical findings.

5.3. Discussion Pertaining to Proposition 2

Proposition 2: *Addressing Challenges that hinder Adoption is Essential for Successful Big Data Management.*

5.3.1. Addressing challenges associated with the adoption of big data.

a. Quality Assurance

Participants identified several challenges during the data collection and processing stages, including data discoverability, understanding data elements, accurately transcribing voice to text due to insufficient natural language processing (NLP), establishing a unified data taxonomy, and managing data silos. Collectively, these issues impact overall organisational data quality.

The *Comprehensive Guide to Data Management* (International, 2017) underscores the importance of data discoverability for effective governance and metadata handling. It argues that without proper cataloguing and metadata practices, data assets lose much of their value. Interestingly, participants in this study contradicted this finding by indicating that Bank X actively uses data catalogues and metadata in data interpretation and processing. The challenge of transcribing voice to text raised by participants reflects well-documented limitations in current NLP technology (Alawida et al., 2023).

b. Maturity Levels

Participants noted that Bank X currently lacks a single, organisation-wide “version of the truth” for its data, a challenge also discussed in the literature. Chaoui et al. (2022) highlight that effective lifecycle management requires clear processes to govern data from its creation to eventual archiving or removal. At Bank X, however, the federated operating model and reliance on legacy systems impede the establishment of a unified enterprise-wide data view.

Although “single sources of truth” exist at the business unit level, the lack of consistency across the enterprise reflects fragmented data management. This fragmentation likely stems from inconsistently applied lifecycle processes and siloed operations. Literature also stresses that robust data retention policies, driven by legal and commercial requirements, are crucial. Extending this, the participants’ feedback suggests that processes related not only to retention, but also to creation, storage, and access, are not uniformly managed across Bank X.

5.4. Discussion Pertaining to Proposition 3

Proposition 3: *Effective Big Data management has a Positive Influence on Business Outcomes.*

5.4.1. Customer

Most participants highlighted that Big Data has significantly improved customer experience by enabling the bank to respond more effectively to customer needs. Personalised offerings ensure that customers receive timely, relevant products and services.

This aligns with Indriasari et al. (2019), who examined the application of Artificial Intelligence (AI) and Big Data Analytics in banking. Their study confirmed that these technologies can enhance customer experience but stressed the importance of overcoming legacy systems and fully embracing digital transformation. They identified several factors that condition success. These include: (a) the bank’s current state, (b)

strategy, (c) customer focus, (d) organisational capabilities, (e) brand promise, (f) regulatory requirement and (g) capital constraints.

5.4.2. Revenue and Sales

Participants reported that Big Data has directly increased revenue. Specifically, they cited double-digit product uptake resulting from the “next best product” recommendation engine, compared to traditional approaches. They also emphasised improved identification of sales opportunities, customer retention strategies, upgrades, and cross-selling initiatives.

These findings are consistent with Hitachi (2020), who argue that leveraging Big Data enables banks to develop customer profiles that anticipate behaviours, identify cross-selling opportunities, and deliver proactive, empathetic service, thereby enhancing both customer experience and revenue growth.

5.4.3. Operational Excellence

Participants noted that Big Data models have reduced losses linked to operational inefficiencies, particularly through real-time fraud detection and monitoring.

This is supported by Vivek et al. (2023), who demonstrate how machine learning models can promptly detect fraudulent activity, improving both operational efficiency and customer trust.

5.5. Conclusion

Proposition 1: *Data tools, data privacy as well as addressing challenges that hinder adoption are essential elements for Big Data Management.*

There is clear convergence between the tools reported by participants and those identified in the literature. Hadoop, in particular, stands out as both a widely acknowledged industry standard and Bank X’s enterprise-wide tool of choice. While Bank X’s centralised security strategy is effective at collection and processing,

participants raised concerns about delays, reflecting Jablonowski's (2019) observation that centrally managed models often create bottlenecks. A hybrid security model may therefore strike a better balance between security and efficiency. ***Proposition 2:*** *Effective Big Data management has a positive influence on business outcomes.*

Addressing challenges is critical to successful Big Data management. Participants reported difficulties with data discoverability, silos, and quality assurance. These challenges appear to contradict International (2017), which links poor discoverability to insufficient metadata. However, since participants noted that Bank X uses catalogues and metadata, the issue may lie not in the absence of metadata but in the fragmented and inconsistent application of governance practices across the organisation.

