COMBINING BIG DATA AND TRADITIONAL BUSINESS INTELLIGENCE – A FRAMEWORK FOR A HYBRID DATA-DRIVEN DECISION SUPPORT SYSTEM

by

Lungisa Dotye
(19071516)

Submitted in fulfilment of the requirements for the degree

MIT (Information Systems)

in the

Faculty of Engineering, Built Environment and Information Technology

at the

UNIVERSITY OF PRETORIA

Study leader:

Prof H. Smuts

Date of submission
(21 June 2021)

Declaration regarding Plagiarism

The Department of Informatics emphasises integrity and ethical behaviour with regard to the preparation of all written assignments.

Although the lecturer will provide you with information regarding reference techniques, as well as ways to avoid plagiarism, you also have a responsibility to fulfil in this regard. Should you at any time feel unsure about the requirements, you must consult the lecturer concerned before submitting an assignment.

You are guilty of plagiarism when you extract information from a book, article, web page or any other information source without acknowledging the source and pretend that it is your own work. This doesn't only apply to cases where you quote verbatim, but also when you present someone else's work in a somewhat amended (paraphrased) format or when you use someone else's arguments or ideas without the necessary acknowledgement. You are also guilty of plagiarism if you copy and paste information directly from an electronic source (e.g., a web site, e-mail message, electronic journal article, or CD ROM), even if you acknowledge the source.

You are not allowed to submit another student's previous work as your own. You are furthermore not allowed to let anyone copy or use your work with the intention of presenting it as his/her own.

Students who are guilty of plagiarism will forfeit all credits for the work concerned. In addition, the matter will be referred to the Committee for Discipline (Students) for a ruling. Plagiarism is considered a serious violation of the University's regulations and may lead to your suspension from the University. The University's policy regarding plagiarism is available on the Internet at *http://upetd.up.ac.za/authors/create/plagiarism/students.htm*.

I (full names & surname):	Lungisa Dotye
Student number:	19071516

Declare the following:

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this assignment is my own, original work. Where someone else's work was used (whether from a printed source, the Internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not copy and paste any information directly from an electronic source (e.g., a web page, electronic journal article or CD ROM) into this document.
4. I did not make use of another student's previous work and submitted it as my own.
5. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his/her own work.

Signature

Date

BRIEF TABLE OF CONTENTS

DETAILED TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE OF ABBREVIATIONS

Abbreviation	Full form
IoT	Internet of Things
IT	Information Technology
IS	Information Systems
BI	Business Intelligence
ETL	Extract, Transform, and Load
RP	Research Participant
DBMS	Database Management System
RDBMS	Relational Database Management System
DD-DSS	Data-driven Decision Support System
FVM	Fit-viability model
OIPT	Organisational Information Processing Theory
CIO	Chief Information Officer
ROI	Return on Investment
DIKW	Data, Information, Knowledge, and Wisdom
CSFs	Critical Success Factors
WWW	World Wide Web

COMBINING BIG DATA AND TRADITIONAL BUSINESS INTELLIGENCE – A FRAMEWORK FOR A HYBRID DATA-DRIVEN DECISION SUPPORT SYSTEM

Abstract

Since the emergence of big data, traditional business intelligence systems have been unable to meet most of the information demands in many data-driven organisations. Nowadays, big data analytics is perceived to be the solution to the challenges related to information processing of big data and decision-making of most data-driven organisations. Irrespective of the promised benefits of big data, organisations find it difficult to prove and realise the value of the investment required to develop and maintain big data analytics. The reality of big data is more complex than many organisations' perceptions of big data. Most organisations have failed to implement big data analytics successfully, and some organisations that have implemented these systems are struggling to attain the average promised value of big data. Organisations have realised that it is impractical to migrate the entire traditional business intelligence (BI) system into big data analytics and there is a need to integrate these two types of systems.

Therefore, the purpose of this study was to investigate a framework for creating a hybrid data-driven decision support system that combines components from traditional business intelligence and big data analytics systems. The study employed an interpretive qualitative research methodology to investigate research participants' understanding of the concepts related to big data, a data-driven organisation, business intelligence, and other data analytics perceptions. Semi-structured interviews were held to collect research data and thematic data analysis was used to understand the research participants' feedback information based on their background knowledge and experiences.

The application of the organisational information processing theory (OIPT) and the fit-viability model (FVM) guided the interpretation of the study outcomes and the development of the proposed framework. The findings of the study suggested that data-driven *organisations collect data from different data sources and process these data to transform them into information with the goal of using the information as a base of all their business decisions. Executive and senior management roles in the adoption of a data-driven decision-making culture are key to the success of the organisation. BI and big data analytics are tools and software systems that are used to assist a data-driven organisation in transforming data into information and knowledge.*

The suggested challenges that organisations experience when they are trying to integrate BI and big data analytics were used to guide the development of the framework that can be used to create a hybrid data-driven decision support system. The framework is divided into these elements: business motivation, information requirements, supporting mechanisms, data attributes, supporting processes and hybrid data-driven decision support system architecture. The proposed framework is created to assist data-driven organisations in assessing the components of both business intelligence and big data analytics systems and make a case-by-case decision on which components can be used to satisfy the specific data requirements of an organisation. Therefore, the study contributes to enhancing the existing literature position of the attempt to integrate business intelligence and big data analytics systems.

