Cape Peninsula
University of Technology

THE ARCHITECTURAL DESIGN OF BIG DATA FOR BUSINESS ENHANCEMENT IN ENTERPRISES

by

WANDISA NYIKANA

Dissertation submitted in partial fulfilment of the requirements for the degree

Master of information and Communication Technology

in the faculty of informatics and Design

at the Cape Peninsula University of Technology

Supervisor: Professor Tiko Iyamu

District Six
Submitted on (27 November 2023)

DECLARATION

I, Wandisa Nyikana, declare that the contents of this dissertation represent my own unaided work, and that the thesis/dissertation has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.

11/03/2024

Signed **Date**

ABSTRACT

Increasingly, organisations rely on big data for their business continuity, competitiveness, and sustainability. However, in many organisations, the rapidity at which the big data are generated, retrieved, and used is unprecedented. This contributes to complexity and challenges in many organisations. As a result, some organisations struggle to use big data, to improve business continuity and competitive advantage. Primarily, the complexity and challenges exist because there is no architecture, to govern and manage the big data. In an organisation where there is architecture, it is specifically designed for small data. Big data and small are not the same, hence, the same architecture cannot be used. Thus, organisations need to design an architecture that focuses on big data. This is the motivation for this study.

The design of architecture for big data was problematised, where the implications and consequences of lack of architecture are stated. For example, it states, that a lack of architectural design for big data compromises the quality including management and governance of big data in an organisation. Based on the problem, the study aimed to design big data architecture, which enterprises can employ, to enhance business continuity and advance competitiveness. This includes the objectives of the study. In achieving the aim and objectives, research questions were formulated, as presented in Chapter 1.

Qualitative methods were employed. Data were extracted from relevant literature, which were gathered from various sources, from both academic and business domains. A total of 201 papers were gathered. Activity theory (AT) was employed to guide the analysis of the data in which the hermeneutics technique was applied. From the analysis, the factors that influence the design of big data architecture were revealed. The factors were interpreted following the subjective reason approach. Based on the interpretation, big data architecture was developed, as presented in Chapter 5 (Figure 5.2.). The study was evaluated, to ensure that the study achieved its aim and objectives.

The study is significant and contributes to both business and academics, from technical and non-technical perspectives. This includes the engineering of big data, governance, and management standpoints. The significance and contributions of the study are discussed in Chapter 6.

ACKNOWLEDGEMENTS

I wish to thank:

- God - for the strength and wisdom He provided me through this journey.
- My husband, Stephen - for his unwavering support and encouragement. Your love and understanding have been my pillar of strength. This accomplishment would not have been possible without you.
- My daughter, Misokuhle, for constantly reassuring me that I am capable. Thank you for your confidence in me.
- My sister, Vuyolwethu - for the encouragement and support you provided me through this journey.
- My supervisor, Professor Tiko Iyamu - for his guidance, patience, and commitment. You have encouraged me to improve myself and become a better researcher. I will forever be grateful for the support you provided to me. May the Lord continue to bless you for the good work you are doing.
- My colleagues at the research forum - for their motivation and knowledge sharing.

DEDICATION

I dedicate this study to my late grandmother. Her belief in my potential motivated my academic journey. This work stands as a tribute to her unforgettable influence on my life. Ndiyabulela Ntombi yaseMacirheni.

TABLE OF CONTENTS

DECLARATION ii
ABSTRACT iii
ACKNOWLEDGEMENTS iv
TABLE OF CONTENTS vi
CHAPTER ONE 1
1.1 Introduction 1
1.2 Research problem 2
1.3 Problem statement 3
1.4 Aim of the study 3
1.4.1 Research objectives 3
1.5 Research question 3
1.5.1 Sub questions 4
1.6 Literature review 4
1.6.1 Enterprises 4
1.6.2 Big data in enterprises 5
1.6.3 Architecture of big data 6
1.6.4 Activity theory 6
1.7 Research methodology 8
1.7.1 Research methods 8
1.7.2 Research design 9
1.7.3 Data collection techniques 10
1.7.4 Data analysis 10
1.8 Significance of the study 11
1.9 Contribution of the study 11
1.10 Ethical considerations 12
1.11 Budget and timeline 12
1.12 Conclusion 12
CHAPTER TWO 13
2.1 Introduction 13
2.2 Enterprises 13
2.3 Big data in enterprises 15
2.4 Architecture of big data 18
2.5 Theoretical framework 20
2.5.1 Activity theory 20
2.5.2 Activity theory and Information Systems Research 22
2.6 Summary 23
CHAPTER THREE 24
3.1 Introduction 24
3.2 Research approach 24
3.2.1 Deductive Approach 24

3.2.2 Inductive Approach 25
3.3 Research methods 25
3.3.1 Quantitative method 26
3.3.2 Qualitative method 26
3.3.3 Mixed method 27
3.4 Research design 28
3.4.1 Ethnography 28
3.4.2 Grounded Theory 28
3.4.3 Action Research 29
3.4.4 Case Study Research 29
3.5 Data Collection 30
3.5.1 The interview technique 30
3.5.1.1 Structured interview 30
3.5.1.2 Structured interview 30
3.5.1.3 Semi-structured interview 31
3.5.2 Document Analysis 31
3.6 Data analysis 32
3.7 Ethical Consideration 33
3.8 Summary 33
CHAPTER FOUR 35
4.1 Introduction 35
4.2 Overview 35
4.3 Data Classification 36
4.3.1 Big data 36
4.3.2 An Enterprise 37
4.3.3 Big data architecture 38
4.4 Research questions 38
4.5 Data Collection 39
4.5.1 Peer-reviewed articles and books (Academic material) 39
4.5.2 Non-peer-reviewed articles (Non-academic material) 41
4.6 Summary 43
CHAPTER FIVE 44
5.1 Introduction 44
5.2 Overview 44
5.3 Data analysis and discussion 45
5.4 Findings and interpretation 53
5.5 The big data architecture 53
5.5.1 Relationship between actors and technology solution 54
5.5.2 Interaction towards an outcome 54
5.5.3 Allocative enabling support entities 55
5.6 Big data architecture for business continuity 56
5.7 Summary 57
CHAPTER SIX 58
6.1 Introduction 58

6.2 Summary of the chapters 58
6.3 Evaluation of research 59
6.4 Summary of the outcomes 61
6.4.1 To examine how big data are generated, stored, governed and used in the enterprises. 61
6.4.2 To examine and understand the factors that influence the design of big data architecture for enterprises. 61
6.4.3 To understand the architectural components (technical and non-technical factors) that suit big data in the context of enterprises. 62
6.5 Contribution of the study 62
6.5.1 Practical contribution 62
6.5.2 Theoretical contribution 63
6.5.3 Methodological Contribution 63
6.6 Benefits of the study 63
6.7 Limitations of the study 64
6.8 Recommendations 64
6.9 Further study 65
6.10 Summary 65
REFERENCES 66
APPENDIX A: ETHICAL CLEARANCE 84

CHAPTER ONE
INTRODUCTION

1.1 Introduction

Over the years, organisations have generated large volumes of data at an increasing rate. This has led to organisations realising the usefulness and value of their data for business continuity (Cockcroft & Russell, 2018). Organisations from different fields such as health, education, finance, and commerce are embracing data as an asset (Sandhu, 2021; Hassan et al., 2020). These organisations use the data to gain insights (Garoufallou & Gaitanou, 2021) that can be used for better decision-making (Mustapha, 2022), drive business growth, and stay competitive (Iyamu, 2018a). However, the speed, volume, and variety of the data they generate make it difficult for them to collect, store, process and analyse the data using traditional technologies (Jaiswal et al., 2020). Hence, the need to design an architecture for big data that encompasses the characteristics and can be scalable, secure, and handle complex datasets.

Big data is described by Goldstein et al. (2021) as complex data sets that consist of structured, semi-structured, and unstructured data. The author further explains that the complexity of big data comes from its V's, volume, velocity, variety, and veracity. This complexity requires new architectures, algorithms techniques and analytics to extract value and meaning out of big data (Garoufallou & Gaitanou, 2021; Wang et al., 2020). The value extracted from big data is used by organisations for various benefits. According to Barham (2017), big data assists organisations in developing new strategies using the insight gained from it. Hence, (Ravikumar, 2022) believes that big data has the potential to transform organisations to operate efficiently and successfully.

Another benefit highlighted by Nyikana and Iyamu (2022) is that big data is used by organisations to gain a better understanding of their customers. This helps the organisations to know their target market, to be innovative, and to create products and services that are based on customer needs.

While big data provides organisations with many benefits, one cannot shy away from its challenges. According to Mgudlwa and Iyamu (2018a), it is difficult to process, manage, and analyse big data. This is due to the high volume of data generated. As a result, organisations need to invest in big data analytic tools for data analysis. Nyikana and Iyamu (2022) highlight that selecting the appropriate tools is another challenge on its own. Additionally, there is a shortage of skilled data scientists, which poses a challenge since there is a demand for expertise to manage and analyse big data effectively (Mustapha, 2022). Moreover, the security and privacy of the data are another concern, as organisations need to ensure that the data is protected from unauthorised people (Sandhu, 2021). Lastly, data storage is a challenge, whereby traditional data architectures lack the flexibility and scalability to store big data (Jaiswal et al., 2020).

Primarily, many organisations do not architect their big data as they gather and use it. This causes many challenges such as complexity that manifests from security, privacy, storage, and management. Moreno et al. (2018) explain that the security problems result from the fact that, originally, big data was not provisioned to ensure security. Saddad et al. (2020) mention other challenges of not having big data architecture as data availability, reliability, scalability, and query performance. Khine and Wang (2017) argue that traditional data warehouses are not able to store and accommodate big data. Also, traditional data architectures are not capable of meeting the demands of big data (Bahri et al., 2018). Manogaran et al. (2018) explain that the challenges above can be overcome by the architecture of big data. Hence, the need to design a scalable and secure architecture for big data that can handle complex datasets.

Various factors are involved in the design of architecture for big data, from both technical and non-technical perspectives. From the technical front, there are various tools such as software and hardware. Roles, responsibilities, rules, and requirements are some of the factors involved from the non-technical viewpoint. The combination of these factors makes activity theory (AT) the most suitable to underpin the study, as no other theory combines the above factors as a focus. AT is used in Information systems (IS) studies to gain an understanding of human activities carried out in organisations (Engeström et al., 2016). AT is considered a framework that can be used to enhance design practices in studies such as Human Computer Interactions (Iyamu & Shaanika, 2019a). Additionally, AT is a framework that can be used to explain the structure and development of human activities within an environment. Nehemia et al. (2018) explain that the AT is used to study the activities involved in developing expert systems.

1.2 Research problem

Many organisations are increasingly generating large amounts of big data. This is a challenge primarily because the traditional data architectures are not designed to accommodate large volumes of data (Saggi & Jain, 2018), the different types (Oussous et al., 2018), and the velocity at which big data is generated (Chen at al., 2016). The problem is that many organisations do not have architectural design that encompasses the characteristics of big data that are rapidly increasing in their environments (Tschoppe & Drews, 2022). The problem manifests in many ways, including the two critical ones identified herein.

First, big data begins to lose value in the organisation, which affects its usefulness and operation. Second, the unprecedented nature of big data growth in some organisations adds to the complexities of the environment. According to Oussous et al (2018); Mishra and Sharma (2015), the complexity of big data affects the software, hardware, and network infrastructure. The problem threatens business continuity. Avci et al. (2020); Manogoran et al. (2018) argue that the lack of architectural design for big data results in the data being exposed to security risks such as cyber-attacks and data breaches. These problems are prohibitive and negatively affects the efficiency and effectiveness of operations and services where big data are applied. Additionally, a lack of architectural design for big data can compromise quality, including managing and governing big data in an organisation. Hence, organisations should have architectural designs for big data to enhance business continuity, increase the efficiency and effectiveness of operations and services, and improve stability in the use of big data.

1.3 Problem statement

Many organisations do not have architectural designs that can encompass and guide the rapidly increasing characteristics of big data in their environments.

1.4 Aim of the study

This study aims to design a big data architecture for enterprises, purposely to enhance business continuity and improve the efficiency and effectiveness of operations and services in the use of big data in an organisation.

1.4.1 Research objectives

Based on the aim, the research objectives are articulated as

i. To examine how big data are generated, stored, governed, and used in enterprises.
ii. To examine and understand the factors that influence the design of big data architecture for enterprises.
iii. To understand the architectural components (technical and non-technical factors) that suit big data in the context of enterprises. Based on the understanding of the components, an architecture of big data will be designed.

1.5 Research question

The research question and sub-questions for achieving the objectives of the study are as follows:

How can the architecture of big data be designed for enterprises, purposely to enhance business continuity and improve the efficiency and effectiveness of operations and services in the use of big data in an organisation?

1.5.1 Sub questions

i. How are big data generated, stored, governed and used in enterprises?
ii. What are the factors that can influence the design of big data architecture?
iii. What are the architectural components (technical and non-technical factors) that suit big data in the context of enterprises?

1.6 Literature review

This section presents the literature review of the study. It focuses on enterprises, big data in enterprises and the architecture of big data.

1.6.1 Enterprises

An enterprise is an entity with economic activities (Purchase et al., 2011) that combines resources such as humans, equipment and materials to produce a product or provide a service (Striy et al., 2019). According to Aliee et al. (2019), the resources in an enterprise are essential to coordinate tasks and disseminate information. In some quarters, enterprises are viewed as a complex system that consists of different but related domains (Smajevic & Bork, 2021). According to Kay (2017), enterprise and business are used interchangeably. However, a business focuses more on generating profits by selling goods or services (Purchase et al., 2011) while an enterprise is a group of businesses together to achieve a common goal (Jinoria, 2014).

Enterprises are increasingly using data to gain competitive advantage and maintain sustainability (Dezi et al., 2018). Additionally, enterprises use data to discover new insights, gain new ground, and uncover more opportunities (Dhaliwal & Shojania, 2018) to improve business decisions (Necsulescu, 2017). It is difficult to find a sector (or an enterprise) that does not employ data for its processes and activities, operationally or strategically. The education sector uses data to advance teaching and learning opportunities (Broos et al., 2017). The health sector uses data to monitor the health of the patients (Izonin et al., 2021) and to diagnose diseases (Mitani & Haneuse, 2020). The finance sector uses data to detect fraudulent activities (Aboud & Robinson, 2022) and to assess and manage risks (Cornwell et al., 2023).

Even though data is widely used in different sectors, many organisations require more than normal data. This is because normal data does not accommodate unstructured data sets such as images and videos. As a result, some organisations lose out on the potential of their data. Hence, the need to further explore the use of big data is increasingly crucial. Nyikana and Iyamu (2023) argue that normal data and big data are not the same. The authors further argue that the differences between the two concepts include scope, volume, and heterogeneity. Normal data is defined by Ahmed et al. (2017) as data with structured data sets, low volumes and constant velocities, while big data consists of huge volumes; structured, semi-structured and unstructured data sets; and high velocities (Oussous et al., 2018).

1.6.2 Big data in enterprises

New technologies and data revolution have made organisations generate large volumes of data from different sources at various levels of high speed (Avci et al., 2020). This change in data is known as the concept of big data. Big data is widely described using the 5 V's; volume, velocity, variety, veracity and value (Garoufallou & Gaitanou, 2021; Belov & Nikulchev, 2021; Wang et al., 2020; Al-Sai & Abdullah, 2019). Big data consists of structured, semi-structured and unstructured data sets (Nyikana & Iyamu, 2022). Saggi and Jane (2018) define big data as "a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high-velocity capture, discovery, and analysis". Hariri et al. (2019) refer to it as the driving force of innovation, competition, and productivity for organisations.

Big data is redesigning the way individuals and groups think and work, and how organisations conduct their businesses. Many sectors such as health, education, entertainment, finance, and energy are using big data in various ways to advance their operations (Wang et al., 2020; Avci et al., 2020; Mokhtari, 2019). One of the reasons is that potentially, big data helps organisations to streamline their processes and reduce bottlenecks (Dezi et al., 2018). This helps to increase the organisation's operational efficiency and effectiveness and improve its services (Gil et al.,2019). According to Al-Sai and Abdullah (2019), big data provides organisations with opportunities that create business value and growth. However, for an organisation to benefit from these opportunities, it needs to have appropriate tools, applications, resources, and people engagement, which can only be enacted through architecture design. Thereafter, big data can increase an organisation's capabilities to improve its overall performance and competitiveness (Hung et al., 2021).

Despite the numerous benefits that big data brings to organisations, some of which are stated above, it has its challenges. Gil et al. (2019) highlight integration as one of the challenges of big data. Another challenge is the complexity of big data sets, the traditional systems being unable to store, process and analyse big data (Garoufallou & Gaitanou, 2021; Wang et al., 2020). Sandhu (2021) claims that some organisations have moved their big data to the cloud environment to try and resolve the issues around storage and processing. However, that has created other challenges such as the inability to execute queries on the database. According to Al-Sai and Abdullah (2019), three major components can threaten an organisation not to benefit from big data: Technology, People, and Process. In addition, Sivarajah et al. (2017) add data management as another challenge. The author highlights some of the data management challenges as security and privacy issues. According to Sandhu (2021), it is difficult to visualise big data in real-time because of its diverse data sets. For these challenges to be addressed, new architectural designs for big data are required to address these complexity issues (Ruiz et al., 2021).

1.6.3 Architecture of big data

Architecture is the plan and design of a system, relationship and interaction of the components to each other and to the environment (Tschoppe & Drews, 2022). Iyamu (2022a) adds that the principles and the interaction of the components help to guide the design and governance of a system. Tupper (2011) describes architecture as the framework that guides the construction of a system from the beginning to the end. It allows an organisation to design its current business processes and accommodate future business processes.

From big data perspective, architecture is the design of the methods used to collect, store, secure and process big data (Yaseen & Obaid, 2020). The big data architecture provides a plan of how data flows from one point to another (Kalipe & Behera, 2019). Additionally, it focuses on the business objectives and requirements to provide a holistic data strategy. Also, big data architecture helps to improve the performance of an organisation by monitoring its operations and providing governance for interconnected systems (Costa & Santos, 2016). Avci et al. (2020) highlight that big data architecture improves data efficiency by processing and monitoring large amounts of heterogeneous data.

Although organisations are investing in big data, many of them are facing challenges primarily because they lack architecture design (Tschoppe & Drews, 2022). The development and construction of big data architecture and its design can be affected by the speedy evolution of big data technologies (Pääkkönenm & Pakkala, 2020). Furthermore, big data architectures can be complex, and difficult to design, build and maintain. This has hampered the efforts and attempts of many organisations over the years. Another challenge is the integration of the components of big data from a variety of data sources and formats (Bansal et al., 2022). Additionally, finding a big data architecture suitable for the organisation is another challenge. This is due to the requirements that need to be considered, such as business and application requirements (Avci et al., 2020).

1.6.4 Activity theory

Due to the various activities involved in the design of architecture, some of which are highlighted above, the activity theory (AT) was used to underpin the study. Activity theory is a socio-technical theory that has been adopted in IS studies in the last three decades (Iyamu, 2022b). The primary concern of the theory is the development of social activities (Shaanika & Iyamu, 2015). AT focuses on understanding the interactions and relationships that occur as activities are performed by humans (Iyamu & Shaanika, 2019a). Hence, Dennehy and Conboy (2017) refer to activity theory as a framework that is used to understand complex human activities within a social system. Nehemia et al. (2018) refer to AT as a theory of consciousness. The reason for that is because the activities performed are consciously planned.

As shown in Figure 1.1 AT consists of six components, which are subjects, objects, tools, rules, community and the division of labour (Park et al., 2013). The components are interconnected and interrelated, indicative of the arrows in the Figure. The interconnections and relationship of the components help to understand the overall activities of the system (Nehemia et al., 2018). Also, as expressed by AT, activities are not static, they constantly evolve due to the changes in the environment (Makovhololo et al., 2017).

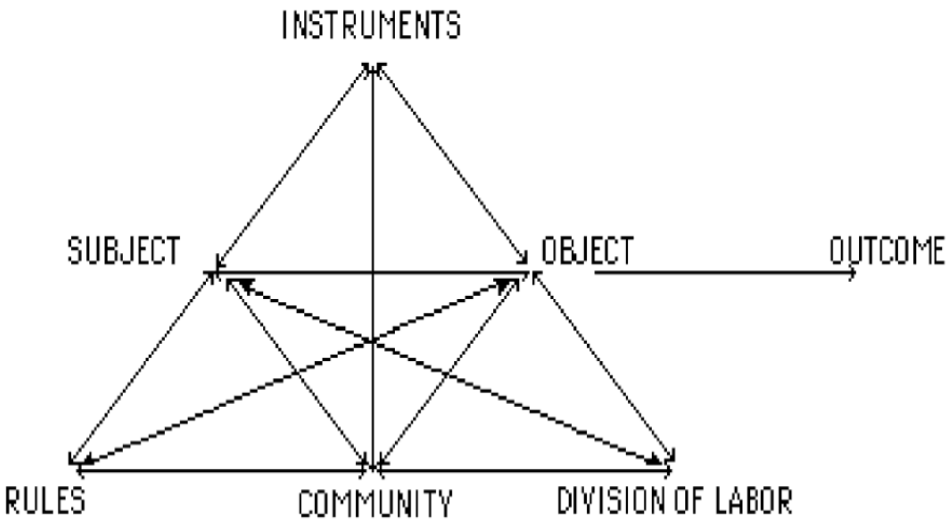

Figure 1.1: Activity Theory Model (Engeström et al., 2016)

The components of AT focus on specific areas but they are dependent on each other and cannot be used in isolation when studying a phenomenon (Iyamu & Shaanika, 2019a). Activity is the main focus of AT and is described by Nehemia et al. (2018) as the sum of actions carried out by human beings to achieve their goals. Makovhololo et al. (2017) define activity as an interaction between a subject and an object. A subject refers to an individual or group of people that perform an activity (Iyamu & Shaanika, 2019b). For a subject to act on an object, it is usually driven by motivation (Gedera & Williams, 2015). The object is the motive to initiate an activity. An object is defined by Iyamu (2022b) as the material or a problem on which the activity focuses.

Tools are the artefacts used in an activity to transform an object into an outcome (Sannino & Engeström, 2018). The tools vary depending on the context of the study. Nehemia et al. (2018)

mention some of the tools as machines, instruments, signs, procedures and laws. 'Community' is a collective of subjects that share the same goal. The community is governed by rules when performing activities (Dennehy and Conboy, 2017). The rules can be policies, procedures and regulations. The division of labour is concerned with assigning responsibilities to the community (Makovhololo et al., 2017).

1.7 Research methodology

Research methodology is a methodological process of gathering information (Patel & Patel, 2019) that can be used to answer a research problem (Kapur, 2018). Mukharjee (2019) describes it as a systematic guideline to conduct research. The research methodology consists of research approaches, methods, design, and data collection techniques (Kothari, 2020; Iyamu, 2022b). The methodology, as it applies to this study is discussed in the remainder of this section.

1.7.1 Research methods

Research methods are the tools used in a research study to achieve the objectives of the study (Mutudi, Nehemia & Iyamu, 2020). Patten (2017) refers to them as the building blocks of discovering new knowledge. The most commonly used research methods are qualitative and quantitative (Shaanika, 2020). These two methods can be combined to form a mixed method (Iyamu, 2022b).

Quantitative research is described by Gravetter and Forzano (2018) as a method that uses statistical instruments to examine relationships between variables and to test theories. Quantitative research focuses on objective reality that is independent of the phenomenon being studied (Yilmaz, 2013). The objective measures help to control bias. Additionally, quantitative research uses rigid and structured procedures to determine the extent of a problem or phenomenon (Kumar, 2018). This makes it difficult for participants to express their experiences, feelings and thoughts in their own words. Lastly, the focus of the quantitative method is more on quantity or numbers (Gupta & Gupta, 2022).