Proposition 3: *Effective Big Management positively influences business outcomes.*

Participants confirmed that Big Data has improved customer experience through personalised offerings and responsiveness, findings that align with Indriasari et al. (2019). Revenue has grown through stronger product uptake and cross-selling opportunities, echoing Hitachi (2020). Furthermore, participants cited reductions in fraud-related losses and improved operational efficiency, a result supported by Vivek et al. (2023).

Taken together, these findings confirm that while significant governance and adoption challenges remain, Bank X is realising measurable benefits from Big Data management, particularly in customer satisfaction, revenue generation, and operational resilience.

6. CONCLUSIONS & RECOMMENDATIONS

6.1. Introduction

This chapter integrates the findings related to the research propositions with the original research questions set out in Chapter 1. It addresses each research question and objective in turn. The objectives were to:

- Evaluate Big Data tools in the banking sector.
- Assess the challenges of Big Data management in the banking sector.
- Evaluate the benefits of adopting Big Data in the banking sector.

6.2. Conclusions regarding which tools are used for Big Data in the banking industry

In response to the first research question, Venkatesh et al. (2019) identified Apache Hadoop as the most influential and established tool for analysing Big Data, particularly effective in uncovering complex patterns and insights beyond the reach of traditional methods.

This study confirmed that the right data tools, together with robust data privacy practices and the ability to address adoption challenges, are essential for effective Big Data management. At Bank X, Hadoop emerged as the primary enterprise-wide tool, supported by a range of in-house solutions and off-the-shelf tools such as SAS, SQL, Databricks, AWS Lambda, and MongoDB.

The findings also highlighted that tool selection depends on several factors, including legacy systems, specific business unit requirements, platform capabilities, and cost considerations. These observations align with Venkatesh et al. (2019), who emphasise that combining multiple tools provides specialised processing for diverse data types, resulting in better data management and more accurate insights.

6.3. Conclusions Regarding the Challenges Faced in the Adoption of Big Data in the Banking Industry

The adoption of Big Data in the banking sector presents several challenges. Participants identified issues such as data discoverability, understanding data elements, limitations in natural language processing for voice-to-text transcription, the absence of a unified data taxonomy, and persistent data silos, all of which undermine overall data quality.

There were differing opinions about data cleaning. Some participants stressed that cleaning data at the source is costly, which is why data is often cleaned in the data lake instead. Others highlighted the importance of ensuring alignment between the Big Data system and source systems, emphasising consistency, data quality tracking, and issue resolution at the source.

Most participants acknowledged that technologies are evaluated for their ability to service Big Data use cases effectively. However, concerns were raised about the constantly evolving technology landscape and the need for adequate training. One participant noted that rapid change can result in a lack of sustained commitment, compounded by varying skill levels among employees. These challenges are being addressed through skills development programmes.

Behl et al. (2019) argue that access to relevant data and employees' technical skillsets are critical enablers for Big Data Analytics adoption. Cabrera-Sánchez and Villarejo-Ramos (2019) further note that robust infrastructure can mitigate implementation challenges, though they caution that organisations already using Big Data Analytics remain sceptical about its overall performance.

While a single version of the truth has not yet been achieved enterprise-wide at Bank X, participants indicated that progress is being made. Legacy systems and the current operating model continue to delay this outcome, though at the business unit and functional level, single sources of truth do exist.

6.4. Conclusions Regarding the Realised Benefits of Big Data in the Banking Industry

Participants confirmed that Big Data improves customer experience by enabling greater personalisation and responsiveness. This finding aligns with Indriasari et al. (2019), who highlight AI and Big Data's potential for enhancing customer experience, while emphasising that success requires digital transformation and strategic alignment.