Keywords: Business intelligence, big data, big data analytics, data warehouse, data-driven organisation, data-driven decisions, decision support system

Acknowledgements

Throughout the writing of this dissertation I have received a great deal of support and assistance.

Firstly, I would like to thank my supervisor, **Professor Hanlie Smuts**, who guided and supported me throughout the study. Her great support and valuable expertise inspired my learning about the academic research process and best practices.

Also, special thanks to all the research participants and my colleagues whose contributions made this study possible. Since they were promised anonymity, I will not mention their names here.

I would also like to specifically thank my line manager, **Adnaan Jones**, for his support and for allowing me to take time off to attend to my studies.

I owe gratitude to my girlfriend, **Pamella Tsheme**, who supported and motivated me throughout the entire study. She had to make personal sacrifices to ensure that I had time to focus on this study.

Finally, I would like to thank my family and friends for all their support and understanding, especially in the final weeks of compiling this paper.

Part I – Introduction

Chapter 1: Introduction

The purpose of this chapter is to give an overview of the background information, problem statement and purpose of the study as guided by the study research questions. The contents of this chapter are outlined in Table 1.

Table 1 – Chapter 1 outline

Section	Section description
1.1	Background Information
1.2	Problem Statement
1.3	Purpose of the Study
1.4	Research Questions
1.5	Research Strategy
1.6	Assumptions
1.7	Limitations
1.8	Outline of the Study

1.1 Background Information

The emergence of digital technologies such as the internet of things (IoT) has forced organisations to change the way they operate. Digital technologies have enabled the automation of data collection and storage through modern tools such as biometric devices, scanners and cell phones (Cai & Zhu, 2015). Many scholars have suggested that the success of most organisations is highly dependent on their ability to process and utilise data that are available to them effectively (Brynjolfsson & McElheran, 2016; Engelbrecht *et al.*, 2016; Merendino *et al.*, 2018; Watson, 2016). Organisations recognise the value of data and they have invested in business intelligence (BI) systems in order to extract value from their data (Ong *et al.*, 2011). However, the amount of data that are available for organisations to exploit has increased to a level that traditional BI systems are unable to handle. According to Merendino *et al.* (2018), these increased volumes and complexity of data are a result of advanced digital technologies that made it possible to collect any kind of data that exists.

Modern information technology (IT) innovations have enabled organisations to collect extensive data from within the organisations' systems as well as from other external data sources such as social media platforms. Multiple formats of data that are large in volume and change very often are referred to as big data. The emergence of big data has changed many organisations' perception of BI and data in general (Lee, 2017; Provost & Fawcett, 2013). Data-driven organisations are exploring new opportunities that can enable them to improve their data-processing capabilities to ensure that they have access to all the data they need to make effective business decisions (Brynjolfsson & McElheran, 2016; Larson & Chang, 2016). Prescriptive and predictive analytics are some of the analytics techniques that data-driven organisations seek to adopt to improve their decision-making. These analytics techniques generally require systems that can handle big data (Sun *et al.*, 2015).

While BI systems use software tools and processes to transform traditional (structured) data into information, big data analytics converts all forms of data into information (Chan, 2013; McAfee *et al.*, 2012). The reliance of BI systems on relational database management systems (RDBMS) to extract, transform and load (ETL) tools has led to major challenges that are related to the processing of big data. BI systems are designed to process structured data. On the other hand, big data analytics systems have gained traction from top management in organisations because they support knowledge discovery of big data. Data-driven organisations are seeking to adopt tools that enable them to create systems that can speedily process big data. Larson and Chang (2016) suggest that data science and fast analytics have emerged in response to the requirement of data-driven organisations to process big data speedily.

Since BI systems are unable to process big data, many organisations are initiating projects to transform BI systems into big data analytics because of the perceived value of big data (Balakrishnan & Rahul, 2018). However, successful transformation and migration of an entire BI system is a challenging exercise and adoption of big data analytics does not guarantee the success of organisational information delivery (Salinas & Lemus, 2017). A number of scholars have indicated that there is a need to investigate ways through which components from big data analytics systems can complementarily be integrated with those from BI systems to deliver results effectively according to the information requirements of a data-driven organisation (Salinas & Lemus, 2017; Santoso, 2017; Tabesh *et al.*, 2019).