The mixed method is known to offer a holistic understanding of the phenomenon being studied by combining and balancing a wide range of knowledge from qualitative and quantitative methods (Baškarada & Koronios, 2018). Combining the qualitative and quantitative methods helps to minimise the weaknesses of each method and highlight their strengths (Kajamaa et al., 2020). For one to use the mixed method, they need to understand and have knowledge of both methods.

While qualitative research focuses on providing an in-depth understanding of the phenomenon being studied (Tungela et al., 2018), the qualitative method allows the researcher to examine and understand the opinions, perspectives and attitudes of human beings (Tümen-Akyildiz & Ahmed, 2021). Lauri (2019) adds that the qualitative method seeks to understand the causes that influence the behaviours and attitudes of human beings. Additionally, It helps to understand people's understanding of the phenomenon being studied. Given the aim, as stated in section 1.4, the qualitative method was considered more suitable for this study because it focuses on textual data (Buseto et al., 2020) to provide rich information about the phenomenon being studied (Mohajan, 2018).

The method was used in this study to understand the experiences and opinions of people about the architecture design of big data. Also, the qualitative method is exploratory in nature (Sovacool et al., 2018), making it suitable for exploring and understanding the factors that influence the architecture design of big data. Primarily, the qualitative method is associated with interpretivism. Thus, it helps to understand the interrelationship and interconnectedness between the influencing factors, based on which an architecture of big data was developed for enterprises.

1.7.2 Research design

Research design is a plan and structure that outlines how the study will be conducted (Leavy, 2022). Tungela (2021) describes the research design as a glue that holds all the elements of the research together to achieve its aim. Creswell and Poth (2016) mention some research designs from the qualitative perspective as grounded theory, ethnography, and case study. The case study design was employed in this study.

A case study is described by Yin (2017) as a design that enables an in-depth exploration of a phenomenon in its natural setting. Algozzine and Hancock (2017) state that the case study enables the researcher to gain an in-depth understanding of situations and the meanings attached to them by the people involved. Additionally, a case study provides a holistic investigation of a phenomenon being studied. Based on the objectives of this study, the case study design was followed to explore and understand the factors that influence the architecture of big data. Also, it was used to investigate how these factors guide the design of big data architecture.

The big data architecture was used as a case in this study. The investigation focused on designing an architecture for big data. Thus, the two concepts, big data and architectural design were investigated as they are the core aspects of the study. The concepts assisted in gaining a deeper understanding of the factors that influence the storing, accessing, and governance of big data in organisations.

1.7.3 Data collection techniques

Data collection is a process of gathering relevant information related to the phenomenon being studied using techniques (Iyamu, 2022b). In a qualitative study, the techniques may include document analysis, observations and interviews (Leech, 2017). For this study, document analysis was selected. The document analysis technique is explained below.

Document analysis is a technique that systematically reviews existing documents related to the phenomenon under investigation (Dalglish, Khalid & McMahon, 2020). The technique focuses on the analysis of documents such as books, newspaper articles, academic articles, and organisational reports (Morgan, 2022). Iyamu (2022b) suggests that documents may exist in either a manual form or in an electronic format. Document analysis can be used as a standalone technique or combined with other techniques (Berner-Rodoreda et al., 2020).

In this study, documents related to big data architecture were collected. They included strategic and operational documents in the areas of big data and architecture. Also, policy documents that focused on areas of big data such as data storage, data security, and data governance were obtained. The documents helped to give a comprehensive and holistic view of big data architecture. The analysis of the documents was two-fold; academic papers (peer-reviewed) and non-academic papers (white papers and green papers).

1.7.4 Data analysis

Data analysis is the systematic way of processing and transforming data into meaningful and valuable information (Taherdoost, 2022). During the process, themes, relationships, links, and key ideas are identified from the data (Mohajan, 2018), from which findings are reached, concisely and logically. According to Dufour and Richard (2019), in a qualitative study, data analysis is a complex stage of the research process because there are no universal methods or procedures to conduct the data analysis; it often relies on subjectivism. This is one of the main rationales for employing AT, to provide a frame and guide the analysis in this study.

The data analysis was conducted to achieve the objectives, as presented in section 1.4.1 above, and revisited here: (i) To examine how big data are generated, stored, governed, and used in enterprises; (ii) To examine and understand the factors that influence the design of big data architecture for enterprises; (iii) To understand the architectural components (technical and non-technical factors) that suit big data in the context of enterprises.

AT was used as a lens to guide the analysis of the data. The theory is explained in section 1.6.4. The analysis focused on three main areas. First, it is to gain an understanding of the activities involved in the use and management of big data in the organisation. This includes the procedure and regulation during the activities. Second, the relationships and interactions between the actors involved in the activities were examined. This includes how roles and responsibilities are assigned and executed in the activities of big data in the organisation. The third focus is on the governance of big data activities. This helped to gain a deeper understanding of how IT solutions are used or can be used at various steps, in trying to maintain uniformity, reduce complexity, and enable flexibility in the activities of big data in the organisation. Based on these focuses, the architecture of big data was designed.

1.8 Significance of the study

The study explores and examines the architecture design of big data that can be used by enterprises for business enhancement. Big data architecture design is significant for business. It provides guidance on how enterprises can effectively structure their big data systems to promote business enhancement. Also, data architects can use the findings of the study as a guide when designing big data architectures that can deal with complex big data sets. Furthermore, it can help enterprises align their big data systems with the aims and objectives of the organisation.

1.9 Contribution of the study

The contributions are threefold; practical, theoretical, and methodological. The practical contribution is at the enterprise (business) level, while the theoretical and methodological are for the academic domain. Thus, the study contributes to the existing knowledge of big data, from both business and academic perspectives.

Practically, the architecture can be used to guide the development of policies, standards, and principles in an enterprise. Based on the policies, standards, and principles big data can be better stored, retrieved, used, and managed. Better management reduces complexity and improves effectiveness and efficiency in the use of big data for service delivery.

From the academic domain standpoint, the architectural design forms part of the enterprise architecture research stream. In addition, the architecture design creates an opportunity for validation through further research. Very importantly it adds to the limited literature in this area of academic work. Researchers and students focusing on understanding the architecture of big data can benefit from the study. As of the time of this proposal, no study seemed to have employed activity theory to examine architectural design for big data. Thus, this contributes to the advancement of theory, from a methodological viewpoint.

1.10 Ethical considerations

Ethics are the values and principles that govern people to reduce risk when engaging in activities that involve humans and animals (Pastzor, 2015). The risks include body injury, psychological effects, and process interruption.

The rules that govern the research at Cape Peninsula University of Technology were observed. The ethical clearance letter was obtained from the Department of Information Technology to grant permission to conduct the research study. Thus, the ethics were applied as follows:

i. This research did not require the participation of individuals. Therefore, no personal information relating to any individual was used.
ii. The materials that were used in the study are in the public domain.
iii. Animals were not involved in the study. Therefore, no possible harm was caused to either the researcher or to animals.
iv. Data was collected from credible sources, which are listed in the reference section to provide transparency.
v. All materials including third-party materials used for this study were appropriately referenced, to avoid academic irregularity.

1.11 Budget and timeline

The budget and timeline are illustrated in Table 1.1. below. The budget and timeline is done to ensure that the research project is within the estimated budget and on schedule.

Table 1.1: Budget and timeline

Task	Item Needed	Estimated Cost	Target Date
Research Proposal	Laptop and WIFI	R35 000	May 2023
Literature Review	-	-	July 2023
Research Methodology	-	-	July 2023
Field Work	-	-	August 2023
Data Analysis	-	-	October 2023
Discussion of Findings	-	-	October 2023
Conclusion	-	-	November 2023
Thesis	Proof Reading	R5000	November 2023
	Printing and binding	R2000	November 2023
Total	R42,000		

1.12 Conclusion

The study looked at unpacking the factors that contribute to the design of big data architecture. Understanding these factors helps data architects in enterprises to know what they need to look at when designing big data architecture that is scalable and capable of accommodating complex data sets of big data. For future research, the study can be expanded to address the development and implementation of the architecture design of big data.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

A literature review is an evaluation of previous studies related to the research topic that has been conducted by other researchers and scholars (Papaioannou et al., 2016). Reviewing the previous research helps to understand what has been done on the topic and to gain a theoretical background about the research topic (Shaanika et al., 2017). Additionally, a literature review helps to identify gaps and provides direction for future research (Paul & Criado, 2020).

This chapter presents a review of the work related to this study. It, therefore, focuses on the core aspects of the study, which are big data and architectural design. In addition, a review of the theory, activity theory (AT), that underpins the study is conducted. It is structured into six main sections. It begins with the introduction, followed by a discussion about enterprises, in the context of the study. The third section covers big data in enterprises. In the fourth and fifth sections, the architecture of big data and activity theory are presented. The last section is the summary of the chapter.

2.2 Enterprises

An enterprise is a collaboration of people who form an organisation and work towards a common goal (Musukutwa, 2022). An organisation can be of any size, ranging from private and public (Sani et al., 2018) to non-profit organisations (NPOs) and non-government organisations (NGOs) (Lang et al., 2021). Brée and Karger (2022) argue that an enterprise is a complex system, that is not only formed by people but also includes other elements. In Striy's et al. (2019) explanation, other elements can include business processes and information technology (IT) solutions such as data, physical building, and equipment. All these elements are combined to achieve the goals and objectives of the organisation. The activities performed by people in an organisation constitute a business. Business is defined by Shafer et al. (2005) as value creation and generating profits from that value. The profits are generated from the activity of buying and selling goods or services (Alfonsius, 2021).

Organisations generate data daily, and the data has become a vital aspect of creating value. Rotondo and Quilligan (2020) define data as raw facts that can be meaningless. According to Dezi et al. (2018), for the data to become meaningful, useful and purposeful, it must be processed and analysed. Hence, Kitchin (2021) refers to data as the building block of information and knowledge. Jones (2019) suggests that data is generated from real-world events such as the activities of an organisation. Organisations use data to perform their daily activities, understand their customers and build new business models and processes (Moreno, 2018). This makes data critically important to many organisations. As a result, most organisations continually re-engineered their business processes, to create value, maintain sustainability, and improve competitiveness (Kitchin, 2021). Also, the use of data helps some organisations to create new insights, promote innovation and efficiency, and drive decision-making for improved efficiency and growth (Dhaliwal & Shojania, 2018; Dezi et al., 2018). Healthcare, education, and finance are some of the sectors (industries) where data are increasingly relied upon due to their unprecedented significance and cruciality (Hariri et al., 2019).

Thus, many organisations increasingly rely on data for various activities, including management, service delivery, and productivity (Iyamu, 2019). For example, in health care, data is used to form health protocols and policies (Svensson & Poveda, 2020) that help define the best practices to deliver effective and efficient healthcare services (Lee-Easton et al., 2022). Also in healthcare, data is relied on to create medical records that provide an overview of a patient’s health. The records are used for clinical decision-making and treatment planning (Mitani & Haneuse, 2020). However, challenges remain in the use of data for healthcare services (Iyamu, 2023). The challenges can be split into two. The first challenge is that many healthcare practitioners do not know what types of data to collect and how to do so (Iyamu & Nunu, 2021). The second problem is how to store, access, and make use of the data (Prasetyo et al., 2019). These challenges can be linked to a lack of understanding, which can be attributed to the fact that most healthcare practitioners are not data scientists or IT specialists.

In the education sector, data is used to assess the performance and progress of the students (Chitpin & Karoui, 2021). According to Mandinach and Abrams (2022), the use of data goes beyond assessing students’ performance and progress, it helps to gain deeper insights into their interests, strengths, and background. Some of these factors influence students’ performances (Fischer et al., 2020). Also, educators increasingly rely on data for teaching and managing activities. For example, during the COVID-19 disastrous period, learners and educators relied on information that they could extract from data, from anywhere and at any given time (Pratsri & Nilsook, 2020).

Financial institutions rely on data to assess the value chain (Mitragotri & Pal, 2019; Nagy et al., 2018), improve sales and marketing strategies (Kim et al., 2016), evaluate sustainability and competitiveness, and generate annual reports (Abdel-Basset et al., 2020; Oussous et al., 2018). Additionally, the banks use data to determine the risk associated with their activities, such as lending money to customers. This includes assessing investment opportunities and their location bounds and strategies (Svensson & Poveda, 2020). Despite this understanding, the use of data continues to pose challenges to many financial institutions. For example, some banks have lost huge amounts of revenue due to the inefficiency of data use, such as analysis or interpretation of the available data (Zeidy, 2022; Soltani Delgosha et al., 2021). Some of the challenges emanate from how data is stored, accessed, assessed, managed or governed in an organisation.

Despite the wide reliance on the use of data, many organisations are frequently or continually confronted with challenges, some of which are detrimental. In addition to the causes of the challenges identified in each of the sectors, other factors are the growing characteristics of data and methods used to ascertain its usefulness and meaningfulness. In recent years, data has grown into big data defined or differentiated by characteristics, which include volume, variety, veracity, and velocity (Nyikana & Iyamu, 2023).

Some of the challenges are due to organisations generating huge, unstructured, and diverse datasets at a high speed (Garoufallou & Gaitanou, 2021). This data is different from normal data, which consists of structured datasets with low volumes generated at a constant speed (Ahmed et al., 2017). The traditional methods are not equipped to store huge datasets and analyse the unprecedented variety, veracity, and velocity of data (Bahri et al., 2018). This is because relational databases which have rigid structures are employed (Belov & Nikulchev, 2021). Consequently, this makes it difficult for organisations to read the hidden patterns within the data. As a result, organisations lose out on the value of their data. Also, the normal data does not have prediction capabilities, it focuses on understanding current situations (Faraway & Augustin, 2018). In addressing these challenges, organisations resolve to explore and invest in the use of big data to enhance and improve service delivery and competitiveness.

2.3 Big data in enterprises

Big data is defined by its characteristics, which are value, volume, variety, veracity, and velocity, often referred to as the 5 Vs, as depicted in Figure 2.1. The volume refers to the huge data generated from multiple sources (Sandhu, 2021). Velocity is the speed at which the data is generated (Garoufallou & Gaitanou, 2021). Variety is the different formats of the data (Avci et al., 2020). Veracity refers to the governance and accuracy of the data (Saddad et al., 2020), while value is the conversion of data into meaningful information (Wang et al., 2020). The distinct focuses of the analytics pose a different type of challenge in selecting the most appropriate tool. Nyikana and Iyamu (2022) state that the value of big data can only be realised once the data is analysed using the most appropriate data analytics tools. Iyamu (2022a) explains how architecture guides the selection and use of IT solutions.

The selection and use of inappropriate analytics can have a detrimental effect on data value. Jones (2019) explains that the value of data depends on how complete and accurate it is, as well as how secure and reliable the storage environment is. Ferraris et al. (2019) argue that the value of data does not only depend on its accuracy and quality, but it also depends on other factors such as actors (humans) that are required to collect and analyse the data. Furthermore, the actors require governance in their use of processes for practices, which architecture provides.

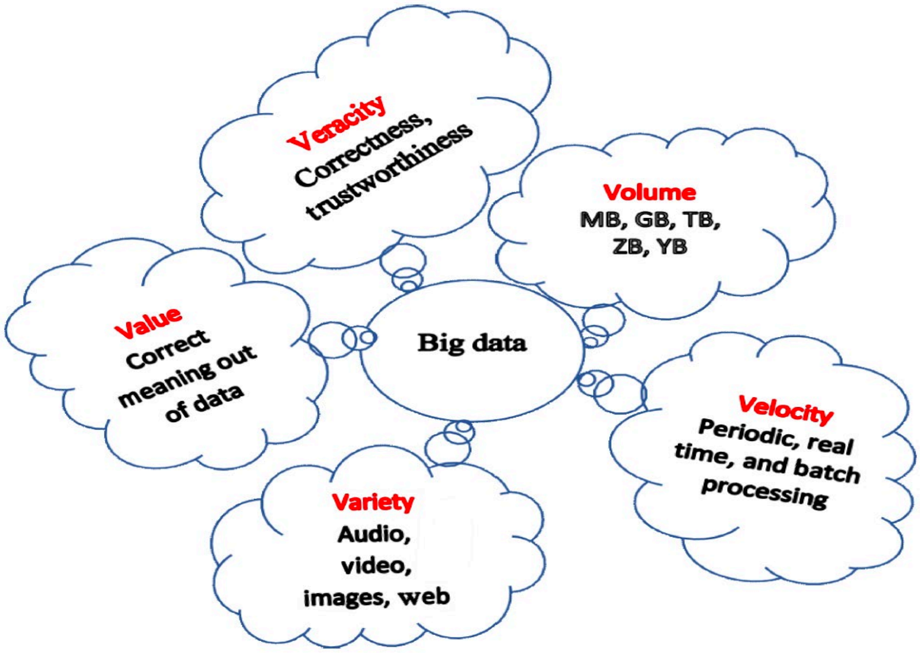

Figure 2.1: Characteristics of big data (Sandhu, 2021)

Big data consists of structured, semi-structured, and unstructured data sets (Tekaya et al., 2020). Structured datasets are data that are organised in rows and columns in a table (Lu et al., 2022). Unstructured data consists of social media content, photos, videos, and audio (Pratsri & Nilsook, 2020). There have been several definitions used to describe big data. For instance, Jones (2019) describes big data as datasets that traditional database software cannot capture, store, and analyse. Khine et al. (2018) consider big data a destructive technology since it has changed how data is used. Garoufallou and Gaitanou (2021) referred to it as the data that requires new architectures, analytic techniques, and algorithms to deal with the scale, diversity, and complexity of datasets.

Data complexity comes from huge datasets (Gil et al., 2019). Huge data sets require unique storage, management, analysis, and visualisation technologies. According to Sandhu (2021), different technologies and models have been designed and developed to deal with the storage and processing of big data to ensure business stability and continuity. Some of the technologies highlighted are NO-SQL for data management (Avci et al., 2020), Hadoop and Apache Spark for data processing and analysis (Jaiswal et al., 2020). Cloud computing resources have been widely recommended for storing, managing, and analysing data over the Internet (Manogaran et al., 2018). This is intended to handle real-time use of data remotely.

Also, with the use of cloud computing, organisations do not have to buy dedicated space and maintenance of software and hardware (Sandhu, 2021). According to Wang et al. (2020), Big data analysis depends on machine learning since it can mine the value from big data.

Big data analysis enables organisations to use their current data to create knowledge that can be transformed to gain a competitive advantage (Hajli et al., 2021). Additionally, it helps to enhance the performance of an organisation, promote business innovation, and increase sustainability (Zhang et al., 2022). According to Iyamu (2020), big data helps to uncover new patterns and make future predictions. Predictive analytics helps to create business value and provide decision support capabilities (Al-Sai & Abdullah, 2019).

Organisations in various industries from the private to the public sector make use of big data for different activities and processes (Avci et al., 2020). For instance, governments use big data to forecast social and economic changes such as unemployment levels and to improve service delivery to citizens (Blazquez & Domenech, 2018). In agriculture, it is used to determine the techniques that can be used to produce agricultural products (Prasetyo et al., 2019). This is done by evaluating data about the type of soil, the temperature and the biodiversity. Healthcare uses big data for diagnosing diseases and reducing healthcare costs (Manogaran et al., 2018) by reducing patient readmissions and preventing frequent emergency room (ER) visits (Pramanik et al., 2022). In marketing, big data generated from transactions are often used to establish customer needs and buy patterns (Al-Sai & Abdullah, 2019). In education, it improves educational effectiveness by promoting data-driven approaches to teaching and learning (Fischer et al., 2020). However, the gathering and use of big data by organisations is not by default; it entails approaches that are governed by standards and principles (Iyamu, 2023).

As identified above, there are many challenges in collecting, using, and managing big data in an organisation (Jeske & Calvard, 2020). The challenges are of both technical and non-technical nature. From the technical perspective, the host infrastructure such as software, security and hardware are discussed (Zhang et al., 2022; Saddad et al., 2020; Moreno et al., 2018). The heterogeneous nature of big data makes the enabling infrastructure more critical and requires more attention such as the architectural guidelines. Jones (2019) argues that some of the challenges can be prohibitive. Therefore, the cost of collecting, analysing, and storing big data should not be taken for granted.

The heterogeneity of big data poses a challenge when it comes to data analysis and integration (Iyamu, 2020). This is due to the differences in the devices used, and the types of data generated. According to Siyal et al. (2019), another challenge posed by big data is security and privacy. Cyber-attacks and the risk of data being hacked and stolen are increasing. This results in personal data being illegally shared without consent (Ravikumar, 2022). Hence, Pratsri and Nilsook (2020) emphasise the importance of determining access rights to protect the use of data by authorised people. Thus, the properties of big data and the challenges it poses require an architecture that is designed to collect, analyse, and store large volumes, velocity, and variety of data (Ruiz et al., 2021; Blazquez & Domenech, 2018).

2.4 Architecture of big data

An architecture comprises development and implementation, which is enabled through its functioning and often umbrellaed by enterprise architecture (EA) (Chitsa & Iyamu, 2020; Zhou et al., 2020). EA often encompasses technical (application and technology) and non-technical (information and business) architectures (Dumitriu & Popescu, 2020; Shaanika & Iyamu, 2015). In the technology domain, architecture constitutes design, structure, and governance (Hajli et al., 2021). The functioning of an architecture entails its relationship with components such as the objectives of the organisation, rules, business processes, data, people, and technology (Shaanika, 2019). This means that each of the components influences how the architecture is designed and employed. Currently, under the umbrella of EA, the technology domains do not include big data architecture.

Enterprise architecture (EA) seeks to align technology and business goals to ensure business units within the organisation work closely together (Zhou et al., 2020). Tschoppe and Drews (2022) describe EA as a holistic approach to designing the entire IT solutions of an organisation including business processes and their interrelationships. This is done by managing and mitigating challenges and complexities that exist within business units (Nehemia & Zondani, 2021). These types of functions make EA an operational and strategic tool or approach. Lapalme et al. (2016) consider EA as a strategy to align, improve and manage complicated business processes and Information Systems (IS) / Information Technology (IT) in an organisation. Chitsa and Iyamu (2020) explain that EA helps to provide governance, principles, and policies that can be used to guide the development of strategic documents of a system.

However, for an organisation to benefit from EA implementation and practice, it is essential that it ensures interrelation between the influencing factors. This has been an ongoing challenge highlighted in many studies over the years as influencers of EA practice in organisations (Iyamu, 2013a; Shaanika & Iyamu, 2018; Ylinen & Pekkola, 2020). This remains a challenge faced by the organisations (Iyamu, 2022a) leading to some organisations outsourcing the development of architectures and relying on consultants (Shaanika & Iyamu, 2018). The challenges associated with EA are revisited because they can extend to the design, development, and implementation of big data architecture, which this study focuses on.

Despite the challenges, EA helps to ensure uniformity and maintain consistency in the evolving IS/IT. According to Lankhorst (2017), evolving technologies force organisations to design and develop new architectures. This is evident with data architectures; whereby new architectures are required to deal with the characteristics of big data (Ruiz et al., 2021). Big data architecture is the design of an approach that structures and governs how to collect, store, secure, process, and visualise big data in an environment (Yaseen & Obaid, 2020). Also, big data architecture provides a plan of how data flows from one point to another (Kalipe & Behera, 2019).