Big Data was also found to have boosted revenue, with participants citing double-digit product uptake driven by the "next best product" recommendation solution, alongside increased sales opportunities, retention, and cross-selling. These findings correspond with Hitachi (2020), who argue that Big Data strengthens customer understanding, facilitates proactive engagement, and drives revenue growth. In addition, participants highlighted that Big Data models have minimised losses by reducing operational inefficiencies, particularly through real-time fraud detection and monitoring. This is supported by Vivek et al. (2023), who demonstrate the effectiveness of machine learning in promptly detecting fraudulent activities and enhancing both efficiency and customer confidence.

6.5. Recommendations

Based on the findings, several recommendations are proposed for Bank X. First, there is a need to strengthen best practice sharing across the organisation. This can be achieved by reinforcing the remit of the central data office and formalising cross-business unit exchange. Establishing structured platforms for knowledge sharing, together with initiatives such as job shadowing programmes, would provide opportunities for practical, hands-on learning and ensure effective transfer of knowledge throughout the organisation.

Secondly, the Bank should adopt a more measured approach to technology rollout. Instead of immediately pursuing the latest innovations, priority should be given to fully realising the value of existing systems. This would promote sustainable

implementation, allow for lessons learned to be integrated into future rollouts, and minimise the disruption often caused by the rapid pace of technological change.

Finally, the study recommends the development and implementation of a comprehensive training and upskilling strategy prior to the rollout of new technologies. Equipping employees with the necessary skills in advance will ensure that they are well prepared, promote more effective adoption of new systems, and maximise the return on investment in technology.

6.6. Suggestions for Further Research

While this study offers valuable insights, further qualitative research is recommended to explore specific challenges such as data governance and technology adoption in more detail. Comparative case studies across banks of different sizes, structures, and geographic locations would provide richer contextual understanding and more nuanced perspectives.

Uncategorized References

Abdullahi, S., Yusuf, N., Zayyad, M. A., Bagiwa, L. I., Zakari, A., Adamu, A., Nura, A., & Shehu, S. (2020). The Big Data Technology: Assessing the Impact in the Banking Industry. *Open Journal of Science and Technology, 3*(2), 71-77.

Adewusi, A. O., Okoli, U. I., Adaga, E., Olorunsogo, T., Asuzu, O. F., & Daraojimba, D. O. (2024). Business intelligence in the era of big data: a review of analytical tools and competitive advantage. *Computer Science & IT Research Journal, 5*(2), 415-431.

AFRICA, B. A. S. (2017). *TRANSFOMATION IN THE FINANCIAL SECTOR*. https://static.pmg.org.za/170314BASA.pdf

Alawida, M., Mejri, S., Mehmood, A., Chikhaoui, B., & Isaac Abiodun, O. (2023). A comprehensive study of ChatGPT: advancements, limitations, and ethical considerations in natural language processing and cybersecurity. *Information, 14*(8), 462.

Almeida, F. L. (2017). Benefits, challenges and tools of big data management. *Journal of Systems Integration, 8*(4), 12-20.

Amakobe, M. (2015). The Impact of Big Data Analytics on the Banking Industry. *Colorado Technical University, 4*.

Armstrong, P. B. (2019). *Business Strategy for a Digital World*

Badshah, A., Daud, A., Alharbey, R., Banjar, A., Bukhari, A., & Alshemaimri, B. (2024). Big data applications: overview, challenges and future. *Artificial Intelligence Review, 57*(11), 290.

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. *Information Processing & Management, 56*(6), 102095.

Barakat, K. A., & Sayegh, M. M. (2021, February 2021). Information Asymmetry in the Age of Big Data Analytics.

BASA. (2023, 4 September 2024). Banking Association of South Africa Integrated Report. https://www.banking.org.za/wp-content/uploads/2024/09/BASA_IAR-2023.pdf

Behl, A., Dutta, P., Lessmann, S., Dwivedi, Y. K., & Kar, S. (2019). A conceptual framework for the adoption of big data analytics by e-commerce startups: a case-based approach. *Information Systems and e-Business Management, 17*(2), 285-318.