1.2 Problem Statement

Many organisations have failed to implement and operate big data analytics systems successfully and most organisations that managed to implement these systems are struggling to attain the average promised value of big data (Tabesh *et al.*, 2019). According to Tabesh *et al*. (2019), 80% of organisations are struggling to adopt big data analytics successfully, and this problem is costing organisations lots of money. The challenges experienced by organisations that are trying to adopt big data analytics have led to many data-driven organisations doubting the practicality and feasibility of the effective adoption of big data-related innovations (Surbakti *et al.*, 2019). Irrespective of the promised and perceived benefits of big data, data-driven organisations still find it difficult to prove and realise the value of their big data investment (Amankwah-Amoah & Adomako, 2019).

Many organisations have senior managers who do not trust the value of big data and are hesitant to invest in big data initiatives. Organisations are struggling to formulate problem definitions whose solutions involve big data adoption, in addition to their lack of justifying the capital required to implement big data analytic systems fully. According to Lee (2017), this is due to the perception that big data analytic systems use emerging technologies. Most executive managers find it risky to invest fully in such initiatives. The reality of big data is much more complex than many organisations' perceptions of big data (Katal *et al.*, 2013). The failure of many organisations to adopt big data analytic systems is problematic, because it forces organisations to rely entirely on traditional BI systems, whose supporting technological tools experience considerable challenges when required to process big data. Therefore, organisations' inability to adopt big data analytics affect the ability to make effective business decisions negatively.

The adoption of big data analytics is perceived to be a challenge for many organisations. However, maintaining traditional BI systems is not a solution either. Data-driven organisations use data to drive business decisions and any challenge encountered during data-processing will affect their business decisions. Effective decision-making determines the success of a data-driven organisation. Because both traditional BI and big data analytics systems support the decision-making of a data-driven organisation, they are perceived to be data-driven decision support systems (DD-DSS). The shortcomings of both BI and big data analytics can lead organisations not to base their decisions fully on data and to use management gut feelings and intuition to drive some decisions (Hedgebeth, 2007). Decisions that are based on gut feelings and intuition are not reliable and can lead to an organisation's failure (Engelbrecht *et al.*, 2016). Given that both traditional BI and big data analytics have their shortcomings, some recent studies have suggested initiating investigations with the aim of combining the two systems (Salinas & Lemus, 2017; Santoso, 2017).

Organisations should reconsider their big data analytics adoption strategy. Currently, most organisations try to replace their BI systems with big data analytics and less work is done to integrate the two systems (Sun *et al.*, 2015). As suggested by many information systems (IS) studies, traditional BI systems pose major challenges that can limit the type and amount of data that an organisation can process (Knabke & Olbrich, 2011; McGlothlin *et al.*, 2017; Surbakti *et al.*, 2019). On the other hand, big data analytics has shortcomings such as skillset shortage, undefined data processes and quality standards, as well as the introduction of increased data security risks (Cai & Zhu, 2015; Taylor-Sakyi, 2016). The need to investigate guidelines and practices for creating a DD-DSS that combines components from both traditional BI and big data analytic systems can therefore not be over-emphasised. The purpose of this study is to investigate a framework that can be used to create a hybrid DD-DSS that effectively integrates traditional BI and big data analytic system components.

1.3 Purpose of the Study

Limited knowledge of and guidelines for the way in which data-driven organisations can create a hybrid DD-DSS are challenging. Organisations have unpleasant experiences when trying to adopt big data analytics because they lack information on the creation of a hybrid DD-DSS that is composed of components from both traditional BI and big data analytic systems (Ruzgas & Dabulyte-Bagdonaviviene, 2017; Salinas & Lemus, 2017). The purpose of this study is to investigate a framework for creating a hybrid DD-DSS; that is, a DD-DSS that contains components from both traditional BI and big data analytic systems.

The following objectives underpin the purpose of this study:

- *To explore what constitutes a data-driven organisation and how data improves the decision-making of a data-driven organisation.*
- *To understand the architectural components of both traditional BI and big data analytic systems.*
- *To propose a framework for creating a hybrid DD-DSS through understanding the common challenges and organisations' perceptions of the integration of traditional BI and big data analytic systems.*

1.4 Research Questions

The emergence of big data demands that data-driven organisations implement effective DD-DSS, which focus on fast information delivery to organisations' decision makers. However, the limitations of both traditional BI systems and big data analytics require organisations to reconsider their big data adoption strategies. Based on the objectives of this study, the following research question guides the investigation and fulfilment of these research objectives:

What are the core elements of a conceptual framework that can be used to design a hybrid data-driven decision-support system that combines traditional business intelligence and big data analytics?