Big data architecture consists of layers that have different functions (Benhlima, 2018). The functions and components of each layer are defined based on the organisational requirements and technological needs (Kalipe & Behera, 2019). Also, the layers and components are influenced by the characteristics of big data. According to Wang et al. (2020), the layers include collection, storage, processing, and visualisation of datasets. Governance is considered an important layer because of its cruciality, such as defining quality standards, security principles, compliance policies, and determining access to data (Saddad et al., 2020; Farooqi et al., 2019). Mustapha (2022) states that when big data architecture does not have a governance layer, the overall effectiveness and performance of the system can be affected. This results in challenging implications such as a lack of compliance, a lack of standardized processes, and security vulnerabilities.

The design of big data architecture that is based on organisational requirements generates value and improves competitiveness (Blazquez & Domenech, 2018). From another perspective, Mustapha (2022) suggests that an organisation that deploys scalable and flexible big data architecture tend to be more responsive to rapid changes in its business environment and has a positive impact on its overall performance. However, organisations struggle to design or find a big data architecture that is suitable for their environment (Ruiz et al., 2021). Avci et al. (2020) link this challenge to fit or align with the application and business requirements of the organisations, which are the technical and non-technical factors that influence the design of a big data architecture.

Also, big data architectures can be difficult to implement (Farooqi et al., 2019) because of various factors required to support the adoption of new architecture (Mustapha, 2022). Additionally, finding the right tools to utilise big data architecture is a challenge (Gökalp et al., 2019). Furthermore, the integration of the existing models with new models such as big data architecture is another challenge highlighted by Bansal et al. (2022) and Iyamu (2013a). It is, therefore, essential to consider these challenging factors in designing an architecture for big data, which did not seem to be in existence at the time of this study (Tschoppe & Drews, 2022; Pääkkönenm & Pakkala, 2020; Saggi & Jain, 2018).

Different big data architectures have been proposed and developed for various purposes. An example is the lambda and kappa architecture, which is concerned with processing real-time scalable big data (Barradas et al., 2022). There is also the Hadoop framework, which consists of a Hadoop distributed file system (HDFS), to store huge amounts of structured and unstructured data sets (Oussous et al, 2018). Another example is MapReduce, whose primary focus is to process data sets (Farooqi et al., 2019). Boumlik and Bahaj (2018) state that HDFS and MapReduce are used in parallel to process, store, and retrieve large data volumes.

In attempts to address existing gaps, Wen et al. (2021) propose a big data architecture that is concerned with integrating big data from different data sources' perspectives. Filaly et al. (2022) and Thota et al. (2017), designed a big data architecture focusing on ensuring the security of the data from attacks and vulnerabilities. Even though there are these different architectures that have been designed or developed, the challenges persist. This means none of the existing architecture seems to cover all the characteristics of big data. Hence, Mostefaoui et al. (2022) highlight the need for an architecture that covers all the characteristics and layers of big data. This is to have a holistic architecture that addresses the huge volume, rapidity of veracity, fluidity of variety, contextualisation of value, and precedented velocity of datasets.

2.5 Theoretical framework

A theoretical framework is defined by Varpio et al. (2020) as logically organised and interconnected concepts that researchers use to plan, structure, and execute their research studies. In Kivunja's (2018) explanation, theoretical frameworks are specialised lenses that help to examine and analyse data and interpret findings. Thus, they are increasingly used to underpin IS studies (Lynch et al., 2020). There are many theoretical frameworks, which include actor-network theory, contingency theory, and diffusion of innovation (Sekgweleo et al., 2017; Mkhomazi & Iyamu, 2013b). The suitability of a theory depends on the objectives of the study (Iyamu, 2022b). Based on the objectives as stated in Chapter One, activity theory is selected to underpin this study.

2.5.1 Activity theory

Activity theory (AT) is a theoretical framework that has been used in many studies for over three decades to understand the relationships and interactions that happen between human beings and the world (Kaptelinin et al., 1995). Iyamu and Shaanika (2019a); and Dennehy and Conboy (2017) highlight that using AT as a framework helps to understand the structure and development of human activities and how they are influenced by social context in an environment. Hence, AT is used in this study to understand the human activities that take place when designing the architecture of big data in an organisation. It is also used to analyse the factors that influence the design of big data architecture.

AT is a socio-technical theory concerned with the development of social activities (Iyamu & Shaanika, 2019a). The theory assists in gaining an understanding of how human beings engage in activities within an environment (Iyamu, 2022b). This helps to understand why certain things happen the way that they do in a social system (an environment). An activity is a collection of actions that are performed by human beings upon objects using tools to achieve an outcome (Engeström & Pyörälä, 2021). According to Kaptelinin and Nardi (2006), an activity is not fixed; it evolves due to environmental changes. Hence, Dennehy and Conboy (2017) highlight the importance of studying an activity in the environment where it is carried out.

As shown in Figure 2.2, the AT model consists of six components which are subjects, objects, tools (instruments), rules, community, and division of labour (Engeström et al., 2016; Park et al., 2013). The components are distinct but cannot be applied in isolation as they are interconnected under the phenomenon being studied, Iyamu (2022b) argued. The subject, object, and tools are often considered as the core concepts of an activity system (Lioutas et al., 2019). This could be attributed to their mediated inseparable connection.

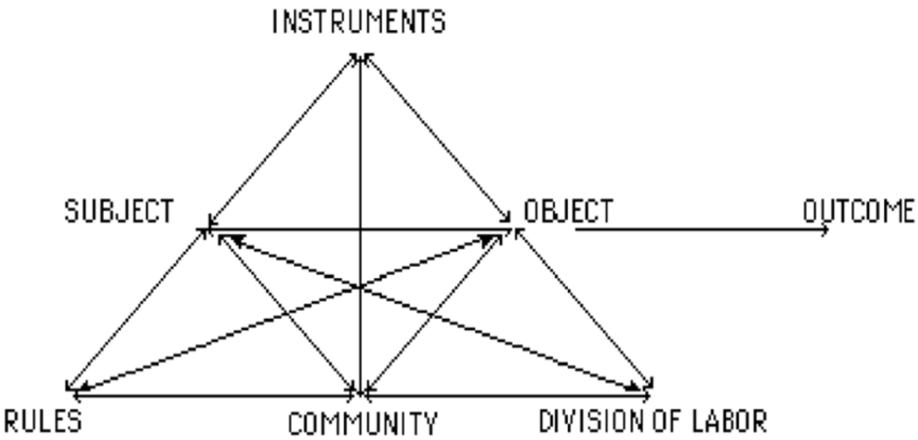

Figure 2.2: Activity Theory Model (Engeström et al., 2016)

In AT, a subject is a human being or a collective of people involved in an activity, also referred to as an actor (Sannino & Engeström, 2018; Shaanika & Iyamu, 2015). The activities carried out by subjects are consciously planned. This is to ensure that the activities have a purpose and are not aimless (Dennehy & Conboy, 2017). Nehemia-Maletzky et al. (2018) state that human consciousness is the basic principle of AT. Nardi (1996) defines consciousness as "a phenomenon that unifies attention, intention, memory, reasoning, and speech". Nehemia-

Maletzky et al. (2018) state that conscious planning does not determine the outcome of the activity. These are critical in gaining an understanding of roles and responsibilities including interrelationships, in designing a big data architecture (Garoufallou & Gaitanou, 2021).

The outcome can either be intended or unintended by the subjects (Hasan & Kazlauskas, 2014). This is attributed to the fact that the outcome is affected by the actions that take place in an activity (Nardi, 1996). The object component of the AT model is the motive for carrying out an activity, and it can be tangible or intangible (Iyamu, 2022b). According to Sannino and Engeström (2018), the object gives an activity a sense of direction and significance.

Subjects depend on tools to mediate with the object. Tools can be technical or psychological artefacts (Hasan & Kazlauskas, 2014). Technical tools such as computers intend to manipulate physical objects, while psychological tools such as language are used by human beings to influence each other (Makovhololo et al., 2017). These are fundament elements and factors that can manipulate or influence an IT solution such as the design of architecture for big data. Er et al. (2010) suggest that tools shape the way subjects interact with objects and influence the outcome of the activity. This is because tools have an enabling and constraining function in them, whereby they can transform or limit the object, depending on motive and how the actor employs them (Kaptelinin et al., 1995).

In AT, rules are control mechanisms, which can include policies, regulations, and legislations that guide and govern how subjects perform their activities (Kelly, 2018). Additionally, Makovhololo et al. (2017) suggest that rules help to maintain order and control conflicts within an activity. Also, rules are used to manage the interactions between actors during the process of allocating tasks, roles and responsibilities among community members (Karanasios & Allen, 2013). Community is defined by Iyamu and Shaanika (2019a) as a collective of individuals in a social system working towards the same goal. Division of labour refers to the allocation of tasks and responsibilities among community members (Lioutas et al., 2019). The components of AT as described and discussed above are usually norms in IT solutions, hence, the theory is increasingly connected with IS research.

2.5.2 Activity theory and Information Systems Research

Activity theory (AT) is categorised as a sociotechnical theory in IS research (Nehemia-Maletzky et al., 2018; Sekgweleo et al., 2017). About five years ago, Makovhololo et al. (2017) claimed that the theory has been applied in more than three million studies. This can be attributed to its relevance in the field of IS research, from various standpoints. Iyamu and Shaanika (2019a) argued that as a sociotechnical theory, AT provides a fresh perspective in IS studies.

From various spheres, AT is increasingly embraced in IS research. According to Nehemia-Maletzky et al. (2018), the complexity of IS/IT solutions requires the sociotechnical theory to dissect some of the challenges. The socio-technical nature of the theory assists in covering factors such as people, processes, and technology. Thus, AT is often used as a lens to guide data analysis and interpretation of findings (Iyamu, 2013b). Primarily, this is because the theory helps to gain a deeper understanding of a social environment that is designed and managed by people using tools (Iyamu & Shaanika, 2019a).

AT is used in IS studies to understand the interactions and relationships of people when carrying out activities of developing, implementing and using IT solutions (Iyamu, 2022b). Additionally, the theory has been used in many studies, to unpack the meaning of technology to people and how technology is affected and affects individuals and groups of people within an environment (Kaptelinin & Nardi, 2018). Some of the studies include those focusing on developing expert systems (Nehemia et al., 2018); enhancing design practices in Human Computer Interactions (Iyamu & Shaanika, 2019a); software development (Dennehy & Conboy, 2017); and the role of IS in mobile technologies (Karanasios & Allen, 2014).

2.6 Summary

The chapter presents a review of the literature that is related to the study. It therefore focuses on the core aspects of the study, which are enterprises, big data in organisations, and big data architecture. The most relevant literature is consulted. Also, a comprehensive review of the activity theory is conducted. This includes the relationship between AT and IS research. The Chapter is well structured based on its sequence for ease of understanding. The next chapter covers the research methodology.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

Research methodology describes the entire process undertaken in a study (Iyamu, 2022b). Kapur (2018) describes it as a systematic process of gathering information that focuses on addressing research questions. According to Vass et al. (2017), research methodology consists of components, which include approaches, methods, and techniques used to study a phenomenon. Based on the objectives of the study as stated in Chapter One, approaches, methods, and techniques were selected and employed. This chapter provides a detailed explanation of how the components were selected and employed in achieving the objectives of the research.

The chapter is organised into eight sections. The first section is introductory. The second and third sections cover the research approach and research methods, respectively. In the fourth section, the research design is discussed. This is followed by an explanation of how the data were collected as dictated by the aim and objectives of the study. The analysis of the data is described in the sixth section. Thereafter, the ethics, which ensure the credibility of the study is covered in the seventh section. Finally, section eight presents a summary of the chapter.

3.2 Research approach

The research approach is concerned with the basic logic and reasoning that a researcher applies in a study (Woo et al., 2017). It is also viewed as the plan a researcher employs, which leads to the type of data collected and analysed (Creswell & Creswell, 2017). The three main research approaches that are commonly applied in the information systems (IS) field are inductive, deductive, and abductive (Saunders et al., 2019). According to Beese et al. (2019), inductive and deductive are the most used approaches in IS studies. The inductive approach is a way of reasoning that focuses on building a theory while the deductive guides testing a theory or theories (Saunders et al., 2019).

3.2.1 Deductive Approach

The deductive approach moves from general explanations to specific case or cases (Behfar & Okhuysen, 2018). It is described by Parmaxi et al. (2016) as a top-down approach. Williamson and Johanson (2017) refer to the deductive approach as a theory of testing. This means that the approach is selected and employed to test existing concepts and patterns using new empirical data (Adekoya et al., 2019). In this study, although new empirical data were gathered, it did not aim to test any theory, as explained in Chapter One.

However, argues that testing a theory is not the only concern of the deductive approach. It also focuses on extending and improving the existing theory. The expansion of the focus of the approach can be narrowed back to testing. This is because a theory can only be extended or improved if it is tested to understand its current state or statute. Such understanding can only be based on facts and figures. Thus, Melnikovas (2018) argues that the deductive approach is objective in nature, and it is employed in quantitative studies seeking to quantify facts. Based on the synopsis provided above, the approach was not deemed appropriate for this study. As discussed in Chapter One, this study focuses on gaining an idea of various factors that influence the governance of big data architecture.

3.2.2 Inductive Approach

The inductive approach allows reasoning that is multidimensional, which means a collective of many realities to form a cohesive whole. Woo et al. (2017) described the approach as a bottom-up approach because it focuses on moving from specific cases to general explanations. The primary concern of the inductive approach is to build a theory (Behfar & Okhuysen, 2018). This means that from the data collected and analysed, a theory emerges in the study (Myers, 2013). Also, the inductive approach allows a researcher to create patterns and relationships by observing a phenomenon (Woo et al., 2017). However, creating patterns and relationships is not only through observation but any other techniques that are associated with the qualitative paradigm can be used. According to Saunders et al. (2019), the inductive approach is mostly employed in qualitative studies.

Thus, the inductive approach was selected for this study. Wolski and Gomolińska (2020) state that inductive reasoning is applied to gain knowledge from data. Primarily, the rationale is twofold. Firstly, the study does not have an interest in testing a theory but in developing an architecture, which adds to knowledge. McAbee et al. (2017) argue that inductive reasoning can also be used to analyse and assess the usefulness of big data in organisations. Secondly, the study seeks to develop a theory by designing an architecture for big data.

3.3 Research methods

Research methods refer to the tools and techniques used to conduct a study (Walliman, 2021). The most used methods are quantitative and qualitative. The quantitative method was developed for natural science studies while the qualitative methods are concerned with social science studies (Myers, 2013). Additionally, the two methods can be combined, which is referred to as a mixed method (Creswell & Creswell, 2017). The methods are associated with various research approaches discussed in section 3.2.

3.3.1 Quantitative method

The quantitative method is a systematic process used to describe and test relationships between variables (Bloomfield & Fisher, 2019). When testing the relationship between the variables, statistical instruments are used to examine and analyse the numeric data (Gravetter & Forzano, 2018). Thus, the quantitative method is objective by nature, and it seeks to minimise biases (Bloomfield & Fisher, 2019). Davies and Fisher (2018) state that the quantitative method is influenced by the positivism paradigm, and it is rooted in a single reality of truth. Hence, it does not allow participants to express their feelings, experiences, and thoughts in their own words (Gupta & Gupta, 2022). This is in line with Bloomfield and Fisher's (2019) argument that the quantitative method is underpinned by assumptions such as objectivity, deduction, and single reality. This is different from those of qualitative method, subjectivity, induction, and multiple realities (Keele, 2010).

The quantitative method was not chosen for this study because it mainly focuses on numbers and measurements to quantify variables, as explained above. This study is interested in understanding people's thoughts, feelings and experiences, which cannot be quantified. Hence, the qualitative method was employed, instead.

3.3.2 Qualitative method

The qualitative method seeks to explore and understand people's experiences and views about the phenomenon (Bloomfield & Fisher, 2019). The method is employed when there is little that is known about the phenomenon, and the researcher wants to gain a deep understanding of that specific phenomenon (Brannen, 2017). According to Saunders (2019), text and images are analysed to understand and make sense of social events, issues, and practices. Lauri (2019) adds that the qualitative method further explores the factors that influence the behaviours and attitudes of human beings.

The qualitative method is associated with the interpretive approach (Iyamu, 2018). This means that the meaning and understanding of the phenomenon being studied are constructed subjectively (Saunders, 2019). This makes continual changes in reality based on the many interpretations that are associated with the phenomenon (Keele, 2010). Table 3.1. presents a summary of the differences between the quantitative and qualitative methods. However, this study or chapter is not about comparing the two methods. This table is intended to help provide clarity on the choice of method selected for this study.

Table 3.1. Characteristics of qualitative and quantitative research (Merriam & Tisdell, 2015)

Point of Comparison	*Qualitative Research*	*Quantitative Research*
Focus of research	Quality (nature, essence)	Quantity (how much, how many)
Philosophical roots	Phenomenology, symbolic interactionism, constructivism	Positivism, logical empiricism, realism
Associated phrases	Fieldwork, ethnographic, naturalistic, grounded, constructivist	Experimental, empirical, statistical
Goal of investigation	Understanding, description, discovery, meaning, hypothesis generating	Prediction, control, description, confirmation, hypothesis testing
Design characteristics	Flexible, evolving, emergent	Predetermined, structured
Sample	Small, nonrandom, purposeful, theoretical	Large, random, representative
Data collection	Researcher as primary instrument, interviews, observations, documents	Inanimate instruments (scales, tests, surveys, questionnaires, computers)
Primary mode of analysis	Inductive, constant comparative method	Deductive, statistical
Findings	Comprehensive, holistic, expansive, richly descriptive	Precise, numerical

The qualitative method was applied to fulfil the objectives of the study. The method helps to gain a better understanding of the meaning associated with the phenomenon under investigation from the documents collected. This helps to gain an in-depth understanding of why things happen the way they do. Furthermore, the qualitative method is exploratory by nature (Sovacool et al., 2018), making it suitable for exploring and understanding the factors that influence the architectural design of big data.

3.3.3 Mixed method

The mixed method is a combination of qualitative and quantitative methods (Creswell & Creswell, 2017). The mixed method is used when the aim of the study and research questions cannot be answered using a single method (Brannen, 2017). Venkatesh et al. (2013) explains that the qualitative and qualitative methods can be used simultaneously, or one after another whereby the results of one method can inform the other method. This makes it crucial for the researcher to understand and have knowledge of both methods. However, this does not apply to this study because the qualitative method alone is deemed suitable.

Combining the qualitative and quantitative methods helps to minimise the weaknesses of each method and highlight their strengths (Kajamaa et al., 2020). In this study, no weakness was identified in the use of the qualitative method. The mixed method is known to offer a holistic understanding of the phenomenon being studied by combining and balancing a wide range of knowledge from qualitative and quantitative methods (Baškarada & Koronios, 2018). The knowledge gathered from the singular use of the qualitative method was sufficient to develop the architecture for big data. The reasons for not selecting the quantitative method are the same for deeming the mixed method inappropriate for this study.

3.4 Research design

Research design is the blueprint for conducting research (Bloomfield & Fisher,2019). The design provides a clear plan that covers the scope or boundaries of the research within which data are collected (Kapur, 2018). Bhattacherjee (2012) emphasises that it is important to select a research design that aligns with the objectives of the study, and not the one the researcher is comfortable with. The research designs associated with the qualitative method include ethnography, grounded theory, action research, and case study (Saunders, 2019).

3.4.1 Ethnography

Ethnography is defined by Mohajan (2018) as the process of studying and documenting people's cultural behaviours to understand them in their natural environment. The environment could be a community or an organisation. Ethnography is also understood to be a design that focuses on society and culture (Merriam & Tisdell, 2015). The primary aim of ethnography is to understand how culture works in practices and rituals performed daily by humans (Madden, 2010). Iyamu (2022b) states that to understand the culture of a specific group of people, the researcher must follow the group for a certain period. During that time, the day-to-day behaviour of the people is observed and interpreted to understand their norms, values, beliefs, and social interactions (Van der Merwe et al., 2020).

This study is concerned with big data, architecture, and enterprises. This means that the study is neither about society nor culture. Also, the study does not focus on values, beliefs, norms, or social interactions. Therefore, ethnography could not be selected as a design for this study.

3.4.2 Grounded Theory

The grounded theory seeks to investigate social processes and interactions that take place in the lives of people (Merriam & Tisdell, 2015). The focus of the grounded theory is to develop a theory from empirical data collected in a looping process (Iyamu, 2022b). Although this research relied on empirical data, a loop process was not appropriate. This is because such a process could imbibe biases in data collection.

Also, the pieces of data collected using grounded theory are constantly analysed and compared to develop a theory (Mohajan, 2018). Based on the objectives of this study, data comparison was not needed. From another perspective, Bhattacherjee (2012) argues that grounded theory requires the researcher to eliminate prior knowledge and biases to allow the theory to be shaped by observed evidence. In developing the architecture for big data, prior knowledge of the influencing factors was critical. In this context, prior knowledge was required to gain a deeper understanding of the implications of the influencing factors.

3.4.3 Action Research

Action research is concerned with finding a solution to an identified practical problem (Kirongo & Odoyo, 2020). The design is often employed to generate or capture actions and interpret their outcomes to create knowledge (Iyamu, 2022b). The results are used to improve or change certain areas in society (Malmi, 2016). However, Kapur (2018) states that it is difficult to generalise the results of an action because they are linked to a specific problem and bound to a particular situation and environment.

Action research was not employed in this study because the study aims to gain an in-depth understanding of a phenomenon, rather than solve a current practical problem. Iyamu (2022b) states that the primary objective of action research is to test a theory or hypothesis. Hence it is not suitable for this study since it aims to develop a theory.

3.4.4 Case Study Research

The case study approach is described by Yin (2017) as a design that enables an in-depth exploration of a phenomenon in its natural setting. A case can be an object (such as metal, mental or technology) or a subject (e.g., a community, an organisation, or a person). Iyamu (2018) explains that the case study approach provides uniqueness to the phenomenon under investigation. The approach can be employed as a design to unpack and gain a better understanding of complex issues in a social unit such as a person, community, or organisation (Kapur, 2018). Algozzine and Hancock (2017) suggest that the case study enables an in-depth understanding of situations and the meanings associated with them by the people involved. This can be linked to the nature of the questions such as why, when, where or how they are employed to get responses from participants (Hancock et al., 2021). Additionally, the in-depth nature of the case study provides a holistic investigation of the phenomenon being studied.

Based on the focus and strength described above, the case study approach was selected as the design for this study. The big data architecture was used as a case to gain a deeper understanding of it and the factors that influence its design. Three factors influenced the selection of big data architecture as a case. The factors are relevance, current discourse, and gap, discussed as follows:

i. Relevance – Big data architecture is relevant to organisations. The relevance of it is associated with its ability to respond to rapid changes in the business environment caused by big data. In today's data-driven world, big data architecture is one of the driving forces of innovation, competition, and productivity for organisations.
ii. Current discourse – Big data architecture is currently a discourse in both business and academic environments. The concept seeks to respond to the challenges, as well as the opportunities that exist in many organisations.

iii. Gap – There is a need for big data architecture in organisations. The gap has been identified from the literature. The existing data architectures are not capable of handling big data with its characteristics.

Based on the justification provided above, data was collected. A sample of the data are presented in Tables 4.1 and 4.2, in Chapter Four.

3.5 Data Collection

The data collection is a process of gathering relevant information related to the phenomenon being studied using techniques (Iyamu, 2022b). In a qualitative study, the techniques may include document analysis, observations, and interviews (Leech, 2017). For this study, document analysis was selected.

3.5.1 The interview technique

An interview is defined by Bihu (2020) as a purposeful conversation between two or more people, the interviewer and the interviewees. According to Taherdoost (2022), interviews are commonly used because they provide a normal and comfortable environment for the participants. Interviews are utilised to gather valuable insights into an individual's subjective experiences and opinions (Busetto et al., 2020). Wahyuni (2012) highlights the importance of using interviews to engage with practitioners to understand the practices that take place in organisations. According to Stuckey (2013), there are 3 types of interviews: structured, unstructured, and semi-structured interviews.