Bertino, E., & Ferrari, E. (2017). Big data security and privacy. In *A comprehensive guide through the Italian database research over the last 25 years* (pp. 425-439). Springer.

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: an empirical investigation. *MIS quarterly*, 169-196.

Bitsch, V. (2005). Qualitative research: A grounded theory example and evaluation criteria. *Journal of agribusiness, 23*(345-2016-15096), 75-91.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Burgess, A. (2017). *The Executive Guide to Artificial Intelligence: How to identify and implement applications for AI in your organization*. Springer.

Burke, M. (2020). *Governance in a Digital Age: Frontier Technologies*.

Cabrera-Sánchez, J.-P., & Villarejo-Ramos, A. F. (2019). Factors affecting the adoption of big data analytics in companies. *Revista de Administração de Empresas, 59*(6), 415-429.

Cele, S. K., & Mlitwa, N. W. (2024). Fintechs in South Africa: Impact on regulation, incumbents and consumers. *South African Journal of Information Management, 26*(1), 1766.

Chaoui, K., Kabachi, N., Harbi, N., & Badir, H. (2022). Data labeling for data security in data lifecycle: A state of the art and issues. *Business Intelligence & Big Data*.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology, 30*(1), 44-57.

Côté, P.-O., Nikanjam, A., Ahmed, N., Humeniuk, D., & Khomh, F. (2024). Data cleaning and machine learning: a systematic literature review. *Automated Software Engineering, 31*(2), 54.

Creswell, J. W. C. J. D. (2018). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches* (H. Salmon, Ed. fifth ed.). SAGE Publications, Inc.

Da Silva, J., & Jensen, R. B. (2022). " Cyber security is a dark art": The CISO as Soothsayer. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), 1-31.

Daly, O. S.-J. J. F. R. (2020, 7 February 2020). The growth of data in banking is unstoppable. https://bankingblog.accenture.com/growth-of-data-in-banking-is-unstoppable

Deighton, K. (2024, 3 April 2024). Chase Bank to Let Advertisers Target Customers Based on Spending Data. *The Wall Street Journal*. https://www.wsj.com/articles/chase-bank-to-let-advertisers-target-customers-based-on-spending-data-275f8226?utm_source=chatgpt.com

Dlamini, Z. (2024). Influence of big data analytics on decision-making processes in financial firms in South Africa. *American Journal of Data, Information and Knowledge Management, 5*(1), 14-25.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63-71.

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

Easton, C., & Steyn, R. (2025). Millennial Leaders 'Preferences for Leadership Development: A Qualitative Analysis. *Administrative Sciences, 15*(4), 135.

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication monographs, 85*(1), 57-80.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of business research, 69*(2), 897-904.

Ertel, W. (2018). *Introduction to artificial intelligence*. Springer.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144.

Gbahabo, L. E. N. a. P. T. (2025). *How can agency banking deepen financial inclusion in South Africa?* (South African Reserve Bank Working Paper Series WP/25/02, Issue. S. A. R. Bank. https://www.resbank.co.za/content/dam/sarb/publications/working-papers/2025/how-can-agency-banking-deepen-financial-inclusion-in-south-africa.pdf

Géczy, P. (2014). Big data characteristics. *The Macrotheme Review, 3*(6), 94-104.

Gray, D. E. (2019). *Doing research in the business world*. Sage.

Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management, 42*, 78-89.

Hall, M. (2021). How the Banking Sector Impacts Our Economy. Retrieved 24 June 2021, from https://www.investopedia.com/ask/answers/032315/what-banking-sector.asp

Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in banking. *Big Data and Cognitive Computing, 2*(3), 18.

Hitachi. (2020). Modern Data Analytics in Banking: Benefits, Outlook & More. https://global.hitachi-solutions.com/blog/big-data-banking/?utm_source=chatgpt.com

IBM. (2020, 28 May 2020). How to manage complexity and realize the value of big data. https://www.ibm.com/blogs/services/2020/05/28/how-to-manage-complexity-and-realize-the-value-of-big-data/

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 sector. 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI),

International, D. (2017). *DAMA-DMBOK: Data management body of knowledge*. Technics Publications, LLC.