The following research sub-questions are formulated to assist in the discovery of information that will answer the main research question of this study:

- *What is meant by a data-driven organisation and data-driven decision-making?*

- *What constitutes traditional business intelligence and big data analytics?*
- *What are the big data and business intelligence adoption and implementation challenges encountered by organisations?*

1.5 Research Strategy

A detailed literature review was conducted to investigate the existing academic publications that are focused on BI, big data, and big data analytics. During the literature review, attention was paid to identifying the common challenges currently experienced by organisations that are trying to adopt big data analytics while still maintaining traditional BI systems. The initial set of definitions of concepts related to traditional BI and big data analytics, as well as challenges linked to these systems, were identified from the literature survey.

The study employed an interpretive philosophy with a single case study design and involved employees from one of the leading insurance companies in South Africa. Established in 1845, the insurance company provides financial solutions to individuals, businesses of all sizes, corporates and institutions across the African continent and other developing countries outside Africa. The reason for choosing this insurance company is that it is currently in the process of adopting big data analytics with the aim of improving its data-driven culture. The company runs multiple BI systems, with each company division having its own set of BI systems. As part of big data analytics adoption, the company is striving towards consolidating all BI systems into a unified effective system. Because of the complexity of big data analytics and traditional BI systems, the company has been experiencing some challenges in its journey of big data adoption. This makes it a relevant setting for this research study.

Qualitative semi-structured interviews were used to collect data from a selected number of research participants. The analysis of qualitative data can take longer than that of quantitative data because of the complexity and the nature of qualitative data; therefore, the sample size for qualitative research is normally smaller than for quantitative research (Coyne, 1997). Therefore, 18 research participants were involved in the study in order to gather adequate insight to assist in answering the research questions of the study.

Qualitative data analysis was done to interpret, understand and explain research participants' feedback. Inductive data analysis is the approach that was used to derive insights that answer the research questions of this study. This approach was chosen because of its flexibility and ability to allow the researcher to address ideas that emerged during data collection (Fusch & Ness, 2015).

1.6 Assumptions

Particular elements and aspects of the study were required to conduct this research, but these could not be guaranteed or proved. This section highlights the assumptions concerning such elements. The study was conducted under these assumptions:

- It is believed that participants who were interested in taking part in the study had a sincere interest in the study and no other motives were involved, for example building a social relationship with the researcher.
- The researcher assumes that all participants have a basic understanding of big data and how it differs from traditional data.
- It is also assumed that participants answered questions honestly.

1.7 Limitations

This section highlights some constraints that may affect the value of the study and are beyond the researcher's control. To avoid any possible misunderstanding and misinterpretation of the study results, these research study limitations are to be noted:

- The study employed a case study research design and was conducted in one South African insurance company; therefore, the findings of this study cannot be generalised across the South African insurance industry.
- An interpretive research paradigm was assumed by the study. Understanding and definitions of context-specific concepts may differ from one individual to another. Therefore, it is worth noting that definitions and perceptions given by the selected research participants might not reflect the perceptions of all the employees who work in that insurance company.

1.8 Outline of the Study

This study is divided into six parts and each part contains one or more chapters (Figure 1). Part I contains the introductory chapter of the study and includes the background information about the study, problem statement and research questions.