3.5.1.1 Structured interview

Structured interviews are guided by a predetermined set of questions that are arranged before the interview (Kumar, 2011). The questions are fixed and closed-ended, and they do not allow follow-ups and probing (Pathak & Intratat, 2012). According to Wethington and McDarby (2015), the questions have predetermined responses such as yes or no. Also, the responses are systematically analysed and quantified. The structured interview technique was not employed in the study because it does not allow the participants to express their views and experiences in their own words as they are guided by the predefined responses.

3.5.1.2 Structured interview

An unstructured interview consists of open-ended questions that are not set before the interview (Low, Saks & Allsop, 2019). Instead, the questions are created as the interview progresses (Vanderstoep & Johnson, 2009). The data collected using this technique lacks structure and analysing it can be time-consuming (Mulcahy, Rossner & Tsalapatanis 2021). Leedy and Ormrod (2014) argue that the lack of structure helps to provide more insights and fruitful data because the participants might provide data that the interviewer was not planning to ask. Even though this technique provides insightful data, it was not employed in this study because the questions for the interviews needed to be set beforehand to ensure the data collected aligns and addresses the research questions and objectives of the study.

3.5.1.3 Semi-structured interview

In using the semi-structured interview technique, the questions are formulated beforehand but may be amended depending on the flow of the interview or responses from the interviewees (McIntosh & Morse, 2015). The flexibility of the technique allows the researcher to probe, rephrase and restructure the questions during the interview (Shaanika & Iyamu, 2018). This type of flexibility enriches the quality of the data. Iyamu (2022b) states that probing requires participants to build on their responses to uncover new knowledge. This provides an opportunity to ask additional follow-up questions or clarifications to responses during the interview process.

In addition, the technique allows the discussion of complex issues to take place and to gain new insights (Iyamu, 2018). The semi-structured interview technique was suitable for the study. However, it was not employed because there was no organisation using big data or that designed architecture for the concept where data could be collected, at the time of this study. This could be attributed to the newness concept of big data architecture in the South African environment. For this reason, the document analysis technique became most appropriate and was employed.

3.5.2 Document Analysis

Document analysis is a technique that systematically reviews existing documents related to the phenomenon under investigation (Dalglish et al., 2020). The technique focuses on the analysis of documents such as books, newspaper articles, academic articles, and organisational reports (Morgan, 2022). Iyamu (2022b) suggests that documents may exist in either a manual form or in an electronic format. Document analysis can be used as a standalone technique or combined with other techniques such as interviews (Berner-Rodoreda et al., 2020). In this study, document analysis was used as a standalone technique.

Qualitative data was collected using a set of criteria that included area of specialisation, publication time frame, and credible sources. The areas of focus were big data and architectural design, which are the core aspects of the study. A period of 10 years was considered, to gain an understanding of the historical background and meanings associated with the concepts (Iyamu, Nehemia-Maletzky & Shaanika, 2016). The documents helped to give a comprehensive and holistic view of big data architecture. Thus, only a small sample of the most appropriate and relevant literature could be gathered (Brereton et al., 2007; Glass, Ramesh & Vessey, 2004).

Materials published in Journal outlets, books, conference proceedings and the internet between 2013 and 2023 were considered. Academic databases such as Ebscohost, IEEE, Emerald, and Google Scholar were used as sources for the collection of data. This helps to ensure the credibility and reliability of the data (Nyikana & Iyamu, 2023). Google was used to search for strategic and operational documents in the areas of big data and architectural design. The analysis of the documents was two-fold; academic papers (peer-reviewed) and non-academic papers (white papers and green papers). As shown in Table 3. 2, a total of 201 papers were collected. In Chapter 4, comprehensive lists of both peer-reviewed and non-peer-reviewed are provided.

Table 3.2: Collected papers

Paper	**Peer-reviewed**	**Non-peer-reviewed**	**Total**
Big data	86	23	109
Big Data Architecture	39	10	49
Enterprise	33	10	43
Total	158	43	201

3.6 Data analysis

Data analysis is the process undertaken to transform the collected data into meaningful and valuable information (Taherdoost, 2022). During the process, themes, relationships, links, and key ideas are identified from the data (Mohajan, 2018), from which findings are reached, concisely and logically. According to Dufour and Richard (2019), in a qualitative study, data analysis is a complex stage of the research process because there are no universal methods or procedures to conduct the data analysis; it often lies on subjectivism.

The Activity Theory (AT) was employed to provide a frame and guide the data analysis in this study. AT is introduced and comprehensively discussed in Chapters 1 and 2, respectively. The data was coded and grouped into themes guided by the six components of AT, which are subjects, objects, tools, rules, community and the division of labour, as shown in Figure 2.2. The data analysis was conducted to achieve the objectives, as presented in Chapter 1 section 1.4. and revisited here:

(i) To examine how big data are generated, stored, governed, and used in enterprises.

(ii) To examine and understand the factors that influence the design of big data architecture for enterprises.

(iii) To understand the architectural components (technical and non-technical factors) that suit big data in the context of enterprises.

AT was used as a lens to guide the data analysis. The analysis focused on three main areas.

i. To gain an understanding of the activities that are involved in the use and management of big data in the organisation. This includes the procedures and regulations that are applied during the activities.
ii. To understand and examine the relationships and interactions that happen between the actors involved in the activities. This includes how and why roles and responsibilities are assigned and executed in the activities of big data in the organisation.
iii. The focus was on the governance of the big data activities. This helped to gain a deeper understanding of how IT solutions are used or can be used at various steps, in trying to maintain uniformity, reduce complexity, and enable flexibility in the activities of big data in the organisation.

The data analysis is presented in Chapter 5 of this thesis.

3.7 Ethical Consideration

Ethics are the values and principles that govern people to reduce risk when engaging in activities that involve humans and animals (Pastzor, 2015). This study adhered to the Cape Peninsula University of Technology's (CPUT's) code of ethics. The student followed the ethical clearance process set by CPUT, for research. The ethical clearance was obtained from the Faculty of Informatics and Design Research Ethics Committee. Some of the requirements included two mandatory documents that the student (researcher) needed to submit: (i) the research proposal approved by the Department in which the research is registered; and (ii) a completed ethics application form.

The university granted permission to conduct and collect the data. Thus, the ethics were applied as follows:

i. This research did not require the participation of individuals. Therefore, no personal information relating to any individual was used.
ii. The materials that were used in the study are in the public domain.
iii. Animals were not involved in the study. Therefore, no possible harm was caused to either the researcher or to animals.
iv. Data was collected from credible sources, which are listed in the reference section to provide transparency.
v. All materials including third-party materials used for this study were appropriately referenced, to avoid academic irregularity.

3.8 Summary

In this chapter, the research methodology undertaken in this study was comprehensively explained and justified. The inductive reasoning approach was employed, which influenced the research method and design that was selected. From an interpretive viewpoint, the qualitative method and case study research design were followed. The data was collected using a document analysis technique. The activity theory was used to guide the data analysis. The ethics that guided this research were also explained. The next Chapter presents the overview of the organisation used in this study.

CHAPTER FOUR
FIELD WORK

4.1 Introduction

This chapter covers the fieldwork that was carried out. Primarily, it focuses on the study's data collection. This includes the types of data that were collected, where and how the data was collected. As discussed in Chapter 3, document analysis is the technique selected for the data collection in this study. In the same chapter, the justifications for selecting the technique are presented. The data collection was guided by the objectives of the study, as presented in Chapter 1, and revisited in Chapter 3 of this thesis.

This chapter is organised into seven main sections. The first section introduces the chapter. In the second section, an overview of the chapter is presented. The section is followed by data classification, which provides a detailed explanation of how the data are categorised based on the focus areas. The fourth section provides an expansive explanation of the research questions. The fifth section presents comprehensive information, which includes the type, source, and a concise summary of the material (data) that was collected. Finally, a summary of the chapter is presented in the sixth section.

4.2 Overview

The document analysis technique is discussed in detail in Chapter 3. Hence, only an overview of the technique is presented in this chapter. The overview provides a further explanation, to awaken our intention in the context of this study. Document analysis is a technique that focuses on reviewing and examining manual and electronic documents to find meaning and gain knowledge (Bowen, 2009). The technique is used to dig into qualitative data, to extract rich content information (Chitsa & Iyamu, 2020). Karppinen and Moe (2012) state that the objectives of the study influence the information that is extracted from the documents being analysed. The documents can include books, newspaper articles, academic research articles, and organisational reports (Morgan, 2022). The document analysis helps to provide details of the circumstances and the origin of the phenomenon that is under investigation (Armstrong, 2020).

In addition, the technique provides insights into sensitive information that could be difficult to access using other data collection techniques (Sherif, 2018). The document analysis technique can be used to track and trace the development of a particular phenomenon over time (Dalglish et al., 2020). Morgan (2022) states that when using document analysis, it is important to understand the credibility and authenticity of the documents being analysed and their relevance to the phenomenon under investigation. In the context of this study, academic material (research peer-reviewed articles and books) and non-academic material (white and green papers) were analysed.

In this study, the document analysis technique is used in three sequential steps as discussed in the sections that follow. In the first step, the technique is used to guide the collection of data, which entails scrutinisation of the materials. Based on this, 201 documents were collected as discussed in section 4.5. The data that were collected were classified in the second step, as presented in Tables 4.1 and 4.2. This helps code the data, accordingly. Data classification is explained in section 4.3.

4.3 Data Classification

Data classification refers to the categorisation and grouping of objects based on their values (Tembusai et al., 2021). This means that data can be grouped based on content, focus, purpose, and relevance (Morgan, 2022). However, common attributes are sometimes used to classify data by comparing the similarities and differences (Khanna et al., 2021). Thus, structure and differentiation become important, which classification brings to the fore. Data classification provides a structure and organised information that can be used to gain a better understanding of complex concepts (Rizk et al., 2018). In this study, it helps to organise the materials into perspectives, in coding the data for analysis purposes. According to Bradford et al. (2022), data classification provides visibility, accessibility, security, and protection to the data.

In addition, locating and retrieving the data becomes easy when it is classified (Khanna et al., 2021). Mphahlele and Iyamu (2010) state that when the data is categorised and defined, it is intended to help an organisation perform better in its processes and activities. Through data classification, the use of big data in enterprises becomes effective and efficient bringing value to the enterprises (Ştefan, 2012). In this study, data were collected and classified into three main categories, which are big data, enterprise as an organisation, and the architecture of big data. The classification is based on the core aspects of the study. Primarily, the classification helps to distinguish the data for analysis purposes. Based on the classification, a format was adopted for the coding of the documents that were gathered. The format constitutes the name of each category and document number (i.e. Name#). The remainder of this section explains the three categories.

4.3.1 Big data

Big data is defined by using its characteristics; volume, velocity, variety, veracity, and value (Garoufallou & Gaitanou, 2021; Wang et al., 2020). Big data contains structured, semi-structured and unstructured datasets (Nyikana & Iyamu, 2022). Some of the attributes of big data include audio, video, images, and text (Sandhu, 2021; Pratsri & Nilsook, 2020). Big data is heterogeneous by nature, making it to be a complex concept that is not easily understood. Data was collected separately for this section to provide a comprehensive investigation of the concept. Also, separating the data collection helped to gain distinct and specific knowledge about big data. This helped to avoid confusion and overlapping definitions of concepts as stated by Nyikana and Iyamu (2023) that characteristics of big data can be easily confused with those of small data when the classification is not done.

The documents (data) collected were in the areas of business and IT activities covering processes, management, governance, operations, and strategy. Collecting data separately contributes towards a focused analysis because of the distinctive approach employed. The documents were coded as BDTDoc01 . . . BDTDocn+1 for peer-reviewed material and NPBDTDoc01 . . . NPBDTDocn+1 for non-peer-reviewed material.

4.3.2 An Enterprise

An enterprise is a legal entity formed to achieve specific objectives (Brée & Karger, 2022). It is often housed in a physical building. Usually, an enterprise is enabled and supported using various means and tools such as business processes, information technology (IT) solutions, and equipment (Striy et al., 2019). Documents were collected from different enterprises. Enterprises are in different sectors with various focus or objectives. It was therefore imperative to understand their focus and group them. This was critical to differentiate and to avoid over-generalisation.

Also, the documents collected from the enterprises were separated. Separation of documents collected from the enterprises was necessary to provide a clear understanding of the contribution of each element towards achieving the goals and objectives of the organisation. Additionally, the separation of data was used to contrast the use of data by different organisations in different fields. This helped to provide the context and relevance of the documents to the study. The grouping was done by establishing the benefits and challenges experienced by organisations when using data. Also, this helped to gain a comprehensive understanding of how some organisations use data to improve their business operations and drive decision-making towards value creation.

The data (documents) gathered from enterprises' perspectives were related to both business and IT processes for big data operations and strategic intent. From each enterprise, the documents focus on processes, engineering, and big data management. The documents were coded as ENTDoc01 . . . ENTDocn+1 for peer-reviewed material and NPENTDoc01 . . . NPENTDocn+1 for non-peer-reviewed material.

4.3.3 Big data architecture

Architecture is the plan and design of a system, relationship and interaction of the components to each other and the environment (Tschoppe & Drews, 2022). The principles and interaction of the components help to guide the design and governance of a system (Iyamu, 2022a). The data is separated in this section to explore the technical and non-technical aspects of big data architecture distinctly. Understanding these aspects is crucial for organisations as it can contribute to the architectural design of big data. Another reason for separating the data collection for the big data architecture section was to provide a structure to the data. The structure helped with the flow and to create the logic of the data collected.

As discussed above, architecture constitutes design and governance (Iyamu, 2022a) including principles and standards (Saddad et al., 2020). Thus, data relating to or focusing on governance, principles, standardisation, and policies of big data were collected and classified as Architecture documents. In alignment with other categories, the documents were coded as ARCDoc01 . . . ARCDocn+1 for peer-reviewed material and NPARCDocn . . . NPARCDocn+1 for non-peer-reviewed material.

4.4 Research questions

The study focuses on answering three research questions. This section explains the questions and what they seek to achieve. The research questions are revisited in this section to ensure that the data collected aligns with and answers them. Yin (2009) states that linking the research questions with data collection provides a chain of evidence to your findings.

RQ01 – How is big data generated, stored, governed, and used in enterprises?

The question focuses on understanding the life cycle of big data in enterprises. It seeks to provide insights into how enterprises generate big data from various sources. Additionally, it helps to understand how this data is stored and managed by the enterprises and to establish how the enterprises use the big data to achieve the business objectives. Zhang et al. (2022) highlight the use of big data to enhance the performance of organisations and increase sustainability. This question also establishes if there are any policies that the enterprises implement to govern the big data in their environments.

RQ02 – What are the factors that can influence the design of big data architecture?

The question seeks to understand the influence of big data characteristics, such as volume, variety, velocity, veracity, and value on the design of big data architecture. Additionally, the question explores other factors such as the business objective, skills, and infrastructure and their role and contribution to big data architectural design. Ylinen and Pekkola (2020) highlight that skills influence the design of big data architecture. This is in line with Dezi et al.'s (2018) view that different skills and infrastructure influence the architecture of big data.

RQ03 – What are the architectural components (technical and non-technical factors) that suit big data in the context of enterprises?

The question focuses on understanding the key components of big data architecture which consists of technical and non-technical components. Exploring these components provides a holistic approach to the suitable components required to design the architecture of big data that aligns with the organisational needs. Kalipe and Behera (2019) state that the components of big data architecture are defined based on technological needs and organisational requirements.

4.5 Data Collection

Based on the research questions and the classification discussed above, data were collected, using existing material from various sources. The existing material consisted of both peer-reviewed and non-peer-reviewed material.

The peer-reviewed material refers to research studies that have been validated by a group of scholars to ensure acceptable quality (Nicholas et al., 2015). According to Wehn and Almomani (2019), peer-reviewed material includes conference papers, journal articles, book chapters and books. The peer-reviewed articles provide credibility, integrity, and value to research studies (Park & Kressel, 2018). Nicholas et al. (2015) state that peer-reviewed material provides trustworthy information. Vahedi et al. (2023) argue that considering only peer-reviewed material may result in omitting important information. Hence, the non-peer-reviewed material was also considered in the study.

Non-peer-reviewed material, also known as grey literature, helps to provide ideas and broader information on concepts (Nicholas et al., 2015). Non-peer-reviewed material includes data collected from sources such as interviews (Poniatowski et al., 2021), general websites, blogs, and tweets (Nicholas et al., 2015), organisational reports and media releases (Cansdale et al., 2022).

4.5.1 Peer-reviewed articles and books (Academic material)

A total number of 158 documents were gathered for peer-reviewed material. The material included 86 documents for big data, 38 for big data architecture and 33 for enterprise. Table 4.1. presents the sample of 26 documents that were collected.

Table 4.1. Selected Peer-Reviewed Material

Data classification	Code name	Type of paper	Source
Big data	BDTDoc38	Conference	Nyikana & Iyamu (2022). Guide for selecting big data analytics tools in an organisation. Hawaii: HICSS, 5451–5461.
	BDTDoc01	Journal	Saggi & Jain (2018). A survey towards an integration of big data analytics to big insights for value-creation. Information Processing & Management, 54(5),758–790.
	BDTDoc02	Journal	Jones (2019). What we talk about when we talk about (big) data. The Journal of Strategic Information Systems, 28(1), 3–16.
	BDTDoc07	Journal	Garoufallou & Gaitanou (2021). Big data: opportunities and challenges in libraries, a systematic literature review. College & Research Libraries, 82(3),410–435.
	BDTDoc39	Book	Iyamu (2023). Advancing Big Data Analytics for Healthcare Service Delivery. Taylor & Francis.
	BDTDoc09	Journal	Sandhu (2021). Big data with cloud computing: Discussions and challenges. Big Data Mining and Analytics, 5(1),32– 40.
	BDTDoc11	Journal	Dezi, Santoro, Gabteni, & Pellicelli (2018). The role of big data in shaping ambidextrous business process management: Case studies from the service industry. Business Process Management Journal, 24(5),1163–1175.
	BDTDoc12	Conference	Singh, Singh & Verma (2022). March. The anatomy of big data: concepts, principles and challenges. India: IEEE, 986-990.
	BDTDoc22	Conference	Al-Sai & Abdullah. (2019) Big data impacts and challenges: a review. Jordan: IEEE, 150–155.
	BDTDoc49	Conference	Tekaya,Feki, Tekaya & Masri (2020). Recent applications of big data in finance. Tunisia: DTUC, 1–6.
Enterprises	ENTDoc05	Book	Musukutwa (2022). Developing an Enterprise Architecture. In SAP Enterprise Architecture: A Blueprint for Executing Digital Transformation. CA: Apress, 51–92.
	ENTDoc06	Journal	Sani, Sharip, Othman & Hussain (2018). Relationship between types of organization with the quality of soft-scape construction work in Malaysia. Asian Journal of Quality of Life, 3(12),137–146.
	ENTDoc07	Journal	Lang, Pott & Shinozaki (2021). Organisations and the production of migration and In/exclusion. Comparative migration studies, 9(60),1–16.
	ENTDoc08	Journal	Brée & Karger (2022). Challenges in enterprise architecture management: overview and future research. Journal of Governance and Regulation, 11(2), 355–367.
	ENTDoc12	Journal	Svensson & Poveda-Guillen (2020). What is Data and What Can it be Used For? Key Questions in the

			Age of Burgeoning Data-essentialism. Journal of Digital Social Research, 2(3):65–83.
	ENTDoc13	Journal	Prasetyo, Aziz, Faqih, Primadi, Herdianto & Febriantoro (2019). A review: evolution of big data in a developing country. Bulletin of Social Informatics Theory and Application, 3(1),30–37.
	ENTDoc23	Conference	Kim, Zimmermann, DeLine & Begel (2016). The emerging role of data scientists on software development teams. Texas: ACM, 96–107.
Big data architecture	ARCDoc01	Journal	Blazquez & Domenech (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting and Social Change, 130(2018),99–113.
	ARCDoc02	Journal	Oussous, Benjelloun, Lahcen & Belfkih (2018). Big Data technologies: A survey. Journal of King Saud University-Computer and Information Sciences, 30(4),431–448.
	ARCDoc21	Journal	Saddad, El-Bastawissy, Mokhtar & Hazman (2020). Lake data warehouse architecture for big data solutions. International Journal of Advanced Computer Science and Applications, 11(8), 417–424.
	ARCDoc25	Conference	Boumlik & Bahaj (2018). Big data and IOT: A prime opportunity for the banking industry. In Advanced Information Technology, Services and Systems: Tangier: Springer International Publishing, 396 – 407.
	ARCDoc24	Journal	Ruiz, Gómez-Romero, Fernandez-Basso & Martin-Bautista (2021). Big data architecture for building energy management systems. Transactions on Industrial Informatics, 18(9),5738–5747.
	ARCDoc17	Journal	Kalipe & Behera (2019). Big Data Architectures: A detailed and application-oriented review. International Journal of Innovative Technology and Exploring. Engineering, 8, 2182–2190.
	ARCDoc15	Journal	Pratsri & Nilsook (2020). Design on Big Data Platform-Based in Higher Education Institute. Higher Education Studies, 10(4),36–43.
	ARCDoc26	Conference	Tschoppe & Drews (2022). Developing Digitalization Strategies for SMEs: A Lightweight Architecture-based Method. Hawaii: HICSS, 1–10.
	ARCDoc27	Book Chapter	Filaly, Berros, Badri, Mendil & EL Idrissi (2023). Security of Hadoop Framework in Big Data. In Farhaoui, Rocha, Brahmia & Bhushab (ed.). Artificial Intelligence and Smart Environment. Cham: Springer International Publishing, 709–715.

4.5.2 Non-peer-reviewed articles (Non-academic material)

A total of 43 documents were gathered, from which 10 focus on big data architecture, 23 for big data and 10 for enterprise. 18 are presented in Table 4.2. as a sample. The non-peer-reviewed documents are further categorised into strategic or operational.