Jablonowski, A., Seitz, R., & Schneider, M. . (2019). Security Models in Big Data: Centralized vs. Decentralized. *Journal of Information Security*, 14-30.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government information quarterly, 37*(3), 101493.

Khajouei, H., Kazemi, M., & Moosavirad, S. H. (2017). Ranking information security controls by using fuzzy analytic hierarchy process. *Information Systems and e-Business Management, 15*(1), 1-19.

Krogstie, J., & Gao, S. (2015). A semiotic approach to investigate quality issues of open big data ecosystems. International Conference on Informatics and Semiotics in Organisations,

Kshetri, N. (2014). Big data ′s impact on privacy, security and consumer welfare. *Telecommunications Policy, 38*(11), 1134-1145.

Ku, P. (2020). Big Data in Banking: Use Cases in 2020 and Beyond. https://blogs.informatica.com/2020/04/13/big-data-banking-use-cases/

Kumar, Y., Marchena, J., Awlla, A. H., Li, J. J., & Abdalla, H. B. (2024). The AI-powered evolution of big data. *Applied Sciences, 14*(22), 10176.

Ladley, J. (2019). *Data governance: How to design, deploy, and sustain an effective data governance program*. Academic Press.

Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New directions for program evaluation, 1986*(30), 73-84.

Lochy, J. (2019). Big Data in the Financial Services Industry - From data to insights. https://www.finextra.com/blogposting/17847/big-data-in-the-financial-services-industry---from-data-to-insights

Mamburu, M. E., de Metz, N., & Davis, A. (2024). Exploring middle manager's identity as strategists within a public sector organisation: a multi-level perspective. *Journal of Strategy and Management, 17*(4), 744-762.

Mandlik, M. A., & Kadirov, D. (2018). Big data approaches and outcome of information asymmetry: opportunities for future research. *International Journal of Business Continuity and Risk Management, 8*(4), 303-318.

Maritz, J. C. C. (2018). Digital disruption in South African banking sector. https://www.pwc.co.za/en/publications/digital-disruption-in-sa-banking-sector.html

Maryville. (2021). Big Data and Artificial Intelligence: How They Work Together. https://online.maryville.edu/blog/big-data-is-too-big-without-ai/

MasterClass. (2020). Information Asymmetry Explained. Retrieved 11 August 2021, from https://www.masterclass.com/articles/information-asymmetry-explained#what-is-information-asymmetry

Maunder, S. (2021). Guide to banks in South Africa. https://www.expatica.com/za/finance/banking/banking-in-south-africa-105795/

Mazumder, S. (2016). Big data tools and platforms. In *Big data concepts, theories, and applications* (pp. 29-128). Springer.

Mazzei, M. J., & Noble, D. (2019). Big Data and Strategy: Theoretical Foundations and New Opportunities. In *Strategy and Behaviors in the Digital Economy*. IntechOpen.

Mikalef, P., Ilias, P. O., Giannakos, M., Krogstie, J., & Lekakos, G. (2016). Big data and strategy: a research framework. MEDITERRANEAN CONFERENCE ON INFORMATION SYSTEMS (MCIS),

Mncube, K., & Viegi, N. (2024). Measuring the unit cost of financial intermediation in South Africa: A measure of bank productivity. *South African Journal of Economics, 92*(1), 47-56.

More, R., & Moily, Y. (2021). Big Data Analysis in Banking Sector.

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative health research, 25*(9), 1212-1222.

MuleSoft. (2021). Financial Services and Legacy Systems. https://www.mulesoft.com/resources/esb/financial-services-legacy-systems-modernization

Mungai, K., & Bayat, A. (2018). The impact of big data on the South African banking industry. 15th International Conference on Intellectual Capital, Knowledge Management and Organisational Learning, ICICKM 2018,

Nisar, Q. A., Nasir, N., Jamshed, S., Naz, S., Ali, M., & Ali, S. (2021). Big data management and environmental performance: role of big data decision-making capabilities and decision-making quality. *Journal of Enterprise Information Management, 34*(4), 1061-1096.