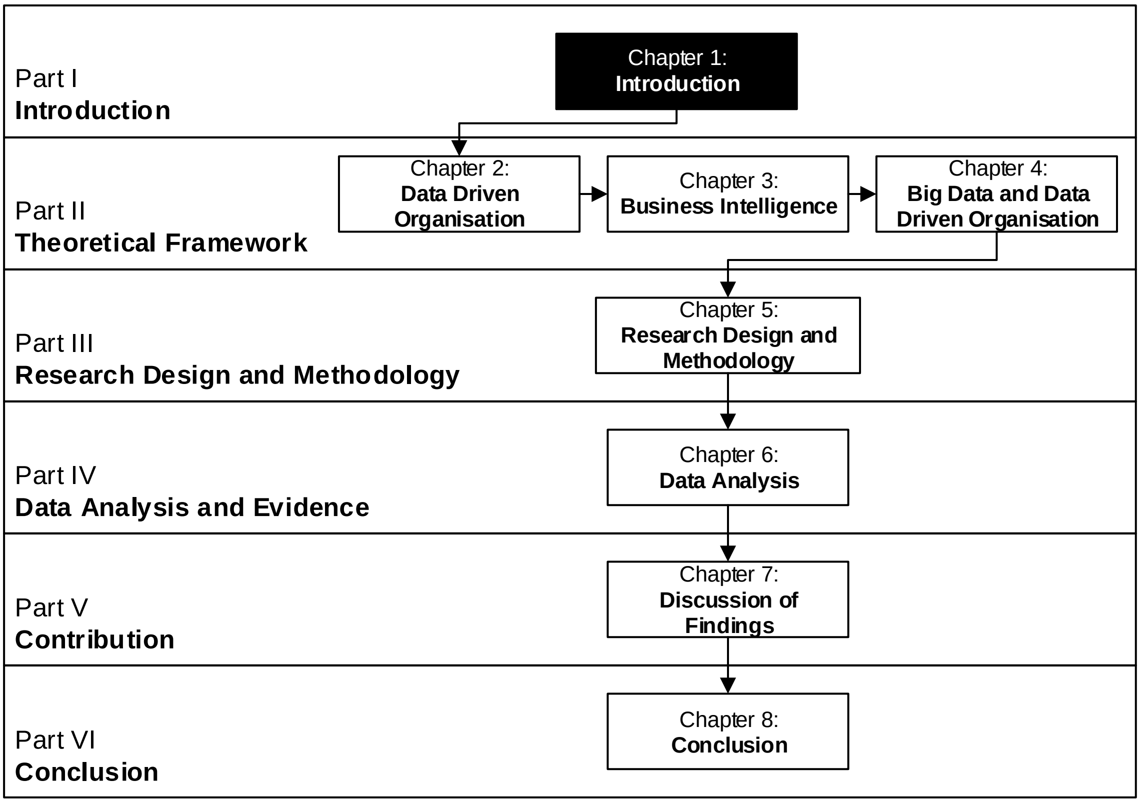

Figure 1 – Part I outline

Part II contains the research literature study. This part has three chapters (Chapters 2, 3, and 4). Chapter 2 describes a data-driven organisation and data-driven decision-making. Chapter 3 is focused on discussing the details of the components of traditional BI systems. Critical success factors and challenges of traditional BI systems are also discussed in Chapter 3. Chapter 4 discusses the relationship between big data and data-driven organisations. All the concepts and aspects of big data that assist a data-driven organisation to make decisions effectively are discussed, in addition to challenges that organisations encounter while adopting big data analytic systems.

Part III contains Chapter 5, which details the methodology and the research design of the study. The chapter includes an overview of commonly used IS, research paradigms and methodologies. Then the details of the methodology and design of this study are described.

Part IV discusses the details of the data analysis and findings of the study. This part of the study contains Chapter 6, whose purpose is to describe the method through which the collected data were analysed, as well as detailing the data analysis findings.

Part V contains Chapter 7, which discusses the contribution of the study. The details of how the findings of the study fulfil the objectives of the study are described in this chapter. The suggested framework for the creation of a hybrid DD-DSS is also described in Chapter 7.

The last part of the study (Part VI) includes the conclusion (Chapter 8) of the study. The summary of the whole study is given in addition to how each of the research questions were answered from the study findings. Summaries of both the practical and theoretical contributions of the study are also included in Chapter 8.

Part II – Theoretical Development

Part II (Figure 2) of the study contains Chapters 2, 3 and 4. This part of the study describes a theoretical framework for a data-driven organisation (Chapter 2), BI (Chapter 3), and big data analytics (Chapter 4).