Table 4.2. Selected Non-Peer Reviewed Material

Data classification	Code	Type of paper	Source	No. of pages	Type of document
Big data	NPBDTDoc22	White paper	Analytics Software & Solutions (2018). SAS®: A Comprehensive Approach to Big Data Governance, Data Management and Analytics. https://sas.com/en_us/whitepapers.html	10	Strategic
	NPBDTDoc19	Policy	African Union (2022). AU Data Policy Framework. www.au.int	84	Strategic
	NPBDTDoc18	Article	Deloitte (2018). Knowledge Management & Big Data. www.deloitte.com/in	28	Operational
	NPBDTDoc17	Report	McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity. www.mckinsey.com/mgi/publications/	156	Strategic
	NPBDTDoc16	Report	World Bank Group (2017). Big Data in Action for Government: Big Data Innovation in Public Services, Policy and Engagement. https://documents1.worldbank.org/curated/en/	20	Operational
	NPBDTDoc12	Policy	Deloitte (2016). Big data analytics for policy making. https://joinup.ec.europa.eu/sites/default/files	122	Strategic
Enterprises	NPENTDoc04	Report	World Economic Forum (2020). A New Paradigm for Business of Data A New. https://www3.weforum.org/docs/	24	Strategic
	NPENTDoc05	E-guide	TechTarget (2019). What is Data Management and Why is it Important? https://www.techtarget.com/searchdatamanagement/definition/data-management	21	Operational
	NPENTDoc06	Report	Universal Education at Brookings (2018). Toward Data-Driven Education System. https://www.brookings.edu/wp-content/uploads/	78	Operational
	NPENTDoc07	Report	Infosys Knowledge Institute (2019). Data-Driven Health Care. https://www.infosys.com/about/knowledge-institute/insights/documents/	6	Operational
	NPENTDoc08	Report	Ernst & Young (2019). Realising the value of health care data: a framework for the future. https://assets.ey.com/content/	36	Strategic
	NPARCDoc01	Whitepaper	Oracle (2016). An Enterprise Architect's Guide to Big Data. https://www.oracle.com/technetwork/topics/entarch/articles/	49	Strategic

Big data architecture	NPARCDoc02	Report	DataVersity (2017). Trends in Data Architecture. https://content.dataversity.net/	66	Strategic
	NPARCDoc03	White paper	Juniper Networks (2012). Introduction to Big Data: Infrastructure and Networking Considerations. https://www.juniper.net/content/dam/www/assets/white-papers/us/en/	11	Operational
	NPARCDoc04	White paper	DataIku (2020). Data Architecture Basics. https://www.dataiku.com/	31	Strategic
	NPARCDoc06	White paper	National Institute of standards and technology (2015). NIST Big Data Interoperability Framework: Volume 5, Security and Privacy. https://www.govinfo.gov/content/	53	Strategic
	NPARCDoc09	eGuide	TechTarget (2019). What is data architecture? A data management blueprint.	18	Operational
	NPARCDoc10	Strategy	Nasa (2021). Data Strategy https://www.nasa.gov/	25	Strategic

4.6 Summary

The document analysis technique used to collect the data was discussed in this chapter. The data collected was classified and categorised based on the focus areas of the study, which are big data, enterprise, and big data architecture. Peer-reviewed and non-peer-reviewed materials were collected from various sources and used as data for the study. The materials were coded whereby documents were assigned codenames based on classifications. The next chapter presents the data analysis and interpretation.

CHAPTER FIVE
DATA ANALYSIS AND INTERPRETATION

5.1 Introduction

This chapter presents the data analysis and interpretation of the findings, including the proposed architecture. The data analysis was guided by the study's aim, which is to design a big data architecture for enterprises, purposely to enhance business continuity and improve the efficiency and effectiveness of operations and services in the use of big data in an organisation. The aim and objectives of the study are presented in Chapter 1 and revisited in Chapter 3. The analysis was conducted using the hermeneutics approach, which was guided by Activity theory (AT) as a lens. The theory is introduced in Chapter 1 and extensively discussed in Chapter 3.

The chapter is organised into six sections. Section one is the introduction. It is followed by section two, which covers the overview of the chapter. This section describes the process that was taken to prepare the collected documents for analysis. The third section provides the data analysis using the activity theory. Section four presents the findings and interpretations. In the fifth section, big data architecture is proposed. The sixth section presents big data architecture for business continuity. The last and seventh section is the summary of the chapter.

5.2 Overview

In Chapter 3, the approach employed for data collection is discussed. Data was collected to achieve the objectives of the study. The fieldwork including a sample of the data is presented in Chapter 4. AT as a lens guides the data analysis as presented in this chapter. Figure 5.1. below presents the activity model which consists of six components, subjects, objects, tools (instruments), rules, community, and division of labour (Park et al., 2013).

Due to various activities involved in the design of architecture, some of which are highlighted above, the activity theory (AT) is used to underpin the study. Activity theory is a socio-technical theory that has been adopted in IS studies in the last three decades (Iyamu, 2022a). The primary concern of the theory is the development of social activities (Shaanika & Iyamu, 2015). AT focuses on understanding the interactions and relationships that occur as activities are performed by humans (Iyamu & Shaanika, 2019a). Hence, in Dennehy and Conboy's (2017) explanation, AT is a framework used to understand complex human activities within a social system. Nehemia et al. (2018) described AT as a theory of consciousness. That is because the activities performed are consciously planned.

As shown in Figure 5. 1, the AT model consists of six components, subjects, objects, tools, rules, community, and the division of labour (Park et al., 2013). The components are interconnected and interrelated, as indicated by the arrows in the figure. The interconnections and relationship of the components help to understand the overall activities of the system (Nehemia et al., 2018). Also, as expressed by AT, activities are not static; they constantly evolve due to the changes in the environment (Engeström et al., 2016).

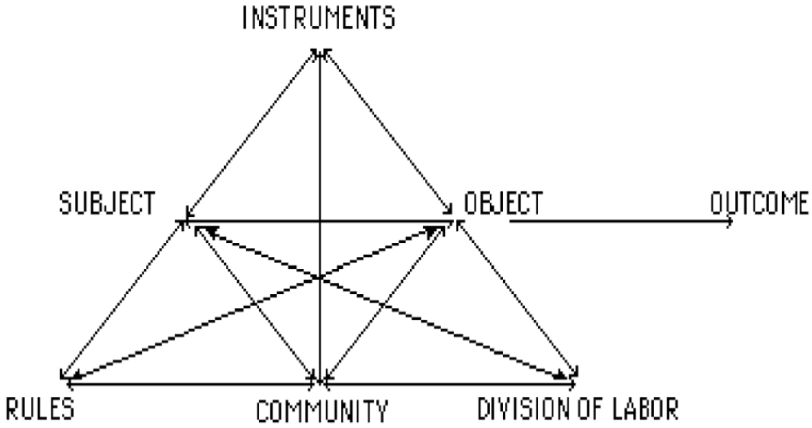

Figure 5.1. Activity model (Engeström et al., 2016)

A format is formulated for referencing the data. An example of the format is as follows: BDTDoc01; Pg#: Ln#. This means the first of big data documents, the page number, and line number. As organised in Chapter 4, the documents (used as data) are categorised into three groups, big data, enterprise, and architecture; the core aspects of the study. The categories are further divided into two, peer-reviewed and non-peer-reviewed. Each of the groups is assigned a codename, BDTDoc01 . . . BDTDocn+1 for big data; ENTDoc01 . . . ENTDocn+1 for enterprise; and ARCDoc01 . . . ARCDocn+1 for architecture. The same format (separated with np in front of the codename, i.e., NPBDTDoc01) is applied for the non-peer-reviewed materials.

5.3 Data analysis and discussion

Qualitative materials (data) were gathered to answer the research problem (Patel & Patel, 2019; Kapur, 2018). This was systematically conducted to ensure that the most appropriate materials were referenced (Iyamu, 2022a; Kothari, 2020; Mukharjee, 2019) as explained in Chapters 3 and 4. The AT model as shown in Figure 5.1 was employed by using the components to guide the analysis. This helps in three ways, in achieving the objective of the study. Firstly, it assisted in gaining a better understanding of how big data are stored and governed in enterprises. Secondly, it fortifies the fathoming of insights and evidence provided in the materials examined, to gain and understand the factors that influence the design of big data architecture for enterprises. Third, it helps to comprehend the relationships between architectural components (technical and non-technical factors) that suit big data in the context of enterprises.

AT is employed as a lens to provide a frame for the use of the hermeneutic approach in the analysis. This helps to gain a fathomed and in-depth view in proposing the architectural design for big data. The analysis is presented following the six components, tools, subject, rules, community, division of labour, and object of AT.

Activity theory: Tools

In AT, tools refer to artefacts used in an activity to transform an object into an outcome (Sannino & Engeström, 2018). Tools differ depending on the objective and the context of the study. Tools may include machines, instruments, signs, procedures, and laws (Nehemia et al., 2018). There are different types of tools when it comes to big data architecture. This includes technical artefacts (software and hardware) and non-technical artefacts (language) that need to be considered when designing the architecture of big data. Scalable storage and processing play a role in the design of big data architecture. Tools can be used in isolation while some may be integrated with others. For instance, HDFS and MapReduce are used in parallel to process, store, and retrieve large volumes of data (Boumlik & Bahaj, 2018). Additionally, when designing an architecture for big data, some strategies and practices need to be followed (Oussous et al., 2018). It is stated as follows in one of the materials used for this study:

> "This scheme was put about cloud computing, whose potential and benefits for storing huge amounts of data and performing powerful calculus are positioning it as a desirable technology to be included in the design of a Big Data architecture" [ARCDoc01; Pg100: Ln11-14].

An understanding of the tools guides an approach to how to apply them towards improving the manageability and governance of big data in an environment (Rao et al., 2019). In the context of this study, the primary architectural tools are big data, rules, structure, and skills. Mikalef et al. (2019) argued that skills are critical for organisations to gain value from big data. According to Jin et al. (2022), deriving rules in the era of big data is of fundamental importance. Pesqueira et al. (2020) revealed that many organisations struggle with structures and skills in storing, processing, and analysing information associated with big data. The skills and structures within an organisation influence decision-making (Kamble et al., 2021). Thus, there exists a default relationship between the actors and the architectural entities such as structures and skills, which forms the basis for essential interactions.

> "Without appropriate organisational structures and governance frameworks in place, it is impossible to collect and analyse data across an enterprise and deliver insights to where they are most needed" [BDTDoc51; Pg417: Ln25-27].

When incorrect tools such as storage and processing software are selected, challenges are sometimes encountered in the areas of performance, reliability, flexibility, and limited features (Chen et al., 2016). Also, selecting an incorrect tool potentially results in an organisation not gaining value from big data technology (Nyikana & Iyamu, 2022). Hence, Iyamu (2022a) argues that the selection of IT solutions should be guided by the architecture. Additionally, there are costs associated with tools, from purchase to support and maintenance perspectives. Thus, how and why tools are selected, used, and managed becomes critical. This helps to avoid prohibitive circumstances that affect the effective and efficient use of the tools.

The tools influence how IT solutions (such as big data) are implemented, managed, and governed (Mutasa & Iyamu, 2023). Interaction is defined by the relationship, and it is instrumental to how an activity is influenced. Interaction with structure produces and reproduces facility to allocate resources (Iyamu, 2010). Wang et al. (2022) explained how interaction with structure facilitates learning about the attributes of information. Thus, architecture cannot be designed without a good understanding of the relationships and interactions that exist between the actors on the one hand, and on the other hand, between the actors and big data.

Activity theory: Subject

Subject refers to a human being or a collective of people involved in an activity (Shaanika & Iyamu, 2015). There are different subjects involved in the activity of designing big data architecture. These subjects need to have the right skills and knowledge that are required to achieve the objective. Boumlik and Bahaj (2018) highlight that developers need to have language query skills to extract and present the correct data that is valuable to an organisation. The skills of the subjects need to align with their roles to perform the big data-related tasks assigned to them (Mohammad et al., 2014). In one of the materials, it is revealed as follows:

> "Organizations need to continuously plan and manage a trained workforce that can handle its Big Data technologies, and as such, this capability too can be considered a critical success factor for sustainable implementation of Big Data" [ARCDoc08; Pg5: Ln6-8].

Subjects (actors) involved in gathering, storing, using, and managing big data in an organisation are specialists with diverse skills. In AT, the subject, an actor engages in various activities within an organisation (Kelly, 2018). This can include developing and integrating IT solutions such as the design and implementation of an architecture for big data. In addition, this can be an individual or a group of actors who can either be technical or non-technical (Mutasa & Iyamu, 2023). Consequently, an actor does not act alone but in collaboration with other colleagues. This makes the allocation of tasks crucial. In doing so, the relationship between the actors must first be established and defined.

Organisations that do not have subjects with the right skills in their environment often struggle to sustain big data technologies such as big data architecture (Mustapha, 2020). This is primarily because the creation and management of the architecture require special or specific types of skills. In an attempt to substitute the specific architecture skill with other types of skills, challenges arise. Lack of the appropriate skills sometimes results in data that cannot be converted into strategic resources, for the operationalisation of the organisational goals and objectives (Dicuonzo et al., 2022). In one of the studies, the implication of lack of skill is stated as follows:

> "Many organisations have not been able to develop and implement architecture primarily because they do not have skilled personnel. What is even more challenging is the availability of the training facilities" [ENTDoc30; Pg52: Ln15-17].

The focal actor ensures appropriateness in the allocation because of its criticality. The appropriate allocation of tasks reduces the cost of operations and maximises the use of time and facilities (Yeon et al., 2022). In the design of architecture, different tasks and skills are required and must be aligned. Pesqueira at al. (2020) explained how skills affect the transformation of big data and ultimately, shape business insights and value creation in organisations. This advances the role of individuals and groups depending on the expertise, in the design and implementation of big data architecture.

Activity theory: Rules

Rules refer to control mechanisms, which can include policies, regulations, and legislations that guide and govern how subjects perform their activities in AT (Kelly, 2018). Additionally, Makovhololo et al. (2017) suggest that rules help to maintain order and control conflicts within an activity. The architecture of big data provides a layer that deals with the governance of the data throughout its lifecycle. The architecture of big data requires standards, laws, and regulation controls to collect, use, share, store and disseminate data (Pratsri & Nilsook, 2020).

> "The governance layer is in charge of applying policies and regulations to the whole data lifecycle, as well as managing the licenses related to the data sets." [ARCDoc01; Pg108: Ln7-9].

Big data is widely employed, yet there seem to be no universal rules dictated by architectural principles to govern how it is stored, retrieved, and managed in many organisations. Some organisations try to adopt the same sets of principles or rules for both small and big data (Todman et al., 2023; Faraway & Augustin, 2018). Consequently, this poses challenges for organisations. One of the challenges emanates because normal data technologies do not accommodate unstructured data sets such as images and videos (Saddad et al.,2020). As a result, some organisations lose out on the potential value and usefulness of their data. Hence the need for further exploring the use of big data is increasingly crucial. Nyikana and Iyamu (2023) argue that normal data and big data are not the same. The authors further posit that the differences between the two concepts include scope, volume, and heterogeneity. From the era of small data in the big data era perspective, a study emphasised the need for rules as follows:

> "In other words, there are strong rules relating to data standardization and compliance within the infrastructure" [BDTDoc52; Pg468: Ln40-42]

Data analysts and other specialists responsible for the governance and management of big data require certain protocols, processes, or regulations for operationalisation, such as gaining access to the data (Yaseen & Obaid, 2020). Also, rules are used to manage the interactions between actors during the process of allocating tasks, roles, and responsibilities. Wang et al. (2020) suggest that organisational structures such as departments are also used to restrict who accesses the resources.

Ahmed et al. (2017) defined normal data as data with structured data sets, low volumes, and constant velocities. Big data consists of huge volumes; structured, semi-structured and unstructured data sets; and high velocities (Oussous et al., 2018). The differences make it difficult to employ the same architectural design for both concepts. Jin et al. (2022) emphasised that the criticality of rules is on extracting the usefulness and providing output for new, and previously unseen value in data. He et al., (2023) suggest that rules enacted by an individual based on their skills are better and more manageable than machine-generated rules. This defines relationships and draws interactions between big data and the actors (users), through comprehensible data-driven decision-making and classification tasks (Jin et al., 2022).

Activity theory: Community

Iyamu and Shaanika (2019a) defined a community as a collective of individuals in a social system working towards the same goal. Big data architecture design involves a group of individuals such as data scientists, architects, analysts, software developers and business users (Chen et al., 2016). Forming teams of developers and data scientists helps to share strategies and ideas that help to increase the effectiveness and efficiency of the data (Kim et al., 2016). These individuals form communities based on their roles and skills which contribute towards achieving the desired goal. In addition, it becomes easy to achieve the goals when there is transparency and clear communication within the teams (Pau et al., 2022). Furthermore, communities are also responsible for making decisions regarding the appropriate big data architecture that aligns with the business goals.

> "Thus, collaboration across teams and workstreams is critical when designing data architecture to help reveal as many areas for improvement or threats as possible" [NPARCDoc04; Pg6: Ln26-29].

Over the years, organisations have been generating large volumes of data at an increasing rate. This has led to organisations realising the usefulness and value of their data for business continuity (Cockcroft & Russell, 2018). This draws interest from more stakeholders in an environment, which can either complicate or improve decision-making. Kamble et al.'s (2021) study reveals the latter, in which it is argued that the involvement of multiple groups (communities) in the decision-making process makes a difference in increasing the quality of output. Additionally, the community extends beyond an organisation through collaboration. For example, organisations from different fields such as health, education, finance, and commerce are embracing data as an asset and collaborating on projects (Hassan et al., 2020; Sandhu, 2021). These organisations use the data to gain insights (Garoufallou & Gaitanou, 2021) that can be used for better decision-making (Mustapha, 2022), drive business growth and stay competitive (Iyamu, 2018).

> "Companies use big data to achieve value creation in collaboration with stakeholders, which is manifested in strengthening connection and interaction, synergistically improving operational performance, and reducing operating costs through platform integration" [BDTDoc53; Pg 6: Ln 23-27]

However, the speed, volume, and variety of the data they generate make it difficult for them to collect, store, process and analyse the data using traditional technologies (Jaiswal et al., 2020). Hence, there is a need for the design of big data architecture that can handle complex data sets. It is, therefore, necessary to involve many persons of diverse skills, from different stakeholder groups in the decision-making process (Kamble et al., 2021).

When an organisation does not have teams with the right skills, it outsources the development of the architecture (Shaanika and Iyamu, 2018). This helps to avoid operational costs and waste of resources that may arise from an architecture that is not designed properly (Pääkkönen & Pakkala, 2020). Hence, the communities need to be guided by the rules when performing activities such as designing the architecture of big data (Dennehy & Conboy, 2017). Some of the rules mentioned by (Pau et al., 2022) are to integrate a security framework into the big data architecture to ensure data is secure and avoid vendor lock-in as it will prevent the future integration of services to the architecture.

Activity theory: Division of labour

Division of labour refers to the allocation of tasks and responsibilities among community members (Lioutas et al., 2019). Division of labour helps to promote accountability of the actions taken by the individuals within a community (Makovhololo et al., 2017). The design of big data architecture requires engagement from various stakeholders (actors) and an understanding of the functional and non-functional requirements of the big data architecture applications and their environment (Pau et al., 2022). The stakeholders perform distinct tasks that are interconnected to achieve a holistic design of big data architecture. For instance, the data scientist's role requires programming and decision-making skills (Uden et al., 2017).

> "On the technical side, data architects create data models themselves and supervise modelling work by others" [NPARCDoc09; Pg15: Ln18-19].

However, there is a shortage of skilled data scientists, which poses a challenge since there is a demand for expertise to manage and analyse big data effectively (Mustapha, 2022). Moreover, the security and privacy of the data are another concern, as organisations need to ensure that the data is protected from unauthorised people (Sandhu, 2021). Lastly, data storage is a challenge, whereby traditional data architectures lack the flexibility and scalability to store big data (Jaiswal et al., 2020). Consequently, the activities involved in big data are broader. This necessitates the inclusion of more individuals to ensure appropriateness in the division of the tasks.

> "Big Data architectures offer remarkable solutions to complex data issues but do not cover the complete flow of information that is required" [ARCDoc40; P1:20-21]

The tasks for the specialists are very specific and specialised. The developers are responsible for the integration of software applications and hardware components (Saggi & Jain, 2018). Some of the tasks of software engineers are to understand how the business and decisionmakers are going to use the data and therefore coordinate the analytical and storage tools needed (Chen et al., 2016). The use of the appropriate tools that align with the organisation's goals helps to deliver value for the business (Uden et al., 2017). Gokalp (2019) states that specific tasks require specific knowledge and experience because of their complexity.

Some of the tasks of architectural design for big data include an understanding of technical and nontechnical (business) requirements, decision-driven processes, and big data. The diversity of the tasks entails different types of skills. People with these skill sets are compulsorily required to work together to achieve a common goal. By implication, the personnel must conjure a working relationship and interact, to fortify their tasks. Skills are critical and useful (Mikalef et al., 2019) and He et al. (2023) emphasise the importance of collective actions in enabling and supporting big data use. Two other important factors arise in the process; task allocative and appropriateness of interaction between the actors involved. These are efforts to guide the outcome because the unintended can happen (Yeon et al., 2022).

Activity theory: Object

The object component of the AT model is the motive of conducting an activity, and it can be tangible or intangible (Iyamu, 2022b). Big data architecture provides organisations with a competitive advantage, improves performance, and generates value (Mustapha, 2022; Blazquez & Domenech, 2018). A well-defined big data architecture drives innovation and provides useful insights to the organisation (Avci et al, 2020). This includes defining the interactions and relationships between the elements of the architecture (Mohammad et al., 2014). A detailed analysis of the characteristics of the existing data architecture is required to help decide on the new architecture (Kalipe & Behera, 2019). Also, it helps to understand the limitations of the existing data architecture to accommodate the new requirements (Uden et al., 2017).

> "The right design of the big data architecture is a vital foundation for building an effective system to be used by the business on an everyday basis" [ARCDoc38; Pg460: Ln30-32].

Enterprises are increasingly using data to gain competitive advantage and maintain sustainability (Dezi et al., 2018). Additionally, enterprises use data to discover new insights, gain new ground, and uncover more opportunities (Dhaliwal & Shojania, 2018) to improve business decisions (Necsulescu, 2017). It is difficult to find a sector (or an enterprise) that does not employ data for its processes and activities, operationally or strategically. The education sector uses data to advance teaching and learning opportunities (Broos et al., 2017). At the same time, the health sector uses the data to monitor the health conditions of patients (Izonin et al., 2021) and to diagnose diseases (Mitani & Haneuse, 2020). The finance sector uses data to detect fraudulent activities (Aboud & Robinson, 2022) and to assess and manage risks (Cornwell et al., 2023).

Big data architecture design is a complex exercise that needs to be tailored based on the organisations' needs, drivers and available resources. Selecting an incorrect big data architecture can result in overlapping functionalities that can hinder the success of the organisation (Kalipe & Behera, 2019). Mustapha (2022) suggests that for an organisation to successfully implement big data architecture, it needs skilled employees who can work in the dynamic environment of big data.

5.4 Findings and interpretation

In designing the big data architecture, which is the objective of the study, the focus is on three main areas that are revealed in the analysis. First, it is to understand the activities involved in the use and management of big data in the organisation. This includes the procedure and regulation during the activities. Second, the relationships and interactions between the actors involved in the activities are examined. This is to mitigate risks in the activities involved in storing, accessing, and managing big data in an organisation. The third focus is on the governance of big data activities. This helps to gain a deeper understanding of how IT solutions are used or can be used at various steps, in trying to maintain uniformity, reduce complexity, and enable flexibility in the activities of big data in an organisation.

From the analysis presented in the preceding section, three factors, interactions, relationship, and allocation are fundamental to the architecture of big data. As shown in Figure 5.2, the factors are interrelated and influence one another in the activities of big data, such as data gathering, retrieval, security, governance, and use.

The factors were interpreted using subjective reasoning. This was done towards achieving the aim of the study, which is to design a big data architecture for enterprises, purposely to enhance business continuity and improve the efficiency and effectiveness of operations and services in an organisation. The architecture is designed and presented in the section that follows.

5.5 The big data architecture

The findings from the analysis, primarily, are interactions, relationships, and allocative factors, as shown in Figure 5.2, the architecture. The factors manifest and are informed by other attributes, as also shown in the architecture. The attributes are from both technical and non-technical standpoints. The factors are discussed in the remainder of this section. The discussion should be read with Figure 5.2 to better understand the big data architecture.