Ochuba, N. A., Amoo, O. O., Okafor, E. S., Akinrinola, O., & Usman, F. O. (2024). Strategies for leveraging big data and analytics for business development: a comprehensive review across sectors. *Computer Science & IT Research Journal, 5*(3), 562-575.

Parashar, S. (2020). Big Data Analytics: An Approach for Banking Industry. *EmErging TrEnds in Big daTa, ioT and CyBEr sECuriTy*, 42.

Parson, G. K. (2021). *Factors Affecting Information Technology Professionals ' Decisions to Adopt Big Data Analytics Among Small-and Medium-Sized Enterprises: A Quantitative Study* Capella University].

Patil, P. S., & Dharwadkar, N. V. (2017). Analysis of banking data using machine learning. 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC),

PWC. (2025). *Adapting with agility South Africa Major Banks Analysis*.

Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management, 38*(1), 187-195.

Riley, J. (2019). *Understanding Metadata: What is Metadata, and What is it For?*

Rossi, R., & Hirama, K. (2022). Characterizing big data management. *arXiv preprint arXiv:2201.05929*.

Salleh, K. A., & Janczewski, L. (2019). Security considerations in big data solutions adoption: Lessons from a case study on a banking institution. *Procedia Computer Science, 164*, 168-176.

SARB. (2025). *SARB ABOUT US*. South African Reserve Bank. https://www.resbank.co.za/en/home/about-us

SAS. (2021). Big Data Analytics: What it is and why it matters. https://www.sas.com/en_us/insights/analytics/big-data-analytics.html

Schindler, P. S. (2019). *Business Research Methods* (thirteenth ed.). McGraw-Hill/Irwin.

Shan, S., Luo, Y., Zhou, Y., & Wei, Y. (2019). Big data analysis adaptation and enterprises 'competitive advantages: the perspective of dynamic capability and resource-based theories. *Technology Analysis & Strategic Management, 31*(4), 406-420.

Srivastav, D. P. V. M. M. R. K. (2021, March 2021). BIG DATA ANALYTICS IN BANKING AND FINANCIAL SERVICES SECTOR [Peer Reviewed Journal]. *DOI: 10.36713/epra2016, 6*(3), 5. https://doi.org/https://doi.org/10.36713/epra6558

Sternkopf, C. (2019, 11 November 2019). LEGACY MODERNIZATION: THIS IS HOW BANKS PUT THEIR IT IN THE FAST LANE. https://www.knowis.com/blog/legacy-modernization-this-is-how-banks-put-their-it-in-the-fast-lane

Surya, L. (2015). An exploratory study of AI and Big Data, and it's future in the United States. *International Journal of Creative Research Thoughts (IJCRT), ISSN*, 2320-2882.

Tole, A. A. (2013). Big data challenges. *Database systems journal, 4*(3), 31-40.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative inquiry, 16*(10), 837-851.

tutorialspoint. (2021). *principles_of_communication_information_theory* https://www.tutorialspoint.com/principles_of_communication/principles_of_communication_information_theory.htm

Venkatesh, K., Ali, M. J. S., Nithiyanandam, N., & Rajesh, M. (2019). Challenges and Research Disputes and Tools in Big Data Analytics. *International Journal of Engineering and Advanced Technology, 6*, 1949-1952.

Vivek, Y., Ravi, V., Mane, A. A., & Naidu, L. R. (2023). ATM fraud detection using streaming data analytics. *arXiv preprint arXiv:2303.04946*.

Wahyuni, D. (2012). The research design maze: Understanding paradigms, cases, methods and methodologies. *Journal of applied management accounting research, 10*(1), 69-80.

Wu, W., & Wang, X. (2024). Navigating strategic balance: CEO big data orientation, environmental investment, and technological innovation in Chinese manufacturing. *Systems, 12*(7), 255.