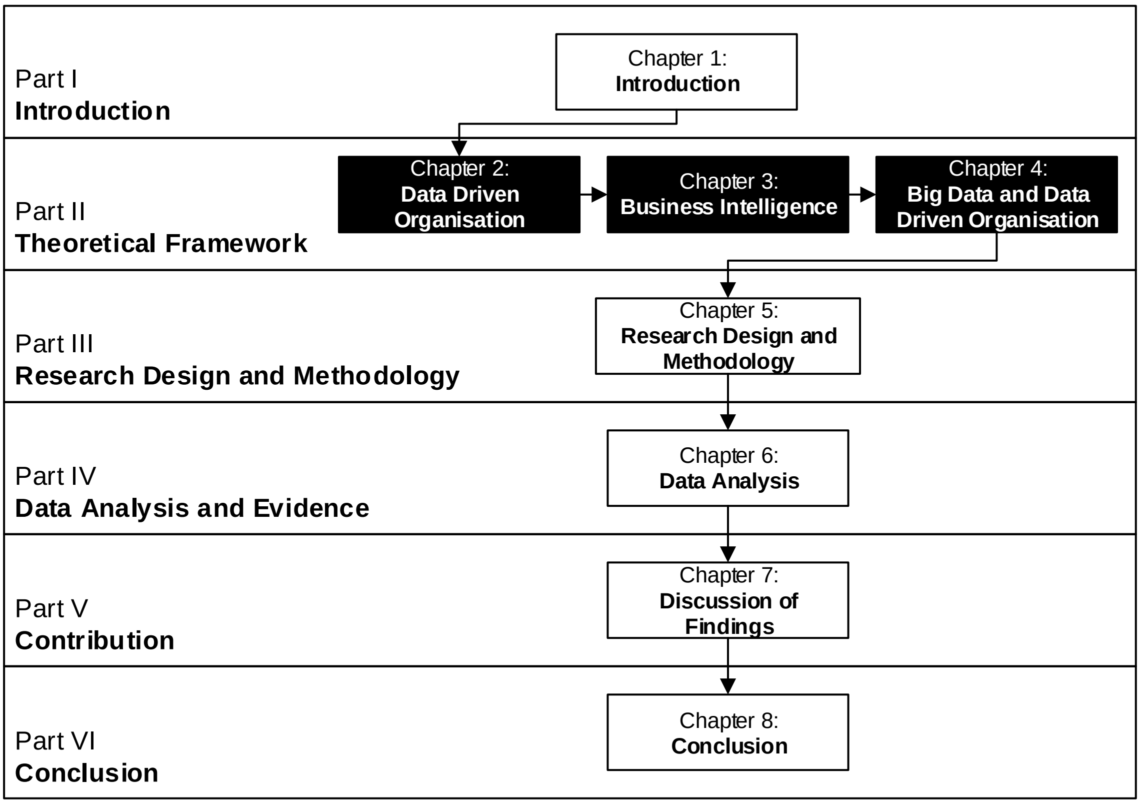

Figure 2 – Part II outline

Chapter 2: Data-driven Organisation

2.1 Introduction

Organisations are forced to change their business model and initiate new product and service offerings because of emerging disruptive technologies. The change in the way organisations do business is not a choice if they want to remain competitive (Smuts & Juleka, 2018); acquisition of advanced technological tools is required to support almost all business changes. The increased amount of data that results from the digitisation of business operations is one of the disrupting forces that requires organisations to enhance their ability to access these data (Hawking & Sellitto, 2010; Ranjan, 2008; Wixom & Watson, 2010). The ability to source and use these data effectively is key to organisations' success. As such, many organisations from across different industries are enhancing their technological systems to ensure that they are able to absorb as much data as possible for use in their decision-making process. Data-driven organisations rely on data to make decisions, which contributes to the success of their business (Brynjolfsson & McElheran, 2016).

This chapter explores the principles of a data-driven organisation. To elevate understanding of the value of data-driven decision-making, intuitive and data-based decision-making techniques are discussed. The data, information, knowledge, and wisdom (DIKW) pyramid is used explain the transformation of data into business knowledge and wisdom. The value of BI in a data-driven organisation is discussed to explore the value of DD-DSS in the success of an organisation's data-driven culture. Common challenges experienced by data-driven organisations as well as critical success factors for adopting a data-driven culture are discussed in the closing sections of the chapter.

A summary of the contents of this chapter is outlined in Table 2.

Table 2 – Chapter 2 outline

Section	Section Description	Sub-section	Sub-section Description
2.1	Introduction		
2.2	Data-driven Culture		
2.3	Data-driven Decision-making	2.3.1	Intuition-driven decision-making
		2.3.2	Data-based decision-making
2.4	Data Analytics and Data-driven Organisation	2.4.1	Data-driven culture skills and competencies
		2.4.2	Data-driven culture processes
2.5	Data-driven Organisational Challenges		
2.6	Critical Success Factors for a Successful Data-driven Organisation		
2.7	Chapter Summary		

2.2 Data-driven Culture

Data-driven organisations use data to make decisions with the aim of improving their performance. These organisations put data-based systems at the centre of their strategic planning (Todorova & Hoeben, 2016; Watson, 2016). The use of statistical and mathematics models to exploit data is key to the success of many data-driven organisations. A data-driven organisation is distinguished from any other organisation by its data-driven culture, which requires executive management of an organisation to commit and support a data analytics initiative (Wang *et al.*, 2019). Watson (2016) suggests that the value of data in the organisation should be demonstrated clearly so that executive management can be willing to invest in data-based initiatives.

Senior management support is key to the success of a data-driven culture (Watson, 2016). However, every employee within an organisation should be willing to be part of a data-driven culture, because in most cases data are handled by most of the organisation's employees. Therefore, all employees should be aware of the importance of data and the reason why data are to be treated with caution. To achieve this, data-driven organisations should embrace the culture of data transparency where access to data is fairly managed and everyone is aware of what data are available for use within an organisation (Provost & Fawcett, 2013). To supplement this argument, Provost and Fawcett (2013) suggest that employees who possess technical skills (such as developers) should work closely with users or data consumers to ensure that everyone is responsibly involved and understands the value of data analytics.

Organisations that use data to derive all decisions are perceived to be more productive and profitable than those that use technology just to support their business operations (Hedgebeth, 2007; Provost & Fawcett, 2013). Data provide business with knowledge that is required to measure the performance of the organisation against its vision. Problems and opportunities can easily be identified through data analytics. Opportunities derived from the effective use of data can range from understanding of business operations to identifying effective customer communication strategies (Wixom & Watson, 2010). The wide use of emerging technological devices by customers introduces opportunities and challenges with regard to organisations' ability to collect such data and use it for effective decision-making. Organisations strive to become data-driven because they need to enhance their customer and product knowledge and these business dimensions are crucial to their success (Nonaka & Takeuchi, 1995; Todorova & Hoeben, 2016). Figure 3 depicts a framework for the culture of data-driven decision-making as suggested by Watson (2016), and the description of different framework components is detailed next.