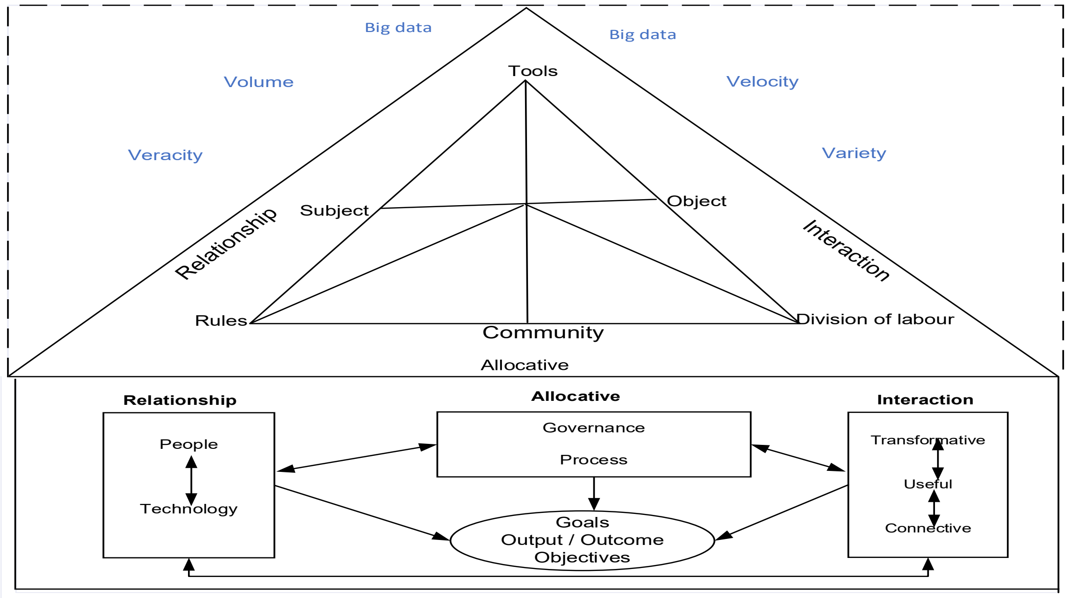

Figure 5.2: Big Data Architecture

5.5.1 Relationship between actors and technology solution

Relationships anchor how humans or non-humans are connected, or the state of being connected. The relationship between actors and big data influences architectural design (Li et al., 2022). The architectural design helps to resolve issues that are durable and long-lasting in the implementation of IT solutions such as big data (Georgiadis & Poels, 2021). The architectural design expresses the relationship between humans (such as business personnel, and IT architects) and technology in employing big data to support and enable organisations.

The relationship allows individuals to contribute expertise that produces relational value. In Giddens' (1984) explanation, systems are patterns of relations categorised into groupings within which relationships exist and interactions are conducted to produce and reproduce actions towards achieving specific goals. The architectural design draws on rules people employ to propel functioning value, which brings fort to an organisation. The relationship between the two entities, people and technology are enclaved in rules and compliance and regulate their interaction and functioning (Iyamu, 2022a). Employees fall back on such architectural value in times of interpretation and challenges. However, big data utilisation becomes a more complex process because relationships define human actions and interactions, Iyamu (2022c) argues.

5.5.2 Interaction towards an outcome

Interaction between actors and interaction with the rules and big data, together, ensure an outcome (architectural design). The interaction manifests into transformative, useful, and connective in the design of big data architecture. Iyamu (2022c) suggests that architecture is a synthetic approach to interacting with big data for organisational purposes. Interaction with rules and IT solutions (such as big data) is often geared towards the transformation of activities and business objectives, which is, however, not always straightforward. Habitually, in many environments, it requires interpretations, which are often subjective, with different meanings. Hence, collaboration among the actors remains a formidable option or solution to bridge the gap created by subjective understandings. It helps to integrate the contributions of individuals with diverse views and perspectives in addressing transformative initiatives.

In the process of architectural design, the actors must be reciprocally active in their interactions, to share requirements, ideas, and knowledge towards usefulness. Van Wessel et al. (2021) emphasised the need for the interaction between business and IT units, to fortify an organisation's better usefulness of IT solutions. Consequently, the rules used are reproduced, various skills are employed to deeply engage with content, and tasks are appropriately distributed. Li et al. (2022) contend that there must be new ways of interacting with big data to promote its usefulness. In the context of this study, interaction from connectivity can be divided into functional and non-functional. Functional interaction occurs between humans (actors), it entails the interpretation of context, and the exchange of ideas among the stakeholders. The non-functional interaction occurs between humans and non-humans, such as rules and IT solutions. Manogaran et al. (2022) argue that the challenges lie in human-computer interaction with big data. The significant role and the diverse nature of interaction make it cumbersome.

5.5.3 Allocative enabling support entities

The allocative factor is shaped by governance and processes, and it constitutes rules that define boundaries within which activities are conducted. Also, the rules do not enforce themselves. Compliance with the use and management of big data is a challenge in many organisations (Georgiadis & Poels, 2021). Allocative efficiency necessitates the distribution of tasks among the actors, in the design and implementation of big data architecture. Primarily, this navigates and fosters the interaction of aligned interests of both business and IT architects (Iyamu, 2022c).

Primarily, one of the challenges in many environments is the inappropriate distribution of tasks in the design, implementation and management of IT solutions including big data. As a result, challenges linger in the efficient use of big data for organisational purposes. Thus, governance becomes fundamentally important. Governance defines the standards, principles, and policies within which events and activities are performed. According to Iyamu (2022c), the allocative system is influenced by policy, and implementation of standards and principles, in the development and implementation of architecture in organisations. Allocative efficiency allows the gathering of holistic technical and business requirements, accessing more accurate information, and more inclusive decision-making, to guide the design of big data architecture.

5.6 Big data architecture for business continuity

The big data architecture is intended to enhance business continuity for an organisation. For the business continuity to happen, there are implications. Ghasemaghaei and Calic (2019) emphasised that the relationship with big data contains significant theoretical and practical implications. As tabulated in Table 5.1., the implications identified are Operationalisation, Innovation, and Integration. Subjective reasoning is applied in identifying the implications. The implications are viewed from both technical (IT unit) and non-technical (business unit) standpoints.

Table 5.1: Implication of the study

Attributes	IT unit	Business unit
Operationalisation	To operationalise the architectural design requires an understanding of the application, to protect and enhance big data in storing, retrieving, and management. Develop an operationalisation approach to enable and support the architectural design (Batyashe & Iyamu, 2020).	How to apply the architectural design to improve big data operations. Understanding the influencing factors to leverage and utilise big data to improve efficiency and performance (Calic & Ghasemaghaei, 2021).
Innovation	The architectural design needs to be measured to gain insights into its innovative value and service to the organisation. This requires metrics by the IT unit. According to Babu et al. (2021), BDA has implications in various stages of innovation, including scope, design, and implementation, to provide meaningful insight into efficiency and effectiveness.	It necessitates business units' understanding of how big data architecture can be used for business innovations, including reducing costs. This includes minimising the total cost of ownership (TCO) and promoting competitiveness (Iyamu, 2022a)
Integration	To integrate big data architectural design with enterprise architecture or any other existing architectures. Wang et al. (2018) suggest that it is critical to employ architectural design in addressing big data component functionalities. It creates opportunities and challenges for skills alignment (Iyamu, 2022a).	How to integrate the big data architectural design with other business objectives to increase efficiency and effectiveness in the organisation. Ethiraj and Posen (2014) explained from the business viewpoint, an implication of architectural design lies in interconnectivity with products.

An understanding of the implications of deploying big data is significant to improve organisational efficiency and performance. From the operationalisation and strategic perspectives, there are factors which influence the practice of big data architectural design. Thus, organisations must develop an operational approach to support architectural design. In the innovation component, the IT unit needs to develop metrics that can be used to measure the value of architectural design to the organisation (Acciarini et al., 2023). The business units need to understand how big data architecture can be used to reduce costs and promote business innovations (Sestino & De Mauro, 2022). Integration remains an iterative approach for the unification and coexistence of IT solutions and business artefacts to reduce complexity, increase effectiveness and efficiency, promote seamlessness of processes, and enable product interconnectivity (Qi, Xu & Rani, 2023; Dwivedi et al., 2022).

5.7 Summary

In this chapter, the components of the activity model of AT were used to guide data analysis. The analysis revealed the factors that influence the design of big data architecture. The findings were interactions, relationships, and allocative factors. The factors constitute technical and non-technical operations and processes involving big data. The architecture of big data was designed based on these factors. The combination of both technical and non-technical factors in the architecture makes it critical for business continuity. Also, theoretical, and practical implications towards business continuity were identified using subjective reasoning. The identified implications were operationalisation, innovation, and integration. The next Chapter concludes the study and offers recommendations for interested academics and enterprises.

CHAPTER SIX
CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

This chapter concludes the thesis and offers recommendations based on the study. The study aimed to design an architecture of big data for enterprises, purposely to enhance business continuity and improve the efficiency and effectiveness of operations and services in the use of big data in an organisation. Based on the aim, data was collected and analysed, and findings were reached. This brings the study to a conclusion with recommendations.

This chapter is divided into ten sections. The first section is the introduction. Sections two and three summarise the chapters and evaluate the research, respectively. The fourth section focuses on the summary of outcomes. The section explains how the objectives of the study were achieved. In the fifth section, the contribution of the study is discussed based on the practical, theoretical, and methodological contributions, in that sequence. Section six presents the benefits of the study, followed by the limitations of the study in section seven. In section eight, the recommendations of the study are presented. Sections nine and ten provide further study and summarise the chapter, respectively.

6.2 Summary of the chapters

This section presents the summary of the chapters covered in the study. Although each chapter was summarised, this is a revisit as part of the conclusion of the thesis. The section provides a structure and preview to help the readers navigate through the study.

Chapter One is the introduction of the thesis. The chapter lays the background of the problem and the rationale for conducting the study. The research questions are also presented. The literature review is introduced covering the concepts of big data, enterprise and architecture. The theory, activity theory (AT) that underpins the study, is also discussed. Furthermore, the chapter explains the research methodology adopted in the study. It includes the research method, research design, data collection technique, and data analysis. The significance and contribution of the study were briefly discussed. The last section of the chapter discussed the ethics that were put into consideration by the researcher when conducting the study.

Chapter Two presents a detailed literature review. The literature focused on enterprises, big data in organisations, and the architecture of big data. The chapter provides a comprehensive and detailed explanation of these concepts. The literature review helped to uncover the gaps that exist in the areas of big data architecture and how big data is currently used in many enterprises. Also, an in-depth understanding of what has been covered by other researchers was gained. An explanation of AT was provided. This includes how the theory relates to the phenomenon being studied.

In Chapter Three, the methodology that was followed to conduct the study was thoroughly discussed and the justifications for selecting each aspect of the methodology were provided. The chapter covered the research approach, research method, research design and data collection technique. The objectives of the study guided the selection of the aspects of this chapter, which make the approaches, methods, and techniques most appropriate. The chapter explains why and how the AT model was employed to guide the data analysis.

Chapter four presented the field work which covers the process undertaken to collect the data. Peer-reviewed and non-peer-reviewed materials (documents) were collected from various sources. The focus was academic databases. The documents were classified based on the focus areas of the study, which are big data, enterprises, and big data architecture. The research questions were revisited in this chapter to ensure their alignment with the data collected. The collected documents were coded for ease of access. A sample of the documents collected is provided in Table 4.1 for peer-reviewed and Table 4.2 for non-peer-reviewed materials.

The data analysis and the interpretation of the findings are presented and discussed in chapter five. The first section of the chapter provides an overview, which briefly explains the process followed to prepare the documents for analysis. Data were analysed guided by the components of AT. The findings were interpreted and a big data architecture is proposed in the chapter. Another important aspect of the chapter is the explanation of how the interpretation was conducted.

Finally, Chapter Six covers the conclusion and recommendations of the study. In this chapter, a summary of the chapters is provided. The research is evaluated based on Dane's (2010) six components of assessing qualitative research. The summary of the outcomes as well as the contributions of the research are also discussed. Also presented are the study's limitations and recommendations.

6.3 Evaluation of research

The evaluation of the study is done to assess its quality. According to Martensson (2016), the evaluation can be done by revisiting the objectives and research questions of the study. Dane (2010) suggests the use of six components for the evaluation of a study. The components are who? what? where? when? how? and why? Table 6.1. below uses the suggested components to evaluate the research.

Table 6.1. Research evaluation

Component	Evaluation of the study
What	**What** is concerned with the phenomenon under investigation and gaining knowledge about it. How to design architecture for enterprises was investigated, purposely to enhance business continuity and improve the efficiency and effectiveness of operations and services in the use of big data. In doing so, how big data are generated, stored, governed, and used in enterprises were examined; the factors that influence the design of big data architecture for enterprises were examined to gain a better understanding; and in the process, the architectural components (technical and non-technical factors) that suit big data in the context of enterprises were better understood. The focus areas for the study were big data, enterprise and architecture. The factors that influence the architectural design of big data were discovered and identified as relationships, interaction and allocative. The theoretical and practical implications towards enhancing business continuity were also identified as operationalisation, innovation, and integration.
Where	**Where** refers to the geographical location where the phenomenon was studied. The problem was identified in organisations in the Western Cape of South Africa. This means that no organisation seems to have designed big data architecture. As a result, no organisation could be used as a case because of lack of experience. Thus, existing materials were collected instead of employing the interview technique. Materials published in Journal outlets, books, conference proceedings and the internet were collected. Academic databases such as Ebscohost, IEEE, Emerald, and Google Scholar were used as sources of data collection.
Who	**Who** refers to the participants who are fit for the study. However, records of people, objects and organisations can sometimes be considered as participants. In this study, there were no human beings that participated. Instead, documents related to big data and architectural design were collected, which are the core aspects of the study. Peer-reviewed and non-peer-reviewed materials were considered. A set of criteria, which include area of specialisation, publication period, and credibility sources was used to select the most relevant documents. For example, a period of 10 years was considered enough to understand the historical background and meanings associated with the concepts. The process is explained in chapters 3 and 4.
When	**When** refers to the period the study was conducted. The study was conducted in 2023. This means that the study was problematised and the thesis completed in the same 2023. The data were collected between May 2023 and October 2023. The data were analysed in November 2023. In 2023, big data and architecture were high in the discourses in both business and academic domains.
How	This refers to **how** the research was conducted. The research methodology which consists of approaches, methods, and techniques was employed in conducting the study. The aim and objectives of the study are set as criteria for selecting the methodology. Based on the guidance, the most appropriate approaches, methods, and techniques were selected and applied in the study. The rationales for selecting each of the methodology components are detailed in Chapter 3. The inductive approach was employed in the study under the interpretive paradigm. A qualitative research method was adopted. The case study

Component	Evaluation of the study
	design was seen to be fit for the study because the study seeks to gain an in-depth understanding of the architecture of big data. Document analysis was adopted to collect the material. The Hermeneutics approach was used to analyse the data guided by the components of the AT. Findings were interpreted from which big data architecture was designed.
Why	**Why** is concerned with the rationale for conducting the study. Organisations do not have architectural designs that encompass the characteristics of big data that are rapidly increasing in their environments. Thus, a gap existed. Explicitly, the gap is explained in the research problem section of Chapter 1. This study was conducted to address the gap, by designing an architecture of big data that can be used to enhance business continuity and improve the efficiency and effectiveness of operations and services in the use of big data in an organisation. Also, at the time of this thesis, no study seems to have employed activity theory to examine architectural design for big data.

6.4 Summary of the outcomes

The study aimed to design a big data architecture for enterprises, purposely to enhance business continuity and improve efficiency and effectiveness of operations and services in the use of big data in an organisation. Three objectives were formulated from the aim. Below is an explanation of how the objectives were achieved.

6.4.1 To examine how big data are generated, stored, governed and used in the enterprises.

In achieving this objective, the question is: How are big data generated, stored, governed and used in enterprises? was posed. Data were collected from various sources, to answer the question using the document analysis technique. Relevant extracts from the data were combined and examined (analysed) using the AT model, which consists of six components (tools, subject, rules, community, division of labour, and object).

It was discovered that the use and management of big data is shaped by governance and processes. How the components of AT were applied are explained in Chapter 3 and an overview of it is presented in Chapter 5. The analysis revealed that governance can be used to define the standards, principles, and policies within which events and activities are performed when developing and implementing architectures in organisations. The governance of activities helps to maintain uniformity, reduce complexity, and enable flexibility in the activities such as how big data are generated, stored, governed and used in the enterprises.

6.4.2 To examine and understand the factors that influence the design of big data architecture for enterprises.

The question is, what are the factors that can influence the design of big data architecture? helps to achieve this objective. From the analysis, the factors that influence the design of big data architecture were revealed. As presented in Chapter 5, three factors were understood to influence the design of big data architecture. The factors are relationship, interactions and allocation. The relationship is between humans such as business personnel, and IT architects and technology.

The relationship between these entities is guided by rules they need to comply with during the design of big data architecture. Relationships define the activities and interactions that take place when utilising big data. Humans interact with rules and IT solutions such as big data to transform business activities and objectives. During the process of designing big data architecture, the interactions between humans allow them to share requirements, ideas, and knowledge, and allocate tasks. The allocation of tasks promotes the alignment of interests between business and IT architects.

6.4.3 To understand the architectural components (technical and non-technical factors) that suit big data in the context of enterprises.

It is revealed from the data analysis that the use of big data to enhance business continuity is influenced by theoretical and practical implications. The implications were viewed from the business and IT units. Operationalisation, innovation, and integration were identified as significant factors for improving organisational efficiency and performance. From the operationalisation perspective, organisations need to develop an operational approach to support the architecture of big data. In the innovation component, the IT units need to develop metrics that can be used to measure the value of big data architecture to the organisation. The business units need to understand how big data architecture can be used to reduce costs and promote business innovations. Integration ensures the unification of IT solutions and business artefacts, to reduce complexity, increase effectiveness and efficiency, promote seamlessness of processes, and enable product interconnectivity.

6.5 Contribution of the study

This section covers the contributions of the study which are threefold, practically, theoretically, and methodologically. The practical contribution is at the enterprise (business) level. While the theoretical and methodological are for the academic domain. Thus, the study contributes to the existing knowledge of big data, from both business and academic perspectives.

6.5.1 Practical contribution

The practical contribution of the study can be viewed from three perspectives. Firstly, the designed architecture can be used to guide the development of policies, standards, and principles in an enterprise. Based on the policies, standards, and principles, big data can be better stored, increase retrieval, and promote usability and manageability. Better management reduces complexity and improves effectiveness and efficiency in the use of big data for service delivery. Secondly, in practice, the study reveals the factors that influence the design of big data architecture, which can be used to develop a research stream. Thirdly, understanding these factors can be useful to business and technology solutions managers, data architects and data scientists to ensure effective and efficient big data architectures, towards achieving business continuity.

6.5.2 Theoretical contribution

Theoretically, the study contributes to both enterprises and academic domains. The architecture enables enterprises to align with the evolving needs and challenges faced by organisations when dealing with the characteristics of big data. Secondly, the study highlights the architectural components that influence the use and management of big data in an environment. From an academic domain standpoint, the architectural design forms part of the enterprise architecture research stream. In addition, the big data architecture provided in this study has not been tested, which makes it theoretical. This creates an opportunity for validation through further research studies. Researchers and students focusing on an understanding of the architecture of big data can benefit from the study. Very importantly it contributes to the body of knowledge in areas such as big data architecture and architectural design where literature is currently limited.

6.5.3 Methodological Contribution

The use of activity theory to underpin the study contributes methodologically. Firstly, at the time of this thesis, no study seemed to have employed activity theory to examine data, towards a design of big data architecture. The components of AT were used to understand the activities that relate to big data and architecture that take place in enterprises. The theory was used to navigate existing materials, which makes the study a good methodological contribution. Thus, it contributes to advancing the application of the theory, in IT studies.

6.6 Benefits of the study

The study focused on designing an architecture for big data that can be used by organisations to enhance business continuity and improve efficiency. It is intended to benefit businesses and academics that focus on big data architecture, or both. This includes businesses in any sector or industry that employ big data or intend to do so. Also, the benefit of the study can be realised at both operational and strategic levels. From an academic perspective, the study can be of benefit to both students and educators (researchers).

Organisations do not have big data architectures that can encompass and guide the rapidly increasing characteristics of big data in their environments. The architecture proposed in this study can benefit data and business architects in understanding the factors that influence the design of big data architecture. The study adds to the existing literature in the field of big data and architectural design including the advancement of the use of AT in IT research.

Additionally, this study can benefit many organisations on the African continent. In Africa, many organisations either struggle to develop an architecture for big data usage and management or have no resources to do so. The design proposed in this study can be used as a reference point to guide the development of big data architecture in any organisation.

6.7 Limitations of the study

There was no organisation with big data that could be used as a case, where data could have been collected. As a result, interviews could not be conducted. Instead, documents were collected from academic databases.

6.8 Recommendations

This section presents the study's recommendations. The recommendations focus on how the study can be put to practice by operationalising the proposed big data architecture. Three recommendations are reached, which are the development of a template, skill development, and teamwork, as discussed below:

i. Template development

The organisations interested in using the proposed architecture need to develop a template for each of its components. It serves as a starting point to describe the boundaries in applying the architecture. The template needs to be created within context, based on the requirements of the organisation. This is to ensure that the initiative to develop the big data architecture aligns with the organisational goals. In addition, the template can be used to identify the skills and infrastructure required to develop the architectural design.

ii. Skills development

Big data is an emerging technology that requires specialised skills. People need to have a good understanding of the tools (technology solutions and processes) that are required to generate, store, govern and use big data in their enterprises. Another important aspect is that people need to have a relationship with the tools that they use and know how to use them. Thus, it is critical for organisations to upskill their employees for these highlighted purposes. Also, architecture is a specialised area that requires specific skills.

iii. Teamwork

Collaboration and communication between humans (business personnel, and IT architects) is crucial when developing an architectural design. This helps to ensure that individuals contribute their expertise and share ideas towards achieving specific tasks. Additionally, collaboration between departments (business and IT units) is required to achieve the efficient and effective development of big data architecture.

6.9 Further study

The architecture proposed in the study has not been tested. Therefore, further studies can be conducted to evaluate the architecture. Also, for future research, the study can be expanded to address the development and implementation of big data architecture design.

6.10 Summary

The chapter presented the conclusion and recommendations of the study. The chapters in the thesis are summarised. The research was evaluated to show the quality and usefulness of the study. The summary of the outcomes was presented to ensure the objectives of the study were achieved. Also, in this chapter, the contributions and benefits of the study were discussed. Recommendations on how the organisations can use the proposed big data architecture were provided.

REFERENCES

Abdel-Basset, M., Ding, W., Mohamed, R. & Metawa, N. 2020. An integrated plithogenic MCDM approach for financial performance evaluation of manufacturing industries. *Risk management*, 22: 92–218.

Aboud, A. & Robinson, B. 2022. Fraudulent financial reporting and data analytics: an explanatory study from Ireland. *Accounting Research Journal*, 35(1):21–36.

Acciarini, C., Cappa, F., Boccardelli, P., & Oriani, R. 2023. How can organizations leverage big data to innovate their business models? A systematic literature review. *Technovation*, *123*: 102713.

Adekoya, O.D., Jimoh, I., Okorie, G. & Olajide, M. 2019. Significance of Employee Engagement and Individual Well-Being on Organisational Performance in Nigeria, *International Journal of Science and Management Studies*, 2(5): 35–47.

Ahmed, V., Tezel, A., Aziz, Z. & Sibley, M. 2017. The future of big data in facilities management: opportunities and challenges. *Facilities*, 35(13):725–745.

Aliee, F.S., Kashfi, H. & Farahani, B. 2019, May. The evolving enterprise architecture: A digital transformation perspective. In *Proceedings of the International Conference on Omni-Layer Intelligent Systems*, 5-7 May 2019. Greece: ACM, 179 – 183.

Algozzine, B. & Hancock, D. 2017. *Doing Case Study Research: A Practical Guide for Beginning Researchers*. 3rd ed. New York: Teachers College Press.

Alfonsius, G. 2021. What is business? *Universal Journal*, 6(2):1–9.

Al-Sai, Z.A. & Abdullah, R. 2019. Big data impacts and challenges: a review. In *2019 IEEE Jordan international joint conference on electrical engineering and information technology*, 09-11 April 2019.Jordan: IEEE, 150–155.