APPENDIX A - Participant information sheet

INFORMATION SHEET FOR PARTICIPANTS

01 February 2024

Dear Sir/Madam

My name is Phumelele Kubheka, a Master on Management in the field of Digital Business student at the University of Witwatersrand, Faculty of Commerce, Law and Management, in the Business School. I kindly request you to participate in my study titled: ***The case of a South African Bank's management of Big Data*** under the supervision of Dr Tebogo Sethibe.

The purpose of this qualitative study is to investigate a South African Bank's management of Big Data. With a focus on Big Data Tools used, the impact of Big Data and the realised benefits or challenges of Big Data.

I am inviting you to take part in a single interview. If you decide to take part, your participation in this research will involve answering questions and will take around 60 minutes. The interview activity will take place online or in person, depending on your availability.

With your permission, I would like to audio record the interview using a digital device. This data will be stored in a password protected computer and will be deleted after 1 year. Only my supervisor and I will have access to this recording. There will be no personal costs to you if you participate in this project. You will not receive any direct benefits from participation but there are no disadvantages or penalties if you do not choose to participate or if you withdraw from the study. You may withdraw at any time or not answer any question if you do not want to.

If you decide to take part in the research study, it should be because you want to volunteer. You do not have to take part. You do not have to answer any questions if you do not want to. Your involvement in this study is voluntary and you may withdraw at any point without any disadvantage to you if it becomes impossible for you to continue to participate. Your identity will remain confidential throughout the research project; no personal data is required. I will be using a pseudonym (false name) to represent your participation in my final research report.

If you have any questions during or afterwards about this research, feel free to contact me on the details listed below. This study will be written up as a research report which will be available online through the university library website. If you would like to receive a summary of this report, I will be happy to send it to you. With your permission the data collected from this research project may be used by other researchers in an anonymized format. If you have any concerns or complaints regarding the ethical procedures of this study, you are welcome to contact the University Human Research Ethics Committee (Non-Medical), telephone +27(0) 11 717 1408, email hrecnon-medical@wits.ac.za. Your participation will be appreciated. If you have further questions about the study, you can contact me or my supervisor.

Thanking you in anticipation

Researcher: Miss Phumelele Kubheka
Cell: 0833502865
Email: 2280561@students.wits.ac.za

Supervisor: Dr Tebogo Sethibe
Email: SethibeT@arc.agric.za

APPENDIX B - Participant agreement form

APPENDICES

Appendix A: Consent Form

Wits Business School
Sculpting global leaders

The case of a South African Bank's management of Big Data

By

Phumelele Kubheka

I…… agree to participate in this research project. The research has been explained to me and I understand what my participation will involve. I agree to the following:

(Please circle the relevant options below).

• I agree that my participation will remain anonymous	YES	NO
• I agree that the researcher may use anonymous quotes in his / her research report	YES	NO
• I agree that the interview may be audio recorded	YES	NO
• I am aware that I can skip any questions(s) I am uncomfortable with and that I have the right to stop the interview at any time.	YES	NO

…… (Signature)

…… (Initial of participant)

…… (Date)

Many thanks,

Phumelele Kubheka

APPENDIX C – Consistency Matrix

Table 6: Consistency table: Research Questions, Propositions, Data Collection and Data Analysis

Research Questions: What is the influence of Big Data management on a South African bank?						
RQ #	**Research Question**	**Research Objective**	**Prop #**	**State Proposition or Hypothesis**	**Data collection detail**	**Data analysis method**
1	Which tools are used for Big Data at Bank X?	Evaluate Big Data Tools at Bank X.	1	*Data tools and Data privacy are essential elements for Big Data Management.*	Interview guide questions a, b, c, d, e, f, g, h	Thematic analysis
2	What challenges are faced in the adoption of Big Data at Bank X?	Evaluate Big Data Management Challenges for Bank X.	2	*Addressing challenges that hinder adoption is essential for successful Big Data Management.*	Interview guide questions i, j, k, l, m, n, o	Thematic analysis
3	What are the realised benefits of Big Data at Bank X?	Evaluate the benefits of adopting Big Data at Bank X.	3	*Effective Big Data management has a positive influence on business outcomes.*	Interview guide questions p, q, r, s, t, u, v, w	Thematic analysis