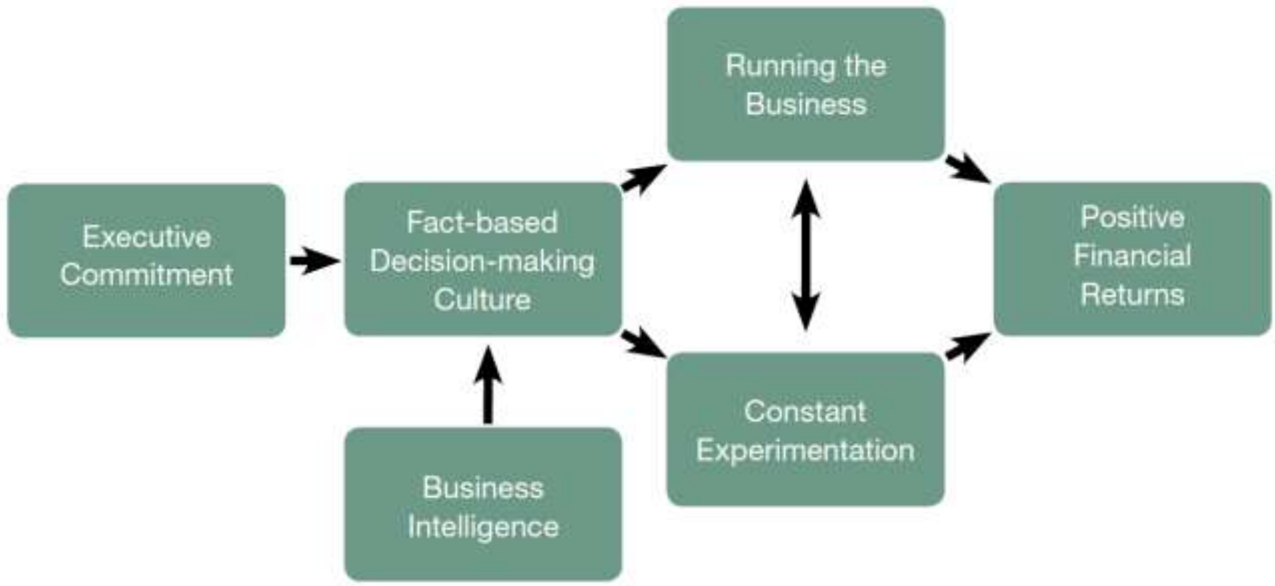

Figure 3 – Framework for a fact-based decision-making culture (Watson, 2016)

As shown in Figure 3, executive managers' responsibility is to define how the organisation should operate and in the case of the adoption of a data-driven or fact-based culture, they can enforce the organisation's data-related processes and governance practices. However, they need to be committed to this transformation. BI deals with all the processing and storage of data, in addition to creating reports. The entire business of a data-driven organisation is run based on reports, dashboards, scorecards and other fact-based measures. Constant experimentation is used to test what works best and what needs to be changed so that the business can be operated optimally. Effective decisions and optimal business operations result in positive financial returns.

A data-driven culture means all decisions of an organisation are based on insights extracted from data. The success of a data-driven organisation is underpinned by the support of its senior management as well as all employees who interact with data in their day-to-day operations. To explore the details of a data-driven organisation, the next section compares the benefits of data-driven decision-making over other decision-making techniques and methods.

2.3 Data-driven Decision-making

Increased interest in data shown by organisations has been the highlight across many industries. The use of data in decision-making and ensuring that these decisions contribute to the success of the organisation have been focus points of many organisations. The emergence of big data has given organisations opportunities to revise their ability to exploit data through the use of the emerging data-processing tools. All these endeavours are underpinned by the aim to use data as a driver of all organisations' decisions (Wang *et al.*, 2019). Many scholars have conducted studies that agree with these statements and suggest that the adoption of a data-driven decision-making culture improves the performance of the organisation (Brynjolfsson & McElheran, 2016; Engelbrecht *et al.*, 2016; Provost & Fawcett, 2013).

Traditional decision-making has been suggested not to be effective anymore; data-driven decision-making is taking over (Wang *et al.*, 2019). There are always questions about what constitutes traditional decision-making and Wang *et al*. (2019) suggest that traditional decision makers use intuition in their decision-making, whereas in data-driven decision-making, every decision made is underpinned by data insights. The details of both data-driven and intuition-based decision-making techniques are discussed next.