Avci, C., Tekinerdogan, B. & Athanasiadis, I.N. 2020. Software architectures for big data: a systematic literature review. *Big Data Analytics*, 5(1):1–53.

Babu, M.M., Rahman, M., Alam, A. & Dey, B.L. 2021. Exploring big data-driven innovation in the manufacturing sector: evidence from UK firms. *Annals of Operations Research*, 1–28.

Bahri, S., Zoghlami, N., Abed, M. & Tavares, J.M.R. 2018. Big data for healthcare: a survey. *IEEE access*, 7:7397–7408.

Bansal, B., Jenipher, V.N., Jain, R., Dilip, R., Kumbhkar, M., Pramanik, S., Roy, S. & Gupta, A. 2022. Big Data Architecture for Network Security. *Cyber Security and Network Security*,233–267.

Barham, H. 2017. Achieving competitive advantage through big data: A literature review. In *Portland international conference on management of engineering and technology*, 9 – 13 July 2017. USA: IEEE,1–7.

Barradas, A., Tejeda-Gil, A. & Cantón-Croda, R.M. 2022. Real-time big data architecture for processing cryptocurrency and social media data: A clustering approach based on k-means. *Algorithms*, 15(5): 1–11.

Baškarada, S. & Koronios, A. 2018. A philosophical discussion of qualitative, quantitative, and mixed methods research in social science. *Qualitative Research Journal*, 18(1):300–315.

Batyashe, N. R., & Iyamu, T. 2020. Operationalisation of the Information Technology strategy in an organisation. *Journal of Contemporary Management*, *17*(2):198–224.

Beese, J., Haki, M.K., Aier, S. & Winter, R. 2019. Simulation-based research in information systems. *Business & Information Systems Engineering*, 61(4):503–521.

Behfar, K. & Okhuysen, G.A. 2018. Perspective—Discovery within validation logic: Deliberately surfacing, complementing, and substituting abductive reasoning in hypothetico-deductive inquiry. *Organization Science*, 29(2):323–340

Belov, V. & Nikulchev, E. 2021. Analysis of Big Data Storage Tools for Data Lakes based on Apache Hadoop Platform. *International Journal of Advanced Computer Science and Applications*, 12(8):551–557.

Benhlima, L. 2018. Big data management for healthcare systems: architecture, requirements, and implementation. *Advances in bioinformatics*, 2018: 1–11.

Berner-Rodoreda, A., Bärnighausen, T., Kennedy, C., Brinkmann, S., Sarker, M., Wikler, D., Eyal, N. & McMahon, S.A. 2020. From doxastic to epistemic: a typology and critique of qualitative interview styles. *Qualitative inquiry*, 26(3–4): 291–305.

Bhattacherjee, A. 2012. *Social Science Research: Principles, Methods, and Practices*. 2nd ed. USA.

Bihu, R. 2020. Using unstructured interviews in educational and social science research: The process, opportunity and difficulty. *Global Scientific Journals, 8*(10): 712–721.

Blazquez, D. & Domenech, J. 2018. Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, 130(2018): 99–113.

Bloomfield, J. & Fisher, M.J. 2019. Quantitative research design. *Journal of the Australasian Rehabilitation Nurses Association*, *22*(2):27–30.

Bradford, M., Taylor, E.Z. & Seymore, M. 2022. A view from the CISO: insights from the data classification process. *Journal of Information Systems*, *36*(1), pp.201– 218.

Brannen, J. 2017. *Mixing methods: Qualitative and Quantitative Research*. 2nd ed. London: Routledge.

Brée, T. & Karger, E. 2022. Challenges in enterprise architecture management: overview and future research. *Journal of Governance and Regulation*, 11(2): 355–367.

Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., & Khalil, M. 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4):571–583.

Broos, T., Verbert, K., Langie, G., Van Soom, C. & De Laet, T. 2017. Small data as a conversation starter for learning analytics: Exam results dashboard for first-year students in higher education. *Journal of Research in Innovative Teaching & Learning*, 10(2):94–106.

Boumlik, A. & Bahaj, M. 2018. Big data and IOT: A prime opportunity for banking industry. In Advanced Information Technology, Services and Systems: In *Proceedings of the International Conference on Advanced Information Technology, Services and Systems,* April 14–15 April 2017. Tangier: Springer International Publishing, 396 – 407).

Bowen, G.A. 2009. Document analysis as a qualitative research method. *Qualitative research journal*, *9*(2): 27– 40.

Busetto, L., Wick, W. & Gumbinger, C. 2020. How to use and assess qualitative research methods. *Neurological Research and practice*, 2(14):1–10.

Calic, G., & Ghasemaghaei, M. 2021. Big data for social benefits: Innovation as a mediator of the relationship between big data and corporate social performance. *Journal of Business Research*, *131*:391–401.

Cansdale, L.G., Kelly, G., Khashan, A., Kachale, F., Lissauer, D., Yosefe, S., Roberts, J., Woodworth, S., Mmbaga, B., Redman, C. & Hirst, J.E. 2022. Use of mHealth tools to register birth outcomes in low-income and middle-income countries: a scoping review. *BMJ open*, *12*(10): e063886.

Chen, H.M., Kazman, R. & Haziyev, S. 2016. Agile big data analytics for web-based systems: an architecture-centric approach. *IEEE Transactions on Big Data*, 2(3):234–248.

Chitpin, S. & Karoui, O. 2021. Culturally responsive pedagogy: A Canadian perspective. *Journal of Higher Education Policy and Leadership Studies*, 2(1): 45–62.

Chitsa, F. & Iyamu, T. 2020. Towards enterprise technical architecture for the implementation of the South African NHIA. *Advances in Science, Technology and Engineering Systems Journal*, 5(2):724–729.

Cockcroft, S. & Russell, M. 2018. Big data opportunities for accounting and finance practice and research. *Australian Accounting Review*, 28(3):323–333.

Cornwell, N., Bilson, C., Gepp, A., Stern, S. & Vanstone, B.J. 2023. The role of data analytics within operational risk management: A systematic review from the financial services and energy sectors. *Journal of the Operational Research Society*, 74(1):374–402.

Costa, C. & Santos, M.Y. 2016. *BASIS: A big data architecture for smart cities. In 2016 Science and Information Computing Conference*,15–17 July 2016. United Kingdom: IEEE, 1247–1256.

Creswell, J. W. & Creswell, J. D. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. 5th ed. Los Angeles: Sage publications.

Creswell, J.W. & Poth, C.N. 2016. *Qualitative inquiry and research design: Choosing among five approaches*. 2nd ed. London: Sage publications.

Dalglish, S.L., Khalid, H. & McMahon, S.A. 2020. Document analysis in health policy research: the READ approach. *Health policy and planning*, 35(10):1424–1431.

Dane, F. C. 2010. *Evaluating research: Methodology for people who need to read research*. 2nd ed. London: Sage.

Davies, C., & Fisher, M. 2018. Understanding research paradigms. *Journal of the Australasian Rehabilitation Nurses' Association*, 21(3), 21–25.

Dennehy, D. & Conboy, K. 2017. Going with the flow: an activity theory analysis of flow techniques in software development. *Journal of Systems and Software*, 133: 160–173.

Dezi, L., Santoro, G., Gabteni, H. & Pellicelli, A.C. 2018. The role of big data in shaping ambidextrous business process management: Case studies from the service industry. *Business Process Management Journal*, 24(5):1163–1175.

Dhaliwal, G. & Shojania, K.G. 2018. The data of diagnostic error: big, large and small. *BMJ Quality & Safety*, 27(7):499–501.

Dicuonzo, G., Galeone, G., Shini, M. & Massari, A. 2022. Towards the use of big data in healthcare: A literature review. *Healthcare*, 10(7): 1232.

Dufour, I.F. & Richard, M.C. 2019. Theorizing from secondary qualitative data: A comparison of two data analysis methods. *Cogent Education*, 6(1):1–15.

Dumitriu, D. & Popescu, M.A.M. 2020. Enterprise architecture framework design in IT management. *Procedia Manufacturing*, *46*: 932–940.

Dwivedi, A., Moktadir, M. A., Jabbour, C. J. C., & de Carvalho, D. E. 2022. Integrating the circular economy and industry 4.0 for sustainable development: Implications for responsible footwear production in a big data-driven world. *Technological Forecasting and Social Change*, *175*: 121335.

Engeström, Y. & Pyörälä, E. 2021. Using Activity Theory to transform medical work and Learning. *Medical Teacher*, 43(1):7–13.

Engeström, Y., Lompscher, J. & Rückriem, G. (eds.). 2016. *Putting activity theory to work: Contributions from developmental work research*. 13th ed. Berlin: Lehmanns Media.

Er, M., Kay, R. & Lawrence, E. 2010 Information systems and activity theory: a case study of doctors and mobile knowledge work. In *2010 Seventh International Conference on Information Technology: New Generations*, 12–14 April, Las Vegas: IEEE, 603–607.

Ethiraj, S. K., & Posen, H. E. 2014. Do product architectures affect innovation productivity in complex product ecosystems?. In *Collaboration and competition in business ecosystems*. Bingley: Emerald Group Publishing Limited.

Faraway, J.J. & Augustin, N.H. 2018. When small data bits big data. *Statistics and probability letters*, 136(2018):142–145.

Farooqi, M.M., Shah, M.A., Wahid, A., Akhunzada, A., Khan, F., ul Amin, N. & Ali, I. 2019. Big data in healthcare: A survey. In Khan, F. (ed.). *Applications of intelligent technologies in healthcare*. Switzerland: Springer, 143–152.

Ferraris, A., Mazzoleni, A., Devalle, A. & Couturier, J. 2019. Big data analytics capabilities and knowledge management: impact on firm performance. *Management Decision*, *57*(8):1923 –1936.

Filaly, Y., Berros, N., Badri, H., Mendil, F.E. & EL Idrissi, Y.E.B. 2023. Security of Hadoop Framework in Big Data. In Farhaoui, Y., Rocha, A., Brahmia, Z. & Bhushab, B. (ed.). *Artificial Intelligence and Smart Environment.* Cham: Springer International Publishing, 709–715.

Fischer, C., Pardos, Z.A., Baker, R.S., Williams, J.J., Smyth, P., Yu, R., Slater, S., Baker, R. & Warschauer, M. 2020. Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1):130–160.

Garoufallou, E. & Gaitanou, P. 2021. Big data: opportunities and challenges in libraries, a systematic literature review. *College & Research Libraries*, 82(3):410–435.

Gedera, D.S. & Williams, P.J. 2015. *Activity theory in education: Research and practice*. In Dilani, S.P., Gedera, D.S. & William, P.J, (eds.). Activity theory in education Rotterdam: Sense publishers, 1-15.

Georgiadis, G., & Poels, G. 2021. Enterprise architecture management as a solution for addressing general data protection regulation requirements in a big data context: a systematic mapping study. *Information Systems and e-Business Management*, *19*: 313–362.

Ghasemaghaei, M., & Calic, G. 2019. Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, *104*: 69–84.

Gil, D., Johnsson, M., Mora, H. & Szymański, J. 2019. Review of the complexity of managing big data of the internet of things. *Complexity*, 2019:1–12.

Glass R., Ramesh V., & Vessey I. 2004. An analysis of research in computing disciplines. *Communications of the ACM, 47* (6): 89–94.

Gökalp, M.O., Kayabay, K., Zaki, M., Koçyiğit, A., Eren, P.E. & Neely, A. 2019. Open-source big data analytics architecture for businesses. In *1st International Informatics and Software Engineering Conference,* 6-7 November 2019. Turkey: IEEE, 1–6.

Goldstein, I., Spatt, C.S. & Ye, M. 2021. Big data in finance. *The Review of Financial Studies*, *34*(7):3213–3225.

Gravetter, F.J. & Forzano, L.A.B. 2018. *Research methods for the behavioral sciences*. 6th ed. United States: Cengage learning.

Gupta, A. & Gupta, N. 2022. *Research methodology*. 1st ed. India: SBPD Publications.

Hajli, N., Shirazi, F., Tajvidi, M. & Huda, N. 2021. Towards an understanding of privacy management architecture in big data: an experimental research. *British Journal of Management*, *32*(2): 548–565.

Hancock, D.R., Algozzine, B. & Lim, J.H. 2021. *Doing case study research: A practical guide*

for beginning researchers. 4th ed. New York: Teachers College Press
Hariri, R.H., Fredericks, E.M. & Bowers, K.M. 2019. Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, 6(1):1–16.
Hasan, H. & Kazlauskas, A. 2014. Activity Theory: who is doing what, why and how. In H. Hasan (Eds.). *Being Practical with Theory: A Window into Business Research*. Wollongong, Australia: 9–14.
Hassan, F., Shaheen, M.E. & Sahal, R. 2020. Real-time healthcare monitoring system using online machine learning and spark streaming. *International Journal of Advanced Computer Science and Applications*, *11*(9): 650-658.
He, W., Hung, J. L., & Liu, L. 2023. Impact of big data analytics on banking: a case study. *Journal of Enterprise Information Management*, 36(2): 459–479.
Hung, S.Y., Chen, C., Choi, H.S. & Ractham, P. 2021. A holistic framework to examine the impact of user, organizational and data factors on the use of big data analytics systems. *Information Research*, 26(4):2015–2219.
Iyamu, T. 2023. *Advancing Big Data Analytics for Healthcare Service Delivery*. Taylor & Francis.
Iyamu, T. 2022a. *Enterprise Architecture for Strategic Management of Modern IT Solutions*. 1st ed. CRC Press.
Iyamu, T. 2022b. *Applying Theories for Information Systems Research*. 1st ed. New York: Routledge.
Iyamu, T. 2022c. *Advancing Big Data Analytics for Healthcare Service Delivery*. London: Taylor & Francis.
Iyamu, T. 2020. A framework for selecting analytics tools to improve healthcare big data usefulness in developing countries. *South African Journal of Information Management*, 22(1):1–9.
Iyamu, T. 2019. The architectures of data and information: Their confounded confusion. *South African Journal of Information Management*, *21*(1):1–9.
Iyamu, T. 2018a. A multilevel approach to big data analysis using analytic tools and actor network theory. *South African Journal of Information Management*, 20(1):1–9.
Iyamu, T. 2018b. Collecting qualitative data for information systems studies: The reality in practice. *Education and Information Technologies*, 23:2249–2264.
Iyamu, T. 2013a. Institutionalisation of the enterprise architecture: The actor-network perspective. In *Social and Professional Applications of Actor-Network Theory for Technology Development*: 144–155. IGI Global.
Iyamu, T. 2013b. Underpinning theories: order-of-use in information systems research, *Journal Of Systems and Information Technology*, 15(3):224–238.
Iyamu, T. 2010. Human Interaction with Structure in the Computing Environment. In

Proceedings of the Sixteenth Americas Conference on Information Systems, 12–15 August 2010. Peru: AIS Press.

Iyamu, T. & Nunu, K. 2021. Healthcare data management conceptual framework for service delivery. *Education and Information Technologies*, *26*: 3513–3527.

Iyamu, T. & Shaanika, I. 2019a. The use of activity theory to guide information systems research. *Education and Information Technologies*, 24 (2019):165–180.

Iyamu, T. & Shaanika, I.N. 2019b. Deployment of Enterprise Architecture from the Activity Theory Perspective. In *Advanced Methodologies and Technologies in Business Operations and Management*. IGI Global :790–801.

Iyamu, T., Nehemia-Maletzky, M. & Shaanika, I. 2016. The overlapping nature of business analysis and business architecture: What we need to know. *Electronic Journal of Information Systems Evaluation*, *19*(3):169–179.

Izonin, I., Tkachenko, R., Dronyuk, I., Tkachenko, P., Gregus, M. & Rashkevych, M. 2021. Predictive modeling based on small data in clinical medicine: RBF-based additive input- doubling method. *Mathematical Biosciences and Engineering*, 18(3):2599–2613.

Jaiswal, A., Dwivedi, V.K. & Yadav, O.P. 2020. Big data and its analyzing tools: A perspective. In *Proceedings of 6th International Conference on Advanced Computing and Communication Systems*, 6-7 March 2020. Coimbatore: IEEE, 560–565.

Jeske, D. & Calvard, T. 2020. Big data: lessons for employers and employees. *Employee Relations: The International Journal*, 42(1):248–261.

Jin, C., Li, F., Ma, S., & Wang, Y. 2022. Sampling scheme-based classification rule mining method using decision tree in big data environment. *Knowledge-Based Systems*, 244: 108522.

Jinoria, K. 2014. The definition of business enterprise overview of legal and economic approaches. In Proceedings of European Scientific Journal, 23–26 October 2014. Tbilisi: *European Scientific Journal*, 1857–7881.

Jones, M. 2019. What we talk about when we talk about (big) data. *The Journal of Strategic Information Systems*, 28(1): 3–16.

Kajamaa, A., Mattick, K., & De-la-Croix, A. 2020. How to do mixed-methods research. The clinical Teacher, 17(3):267–271.

Kalipe, G.K. & Behera, R.K. 2019. Big Data Architectures: A detailed and application oriented review. *International Journal of Innovative Technology and Exploring. Engineering*, 8(2019): 2182–2190.

Kallio, H., Pietilä, A.M., Johnson, M. & Kangasniemi, M. 2016. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, 72(12):2954–2965.

Kamble, S. S., Belhadi, A., Gunasekaran, A., Ganapathy, L., & Verma, S. 2021. A large multigroup decision-making technique for prioritizing the big data-driven circular economy practices in the automobile component manufacturing industry. *Technological Forecasting and Social Change*, 165*:*120567.

Kaptelinin, V. & Nardi, B. 2018. Activity theory as a framework for human-technology interaction research. *Mind, Culture, and Activity*, *25*(1):3–5.

Kaptelinin, V. & Nardi, B.A. 2006. *Acting with technology: Activity theory and interaction design*. 1st ed. Massachusetts Institute of Technology Press.

Kaptelinin, V., Kuutti, K. & Bannon, L. 1995. Activity Theory: Basic Concepts and Applications: A summary of a tutorial given at the East West HCI 95 Conference. In *International Conference on Human-Computer Interaction,* 03 July 1995. Berlin: Heidelberg, Springer, 79–90.

Kapur, R. 2018*. Research methodology: Methods and strategies*. New Delhi: Department of Adult Education and Continuing Extension.

Karanasios, S. & Allen, D. 2014. Mobile technology in mobile work: Contradictions and congruencies in activity systems. *European Journal of Information Systems*, 23(5): 529–542.

Karanasios, S. & Allen, D. 2013. ICT for development in the context of the closure of Chernobyl nuclear power plant: an activity theory perspective, *Information Systems Journal*, 23(4):287–306.

Karppinen, K. & Moe, H. 2012. What we talk about when we talk about document analysis. *Trends in communication policy research: New theories, methods and subjects*,177–193.

Kay, M. 2017. The elusive definition of an enterprise CMS. *EContent*, 40(2):10–11.

Keele, R. 2010. Quantitative Versus Qualitative Research, or Both? In *R. Keele, ed. Nursing research and evidence-based practice: Ten steps to success*. 35–52.

Kelly, P. 2018. An activity theory study of data, knowledge, and power in the design of an international development NGO impact evaluation. *Information Systems Journal*, 28(3): 465–488.

Khanna, A., Gupta, D. & Dey, N. (eds.). 2021. *Applications of big data in healthcare: theory and practice*. Academic Press.

Khine, P.P. & Wang, Z.S. 2017. Data lake: a new ideology in big data era. In *Proceedings of 4th Annual International Conference on Wireless Communication and Sensor Network*, 15 – 17 December 2017.China: EDP Sciences, 1–11.

Kim, M., Zimmermann, T., DeLine, R. & Begel, A. 2016. The emerging role of data scientists on software development teams. In *Proceedings of the 38th International Conference on Software Engineering*, 14–22 May 2016. Texas: ACM, 96–107.

Kirongo, A. & Odoyo, C. 2020. Research philosophy design and methodologies: A systematic review of research paradigms in information technology. *Global Scientific Journals*, 8(5):33–38

Kitchin, R. 2021. *Data lives: How data are made and shape our world*. Bristol University Press. Britain.

Kivunja, C. 2018. Distinguishing between theory, theoretical framework, and conceptual framework: A systematic review of lessons from the field. *International journal of higher education*, *7*(6): 44–53.

Kothari, C.R. 2020. *Research methodology methods and techniques*. 2nd ed. New Delhi: New Age International (P) Ltd., Publishers.

Kumar, R. 2018. *Research methodology: A step-by-step guide for beginners*. 5th ed. London: Sage.

Lang, C., Pott, A. & Shinozaki, K. 2021. Organisations and the production of migration and I n/exclusion. *Comparative migration studies*, 9(60):1–16.

Lankhorst, M. 2017. *Enterprise architecture at work: Modelling, communication and analysis*. 4th ed. Springer.

Lapalme, J., Gerber, A., Van der Merwe, A., Zachman, J., De Vries, M. & Hinkelmann, K., 2016. Exploring the future of enterprise architecture: A Zachman perspective. *Computers in Industry*, *79*:103–113.

Lauri, M.A. 2019. Write a Scientific Paper: Collecting qualitative data using focus groups. *Early human development*, 133:65–68.

Leavy, P. 2022. *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*. 2nd ed. London: Guilford Publications.

Leech NL, Collins KM, & Onwuegbuzie AJ. 2017. *Collecting Qualitative Data to Enhance Social Network Analysis and Data Mining*. In: Alhajj R, Rokne J, (eds.). Encyclopedia on Social Network Analysis and Data Mining. New York: Springer,1–11.

Leedy P.D. & Ormrod, J.E. 2014. *Practical Research: Planning and Design*. 10th ed. Upper Saddle River, NJ: Pearson.

Lee-Easton, M.J., Magura, S. & Maranda, M.J. 2022. Utilization of evidence-based intervention criteria in US Federal grant funding announcements for behavioral healthcare. *The Journal of Health Care Organization, Provision, and Financing*, 59: 1–16.

Li, J., Ye, Z., & Zhang, C. 2022. Study on the interaction between big data and artificial intelligence. *Systems Research and Behavioral Science*, *39*(3):641–648.

Lioutas, E.D., Charatsari, C., La Rocca, G. & De Rosa, M. 2019. Key questions on the use of big data in farming: An activity theory approach. *NJAS-Wageningen Journal of Life Sciences*, 90:1–12.

Low, J., Saks, M. & Allsop, J. 2019. *Unstructured and Semi-structured interviews in Health*

Research. Researching Health: Qualitative, Quantitative and Mixed methods. 3rd ed. London: Sage publications.

Lu, Y., Kandula, S., König, A.C. & Chaudhuri, S. 2021. Pre-training summarization models of structured datasets for cardinality estimation. *Proceedings of the VLDB Endowment*, 15(3):414–426.

Lynch, J., Ramjan, L.M., Glew, P.J. & Salamonson, Y. 2020. How to embed a conceptual or theoretical framework into a dissertation study design. *Nurse researcher*, 28(3): 24 –29.

Madden, R. 2010. *Being ethnographic: A guide to the theory and practice of ethnography*. 1st ed. London: Sage Publications.

Makovhololo, P., Shaanika, I.N., Sekgweleo, T. & Okigui, H.H. 2017. Use of Activity Theory to comprehend the impacts of communication barriers in healthcare data analytics. *International Journal of Computer Science and Information Security*, 15(9):70–78.

Mandinach, E.B. & Abrams, L.M. 2022. Data literacy and learning analytics. *Handbook of learning analytics*,196–204.

Malmi, T. 2016. Managerialist studies in management accounting: 1990–2014. *Management Accounting Research*, 31: 31–44.

Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P.M., Sundarasekar, R. & Thota, C. 2018. A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. *Future Generation Computer Systems*, 82(2018):375–3 87.

Mårtensson, P., Fors, U., Wallin, S. B., Zander, U., & Nilsson, G. H. 2016. Evaluating research: A multidisciplinary approach to assessing research practice and quality. *Research Policy*, 45(3): 593-603.

McAbee, S.T., Landis, R.S. & Burke, M.I. 2017. Inductive reasoning: The promise of big data. *Human Resource Management Review*, 27(2):277–290.

McIntosh, M. & Morse, J.M. 2015. Situating and constructing diversity in semi-structured interviews. *Global Qualitative Nursing Research*. 1–12.

Melnikovas, A. 2018. Towards an explicit research methodology: Adapting research onion model for future studies. *Journal of futures Studies*,23(2): 29–44.

Merriam, S.B. & Tisdell, E.J. 2015. *Qualitative research: A guide to design and implementation*. 4th edition. John Wiley & Sons, Incorporated.

Mgudlwa, S. & Iyamu, T. 2018. Integration of social media with healthcare big data for improved service delivery. *South African Journal of Information Management*, 20(1):1–8.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. 2019. Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98: 261–276.

Mishra, R. & Sharma, R. 2015. Big data: opportunities and challenges. *International Journal of Computer Science and Mobile Computing*, 4(6):27–35.

Mitani, A.A. & Haneuse, S. 2020. Small data challenges of studying rare diseases. *JAMA network open*, 3(3):e201965–e201965.

Mitragotri, P. & Pal, V.K. 2019. Overview of Machine Learning Applications in Financial Sector. *International Journal of Management, IT and Engineering*, 9(8):1–6.

Mkhomazi, S.S. & Iyamu, T. 2013. A guide to selecting theory to underpin information systems studies. In *Grand Successes and Failures in IT. Public and Private Sectors: IFIP WG 8.6 International Working Conference on Transfer and Diffusion of IT,* 27–29 June 2013. India: Springer, 525–537.

Mohajan, H. K. 2018. Qualitative research methodology in social sciences and related subjects. *Journal of Economic Development, Environment and People*, 7(1):23–48.

Mohammad, A., Mcheick, H. & Grant, E. 2014. Big data architecture evolution: 2014 and beyond. In *Proceedings of the fourth ACM international symposium on development and analysis of intelligent vehicular networks and applications*, 21–26 September 2014. Canada: ACM, 139–144.

Mokhtari, G., Anvari-Moghaddam, A. & Zhang, Q. 2019. A new layered architecture for future big data-driven smart homes. *IEEE Access*, 7(2019): 19002–19012.

Moreno, J., Serrano, M.A., Fernandez-Medina, E. & Fernandez, E.B. 2018. Towards a Security Reference Architecture for Big Data. In *Proceedings of 20th International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data*, 26 -29 March 2018. Austria: DOLAP, 1–9.

Morgan, H. 2022. Conducting a qualitative document analysis. *The Qualitative Report*, 27(1): 64–77.

Mostefaoui, A., Merzoug, M.A., Haroun, A., Nassar, A. & Dessables, F. 2022. Big data architecture for connected vehicles: Feedback and application examples from an automotive group. *Future Generation Computer Systems*, *134*: 374–387.

Mphahlele, L. & Iyamu, T. 2010. The Role of Principles in the Deployment of Enterprise Information Architecture. In *International Conference on Informatics and Semiotics in Organisations*, UK, 19-21 July 2010. SciTePress: 299–307.

Mukherjee, S.P. 2019. *A guide to research methodology: An overview of research problems, tasks and methods*. London: CRC Press.

Mulcahy, L. Rossner, M. & Tsalapatanis, A. 2021. It's about time: investigating the temporal in socio-legal studies through unstructured interviews. *Journal of Law and Society*, *48*: S104–S117.

Mustapha, S.S. 2022. The UAE Employees' Perceptions towards Factors for Sustaining Big and Continuous Impact on Their Organization's Performance. *Sustainability*, 14(22):15271.

Mutasa, L. S., & Iyamu, T. 2023. Application of activity theory to examine the implementation of e-health in Namibia. *International Journal of Healthcare Management*, 16(2): 157–166.

Musukutwa, S.C. 2022*. Developing an Enterprise Architecture. In SAP Enterprise Architecture: A Blueprint for Executing Digital Transformation*. CA: Apress, 51–92.

Mutudi, M., Nehemia, M. & Iyamu, T. 2020. *Qualitative Methods, Lesson for Information Systems Researchers*. In Arthur T. (ed.). Encyclopaedia of Education and Information Technologies. Australia: Springer International Publishing, 1413–1420.

Myers, M.D. 2013. *Qualitative research in business and management*. 2nd ed. London: Sage Publications.

Nagy, J., Oláh, J., Erdei, E., Máté, D & Popp, J. 2018. The role and impact of Industry 4.0 and the internet of things on the business strategy of the value chain the case of Hungary. *Sustainability*, 10(10): 1–25.

Nardi, B.A. 1996. Activity theory and human-computer interaction. *Context and consciousness: Activity theory and human-computer interaction*, 436: 7–16.

Necsulescu, N. 2017. Focusing on small data to drive big results. *Applied Marketing Analytics*, 2(4): 296–303.

Nehemia, M. & Zondani, T. 2021. The Use of an Enterprise Architecture Framework to Guide the Management of Big Data in Health Organisations. In *Empowering Businesses With Collaborative Enterprise Architecture Frameworks*, 39-50. IGI Global.

Nehemia-Maletzky, M., Iyamu, T. & Shaanika, I. 2018. The use of activity theory and actor network theory as lenses to underpin information systems studies. *Journal of Systems and Information Technology*, 20(2): 191–206.

Nicholas, D., Watkinson, A., Jamali, H.R., Herman, E., Tenopir, C., Volentine, R., Allard, S. and Levine, K. 2015. Peer review: Still king in the digital age. *Learned Publishing*, *28*(1):15–21.

Nyikana, W. & Iyamu, T. 2023. The taxonomical distinction between the concepts of small data and big data. In *Proceedings of the 16th International Association for Development of Information Society*, 11 – 13 March 2023. Lisbon: IADIS, 138–146.

Nyikana, W. & Iyamu, T. 2022. A Guide for selecting big data analytics tools in an organisation. In *Proceedings of the 55th Hawaii International Conference on System Sciences, 04 – 07 January 2022. Hawaii: HICSS*, 5451–5461.

Oussous, A., Benjelloun, F.Z., Lahcen, A.A. & Belfkih, S. 2018. Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4):431–448.

Pääkkönen, P. & Pakkala, D. 2020. Extending reference architecture of big data systems towards machine learning in edge computing environments. *Journal of Big Data*, 7(1):1–29.

Papaioannou, D., Sutton, A. & Booth, A. 2016. Systematic approaches to a successful literature review. *Systematic approaches to a successful literature review*,1–336.

Pau, M., Kapsalis, P., Pan, Z., Korbakis, G., Pellegrino, D. & Monti, A. 2022. MATRYCS—A Big Data Architecture for Advanced Services in the Building Domain. *Energies*, 15(7):2568.

Paul, J. & Criado, A.R. 2020. The art of writing literature review: What do we know and what do we need to know?. *International Business Review*, 29(4): 101717.

Parmaxi, A., Zaphiris, P. & Ioannou, A. 2016. Enacting artefact-based activities for social technologies in language learning using a design-based research approach. *Computers in Human Behaviour*, 63: 556–567.

Park, S.H. and Kressel, H.Y., 2018. Connecting technological innovation in artificial intelligence to real-world medical practice through rigorous clinical validation: what peer-reviewed medical journals could do. *Journal of Korean Medical Science*, *33*(22).

Park, S., Cho, Y., Yoon, S.W. & Han, H. 2013. Comparing team learning approaches through the lens of activity theory. *European Journal of Training and Development*, *37*(9): 788–810.

Pasztor, J. 2015. What Is Ethics, Anyway?. *Journal of Financial Service Professionals*, 69(6):30–32.

Patel, M. & Patel, N. 2019. Exploring Research Methodology. *International Journal of Research and Review*, 6(3):48–55.

Pathak, A. & Intratat, C. 2012. Use of c interviews to investigate teacher perceptions of student collaboration. *Malaysian Journal of ELT Research*, 8(1):1–10.

Patten, M.L 2017. *Understanding research methods: An overview of the essentials*. 10th ed. New York: Routledge.

Pesqueira, A., Sousa, M. J., & Rocha, Á. 2020. Big data skills sustainable development in healthcare and pharmaceuticals. *Journal of Medical Systems*, 44: 1–15.

Poniatowski, M., Lüttenberg, H., Beverungen, D. & Kundisch, D. 2021. Three layers of abstraction: a conceptual framework for theorizing digital multi-sided platforms. *Information systems and e-business management*, 20: 1– 27.

Pramanik, P.K.D., Pal, S. & Mukhopadhyay, M. 2022. Healthcare big data: A comprehensive overview. In *Research anthology on big data analytics, architectures, and applications*,119–147. IGI Global.

Pratsri, S. & Nilsook, P. 2020. Design on Big Data Platform-Based in Higher Education Institute. *Higher Education Studies*, 10(4):36–43.

Prasetyo, B., Aziz, F.S., Faqih, K., Primadi, W., Herdianto, R. & Febriantoro, W.2019. A review: evolution of big data in developing country. *Bulletin of Social Informatics Theory and Application*, 3(1):30–37.

Purchase, V., Parry, G., Valerdi, R., Nightingale, D. & Mills, J. 2011. Enterprise transformation: Why are we interested, what is it, and what are the challenges?. *Journal of Enterprise Transformation*, 1(1): 14–33.

Qi, Q., Xu, Z., & Rani, P. 2023. Big data analytics challenges to implementing the intelligent Industrial Internet of Things (IIoT) systems in sustainable manufacturing operations. *Technological Forecasting and Social Change*, *190*: 122401.

Rao, T. R., Mitra, P., Bhatt, R., & Goswami, A. 2019. The big data system, components, tools, and technologies: a survey. *Knowledge and Information Systems*, 60: 1165–1245.

Ravikumar, T., Sriram, M., & Murugan, N. 2022. Applications and Risks of Big Data in Financial Services. In *Proceedings of 10th International Conference on Emerging Trends in Corporate Finance and Financial Markets*, 13–14 October 2022. India: SDMIMD, 1–7.

Rizk, A., Bergvall-Kåreborn, B. & Elragal, A. 2018. Towards a taxonomy for data-driven digital services. In *Proceedings of the 51st Hawaii international conference on system sciences,* Hawaii, 3 – 6 January 2018. USA: HICSS: 1076–1085.

Ruiz, M.D., Gómez-Romero, J., Fernandez-Basso, C. & Martin-Bautista, M.J. 2021. Big data architecture for building energy management systems. *IEEE Transactions on Industrial Informatics*, *18*(9):5738–5747.

Rotondo, A. & Quilligan, F. 2020. Evolution paths for knowledge discovery and data mining process models. *Computer Science*, 1(2):1–19.

Saddad, E., El-Bastawissy, A., Mokhtar, H.M. & Hazman, M. 2020. Lake data warehouse architecture for big data solutions. *International Journal of Advanced Computer Science and Applications*, 11(8):417–424.

Saggi, M.K. & Jain, S. 2018. A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, 54(5):758–790.

Sandhu, A.K. 2021. Big data with cloud computing: Discussions and challenges. *Big Data Mining and Analytics*, 5(1): 32–40.

Sani, J.A., Sharip, N.A.A., Othman, N. & Hussain, M.R.M. 2018. Relationship between types of organization with the quality of soft-scape construction work in Malaysia. *Asian Journal of Quality of Life*, 3(12):137–146.

Sannino, A. & Engeström, Y. 2018. Cultural-historical activity theory: Founding insights and new challenges. *Cultural-historical psychology*,14(3):43–56.

Saunders, M., Lewis, P. & Thornhill, A. 2019. *Research methods for business students*. 8th ed. England: Pearson Education.

Sekgweleo, T., Makovhololo, P. & Iyamu, T. 2017. The connectedness in selecting socio-technical theory to underpin information systems studies. *Journal of Contemporary Management*, 14(1): 097–1117.

Sestino, A., & De Mauro, A. 2022. Leveraging artificial intelligence in business: Implications, applications and methods. *Technology Analysis & Strategic Management*, *34*(1): 16–29.

Sivarajah, U., Kamal, M.M., Irani, Z. & Weerakkody, V. 2017. Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70(2017):263–286.

Shaanika, I.N. 2022. The Use of Mixed-Methods as a Research Strategy in Information Systems Studies. In *Proceedings of the 13th International Conference on Society and Information Technologies*, 8–11 March 2022. Virtual Event: ICSIT, 50 – 55.

Shaanika, I.N. 2019. A unified architecture framework for healthcare mobile systems to improve big data usefulness in a government environment. Unpublished Doctoral dissertation, Cape Peninsula University of Technology. https://etd.cput.ac.za/handle/20.500.11838/2963.

Shaanika, I. & Iyamu, T. 2018. Developing the enterprise architecture for the Namibian government. *The Electronic Journal of Information Systems in Developing Countries*: 84(3):1–11.

Shaanika, I. & Iyamu, T. 2015. Deployment of enterprise architecture in the Namibian government: The use of activity theory to examine the influencing factors. *The Electronic Journal of Information Systems in Developing Countries*, 71(1):1–21.

Shaanika, I.N., Makovhololo, P., Okigui, H.H. & Sekgweleo, T. 2017. Contextualising Literature Review Challenges in Information Systems Studies. *International Journal of Innovative Research and Advanced Studies*, 4(10): 92–98.

Shafer, S.M., Smith, H.J. & Linder, J.C. 2005. The power of business models. *Business horizons*, 48(3): 199–207.

Sherif, V. 2018. Evaluating preexisting qualitative research data for secondary analysis. Forum Qualitative Sozialforschung/Forum: *Qualitative Social Research*, 19(2): 26– 42.

Siyal, A.A., Junejo, A.Z., Zawish, M., Ahmed, K., Khalil, A. & Soursou, G. 2019. Applications of blockchain technology in medicine and healthcare: Challenges and future perspectives. *Cryptography*, 3(3):1–16.

Smajevic, M. & Bork, D. 2021. Towards graph-based analysis of enterprise architecture models. In *Proceeding of 40th International Conference*, 18–21 October 2021: Canada: Springer International Publishing, 199–209.

Striy, L., Irina, S., Alizade, H., Rizwan, B. & Rustamov, O. 2019. New Types of Activity of Enterprises of Telecommunications. *Education, Law, Business*, 2019: 47–199.

Soltani Delgosha, M., Hajiheydari, N. & Fahimi, S.M. 2021. Elucidation of big data analytics in banking: a four-stage Delphi study. *Journal of Enterprise Information Management*, 34(6): 1577–1596.

Sovacool, B.K., Axsen, J. & Sorrell, S. 2018. Promoting novelty, rigor, and style in energy social science: Towards codes of practice for appropriate methods and research design. *Energy Research & Social Science*, 45(2018):12–42.

Ştefan, R.M. 2012. A comparison of data classification methods. *Procedia Economics and Finance*, *3*:420-425.

Stuckey, H. 2013. Qualitative research methods in social health. *Journal of Social Science Health and Diabetes*, 1(2):56.

Svensson, J. & Poveda Guillen, O. 2020. What is Data and What Can it be Used For?: Key Questions in the Age of Burgeoning Data-essentialism. *Journal of Digital Social Research*, 2(3):65–83.

Taherdoost, H. 2022. Different types of data analysis; data analysis methods and techniques in research projects. *International Journal of Academic Research in Management*, 9(1):1–9.

Tekaya, B., Feki, S.E., Tekaya, T. & Masri, H. 2020. Recent applications of big data in finance. In *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*, 15 – 17 October 2020. Tunisia: DTUC, 1–6.

Tembusai, Z.R., Mawengkang, H. & Zarlis, M. 2021. K-nearest neighbor with k-fold cross validation and analytic hierarchy process on data classification. *International Journal of Advances in Data and Information Systems*, *2*(1):1– 8.

Thota, C., Manogaran, G., Lopez, D. & Vijayakumar, V. 2017. Big data security framework for distributed cloud data centers. In *Cybersecurity breaches and issues surrounding online threat protection*. 288–310. IGI global.

Tschoppe, N. & Drews, P. 2022. Developing Digitalization Strategies for SMEs: A Lightweight Architecture-based Method. In *Proceedings of the 55th Hawaii International Conference on System Sciences, 04 – 07 January 2022. Hawaii: HICSS*, 1–10.

Todman, L. C., Bush, A., & Hood, A. S. 2023. Small Data for big insights in ecology. *Trends in Ecology & Evolution*. 38(7): 615–622.

Tümen-Akyildiz, S.T. & Ahmed, K.H. 2021. An overview of qualitative research and focus group discussion. *International Journal of Academic Research in Education*, 7(1):1–15.

Tungela, N., Mutudi, M. & Iyamu, T. 2018. The roles of e-government in healthcare from the perspective of structuration theory. In *Proceedings of 2018 Open Innovations Conference*, 3 -5 October. Johannesburg: IEEE, 332–338.

Tungela, N. 2021. An information technology readiness assessment framework for the South African national health insurance. Unpublished master's thesis, Cape Peninsula University of Technology, Bellville. https://etd.cput.ac.za/handle/20.500.11838/3527 [November 2021].

Tupper, C. 2011. *Data architecture: from zen to reality*. 1st ed. USA: Elsevier.

Uden, L., Lu, W. & Ting, I.H. 2017. Knowledge Management in Organizations: In *12th International Conference in Knowledge Management in Organisations*, 21–24 August 2017. Beijing: Springer, 460–469.

Vahedi, L., Qushua, N., Seff, I., Doering, M., Stoll, C., Bartels, S.A. & Stark, L. 2023. Methodological and ethical implications of using remote data collection tools to measure sexual and reproductive health and gender-based violence outcomes among women and girls in humanitarian and fragile settings: a mixed methods systematic review of peer-reviewed research. *Trauma, Violence, & Abuse*, *24*(4): 2498–2529.

Van der Merwe, A., Gerber, A. & Smuts, H. 2020. *Guidelines for Conducting Design Science Research in Information Systems*. In: Tait, B., Kroeze, J., Gruner, S. (eds.). ICT Education. SACLA 2019. Communications in Computer and Information Science, vol 1136, 163–178. Switzerland: Springer, Cham

Vanderstoep, S.W. & Johnson, D.D. 2009. *Research Methods for everyday life: Blending Qualitative and Quantitative Approaches.* 1st ed. United states of America.

Van Wessel, R. M., Kroon, P., & De Vries, H. J. 2021. Scaling Agile Company-Wide: The Organizational Challenge of Combining Agile-Scaling Frameworks and Enterprise Architecture in Service Companies. *IEEE Transactions on Engineering Management*, *69*(6): 3489–3502.

Varpio, L., Paradis, E., Uijtdehaage, S. & Young, M. 2020. The distinctions between theory, theoretical framework, and conceptual framework. *Academic Medicine*, 95(7):989–994.

Vass, C., Rigby, D. & Payne, K. 2017. The Role of qualitative research methods in discrete choice experiments. *Medical Decision Making*, 37(3): 298–313.

Venkatesh, V., Brown, S. A. & Bala, H. 2013. Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, 37 (1):21–54.

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

Walliman, N. 2021. *Research methods: The basics*. 3rd ed. London: Routledge.

Wang, J., Zhou, Z., Li, B., & Wu, M. 2022. Attribute Network Representation Learning with Dual Autoencoders. *Symmetry*, 14(9): 1840.

Wang, J., Yang, Y., Wang, T., Sherratt, R.S. & Zhang, J. 2020. Big data service architecture: a survey. *Journal of Internet Technology*, 21(2):393–405.

Wang, Y., Kung, L., & Byrd, T. A. 2018. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological forecasting and social change*, *126*:3-13.

Wehn, C., Yang, J., Gan, L. & Pan, Y. 2021. Big data-driven Internet of Things for credit evaluation and early warning in finance. *Future Generation Computer Systems*, 124: 295–307.

Wehn, U. & Almomani, A. 2019. Incentives and barriers for participation in community-based environmental monitoring and information systems: A critical analysis and integration of the literature. *Environmental Science & Policy*, 101:341-357.

Wethington, E. and McDarby, M.L. 2015. Interview methods (structured, semi structured, unstructured). *The encyclopaedia of adulthood and aging*,1–5.

Williamson, K. & Johanson, G. (eds.). 2017. Research Methods: Information, Systems, and Contexts. UK: Chandos Publishing.

Wolski, M. & Gomolińska, A. 2020. Data meaning and knowledge discovery: Semantical aspects of information systems. *International Journal of Approximate Reasoning*, 119:40–57.

Woo, S.E., O'Boyle, E.H. & Spector, P.E. 2017. Best practices in developing, conducting, and evaluating inductive research. *Human resource review*, 27(2): 255–264.

Yaseen, H.K. & Obaid, A.M. 2020. Big Data: Definition, Architecture & Applications. *International Journal on Informatics Visualization*, 4(1):45–51.

Yilmaz, K. 2013. Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European journal of education*, 48(2):311–325.

Ylinen, M. and Pekkola, S. 2020. Jack-of-all-trades torn apart: Skills and competences of an enterprise architect. In *Proceedings of the 28th European Conference on Information Systems,* 15–17 June 2020. Morocco: Association for Information Systems, 1–14.

Yin, R. K. 2017. *Case study research and applications: Design and methods*, 6th ed. Thousand Oaks: SAGE publications.

Yeon, M. S., Lee, Y. K., Pham, D. L., & Kim, K. P. 2022. Experimental Verification on Human-Centric Network-Based Resource Allocation Approaches for Process-Aware Information Systems. *IEEE Access*, 10:23342–23354.

Zeidy, I.A. 2022. The role of financial technology (FinTech) in changing financial industry and increasing efficiency in the economy. *COMESA Monetary Institute. Available at https://www. comesa. int/wp-content/uploads/2022/05/The-Role-of-Financial-Technology. pdf*.

Zhang, Z., Shang, Y., Cheng, L. & Hu, A., 2022. Big data capability and sustainable competitive advantage: The mediating role of ambidextrous innovation strategy. *Sustainability*, *14*(14):1–17.

Zhou, Z., Zhi, Q., Morisaki, S. & Yamamoto, S. 2020. A systematic literature review on enterprise architecture visualization methodologies. *IEEE Access*, *8*:96404–96427.

APPENDIX A: ETHICAL CLEARANCE

Cape Peninsula University of Technology
creating futures

PO Box 1906, Bellville,7535 Symphony Way, Bellville, Cape Town, South Africa
+ 27 (0)21 959 6767 www.facebook.com/cput.ac.za info@cput.ac.za www.cput.ac.za

Office of the Research Ethics Committee
Faculty of Informatics and Design
Room 2.09
80 Roeland Street
Cape Town
Tel: 021-469 1012
Email: ndedem@cput.ac.za
Secretary: Mziyanda Ndede

06 November 2023

Ms Wandisa Nyikana
c/o Department of Information Technology
CPUT

Reference no: 203168283/2023/24

Project title: Architectural design of big data for business enhancement in enterprises

Approval period: 06 November 2023 – 31 December 2024

This is to certify that the Faculty of Informatics and Design Research Ethics Committee of the Cape Peninsula University of Technology approved the methodology and ethics of Ms Wandisa Nyikana (203168283) for Master of Information Communication and Technology.

Any amendments, extension or other modifications to the protocol must be submitted to the Research Ethics Committee for approval.

The Committee must be informed of any serious adverse event and/or termination of the study.

Prof L.J. Theo
Chair: Research Ethics Committee
Faculty of Informatics and Design
Cape Peninsula University of Technology