APPENDIX D – Initial Codes and Themes

Initial code	Initial Theme	Final Theme
• SQL • SAS • APIs • Amazon Web Services (AWS) • Databricks • Hadoop • MongoDB • Apache Spark • Hive • AWS S3 • Dell ECS • Power BI • ClickView • Tableau • AWS Lambda • AWS Glue • Athena • EventBridge • CloudTrail • CloudWatch • VPC • Terraform • GuardDuty • Unity Catalog • Inhouse Access portal	• Big Data Collection Tools • Data Visualization Tools • Other Tools	• Data Tools

Initial code	Initial Theme	Final Theme

• Logical access management
• Periodic access reviews
• Segregation between environments
• Encryption at rest
• Encryption in transit
• Periodic key changes
• Key management
• Dedicated cloud forum
• Risk reviews
• Cybersecurity reviews
• Physical security reviews
• Compliance reviews
• Strict access governance
• Authorized personnel access
• Internal access tool
• Role-based access controls
• IAM roles and policies
	• Data encryption
	• Access controls
	• Role-based access
	• NDAs for authorized personnel
	• Data security measures
	• Data governance framework
	• Data privacy compliance
	• Secure data transmission
	• Data cleaning methods
	• Data quality assurance
		• Data Privacy
		  ○ Security and Access Management
		  ○ Data Governance

• Standardization of data
• Handling unstructured data
• Transcribing voice to text
• Developing governance system
• Performance issues
• Skill shortages
• Legacy challenges
• Data quality
• Discoverability
	• Data governance principles
	• Quality assurance checks
	• Consistent data format
	• Reduce data misinterpretation
	• Ensure data quality
	• Data accuracy improvement
	• Data cleaning methods
	• Data quality tracking
	• Data validation processes
		• Addressing challenges associated with the adoption of big data.
		  ○ Quality Assurance
		  ○ Maturity Levels

Initial code	Initial Theme	Final Theme
• Slow source systems • Lack of central data library • Data normalization and architecture • Responsiveness and data access • Data governance principles • Quality assurance checks • Consistent data format • Reduce data misinterpretation • Ensure data quality • Data accuracy improvement • Data cleaning methods • Data quality tracking • Data validation processes • Anomaly detection capabilities	• Anomaly detection capabilities • Advanced • No formalized process in place • Basic processes • Some level of discipline • Well-documented and standardized • Processes are measured and controlled • Process improvement • Quantitative feedback Piloting innovative ideas	

• Standardization of data
• Handling unstructured data
• Transcribing voice to text
• Developing governance system
• Performance issues
• Skill shortages
• Legacy challenges
• Data quality
• Discoverability
• Slow source systems
• Lack of central data library
• Data normalization and architecture
	• Understanding customer behavior
	• Improving decision-making
	• Increasing campaign profitability
	• Targeting right people
	• Predictive models
	• Prescriptive models
	• Efficient targeting
	• Higher take-up rates
	• Data-driven decisions
	• Real-time decisions
		• Customer
		• Revenue / Sales
		• Operational Excellence / Cost saving
		  ○ Fraud detection
		  ○ Improved decision making

Initial code	Initial Theme	Final Theme
• Responsiveness and data access • Fraud detection • Risk management • Anomaly detection • Data security • Access controls • Data encryption • Role-based access • Data governance • Compliance monitoring • Unauthorized access prevention		

APPENDIX E - Ethics approval notification

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/DB2280561/629
This certificate is only valid with a legitimate ethics protocol number and signed by the Researcher (below

Project title	The case of a South African bank's management of big data
Investigator / Researcher	Miss Phumelele Kubheka
Nature of Project	MM (Digital Business)
Decision of the Committee	Approved, provided stakeholders and participants are guaranteed confidentialty.
Issue Date of Certificate	2021-11-11
Expiry date	Date of submission of the project 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.

kubheka.
Signature

11 November 2021
Date:

APPENDIX F - Permission letter from organization

[A copy of the permission letter has been sent to the exam office]