2.3.1 Intuition-driven decision-making

In some organisations the management believes that intuition overturns data analytics. In such organisations, decisions are made by people who either rely on their gut feelings or their past individual experiences (Erez & Grant, 2014). Feelings resulting from thoughts that are processed automatically without conscious awareness drive intuition (Vanlommel *et al.*, 2017). Vanlommel *et al*. (2017) and Wang *et al*. (2019) suggest that intuition is instinctive knowing without any logic and evidence to support the thinking. As depicted in Figure 4, the experiences and past causations are the fundamental interruptions destruction that lead to the avoidance of data when decisions are made. Factors that may have led to effective decisions may have different implications after some time and such factors might be outdated. Therefore, intuition-based decision-making might not reflect the reality.

The danger of an intuition-based decision is that the decision must feel good to the decision makers because it is based on the feelings and outcomes of their previous experience. Such feelings might not be good for the organisation and are prone to biasness. Another interesting point about intuition-based decisions is that data can be used only to validate and support the decision. In this case, data are not used as a foundation for a decision, but only as a validation factor. Vanlommel *et al.* (2017) call this a “confirmation bias”. Figure 4 shows the interrelationship between different factors that contribute to intuitive decision-making.

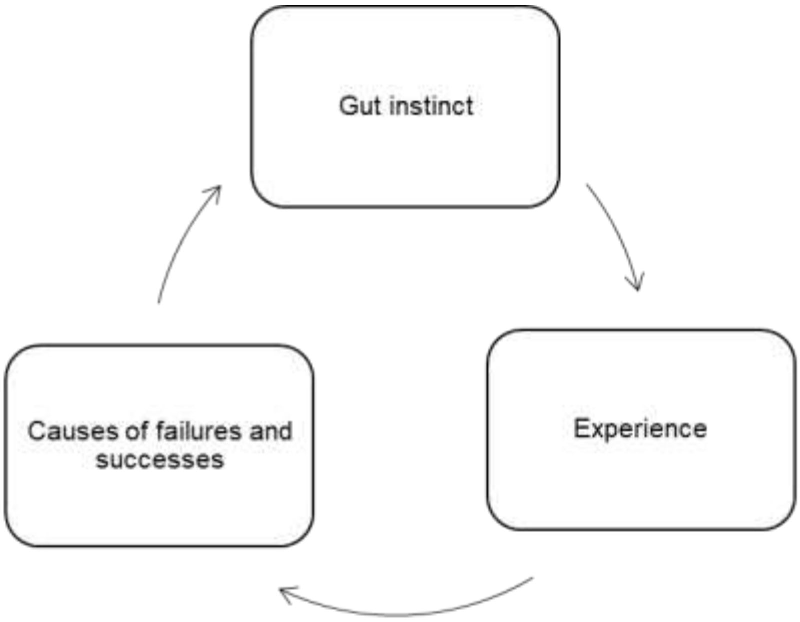

Figure 4 – Main drivers of intuition-driven decision-making (Wang *et al.*, 2019)

Historic experiences that might have led to the success or failure of any business event are used as an input to the decisions that result from gut instincts. It is not practical to base organisations' decisions on gut feelings because such a technique is prone to errors and can only work in favour of individuals, as opposed to the benefit of the entire organisation (Hedgebeth, 2007).

2.3.2 Data-based decision-making

While an intuition-based decision-making culture ignores the results of data and analytics, data-based decision-making uses information that is a result of processed data. The use of data in organisational decision-making ensures that analytics underpins all the decisions made in an organisation (Engelbrecht *et al.*, 2016; Wixom & Watson, 2010). The quality of data that feed the analytics models determines the effectiveness of the decisions made on the basis of these data. Organisations that are embracing data-driven decision-making put data quality measures in place to ensure the reliability of their analytics results. The issue of data quality is always suggested to be one of the major challenges with which organisations that embrace data-driven decision-making struggle (Lee & Kang, 2015).

No data can tell a universal story; all data collected should be used for a specific need and even if good quality data are used, these will not yield good results if not used with caution (Nasser & Tariq, 2015). The information derived from data can be used in a wrong context if data usage is not determined by a business need and a decision requirement. The adoption of a data-based decision-making culture comes with some challenges (Hawking & Sellitto, 2010; Sun *et al.*, 2015). For example, there is a general belief that data-processing systems and tools are expensive. The cost of data-based decision-making can be strongly related to technology that an organisation needs to adopt to acquire data and to ensure speedy transformation of data into information (Knabke & Olbrich, 2011). All data analytics results will lead to effective information if all relevant stakeholders who need access to this information are able to utilise the information (Smuts *et al.*, 2009). Therefore, data-driven information is not useful if the relevant stakeholders who need to use it to drive decisions are unable to access it.

The DIKW pyramid is a common representation used in information science to show how data are converted into wisdom. This pyramid is used in this chapter to describe how business wisdom informs better decision-making of an organisation. Different levels of the transformation of data into knowledge and wisdom are summarised in Figure 5.

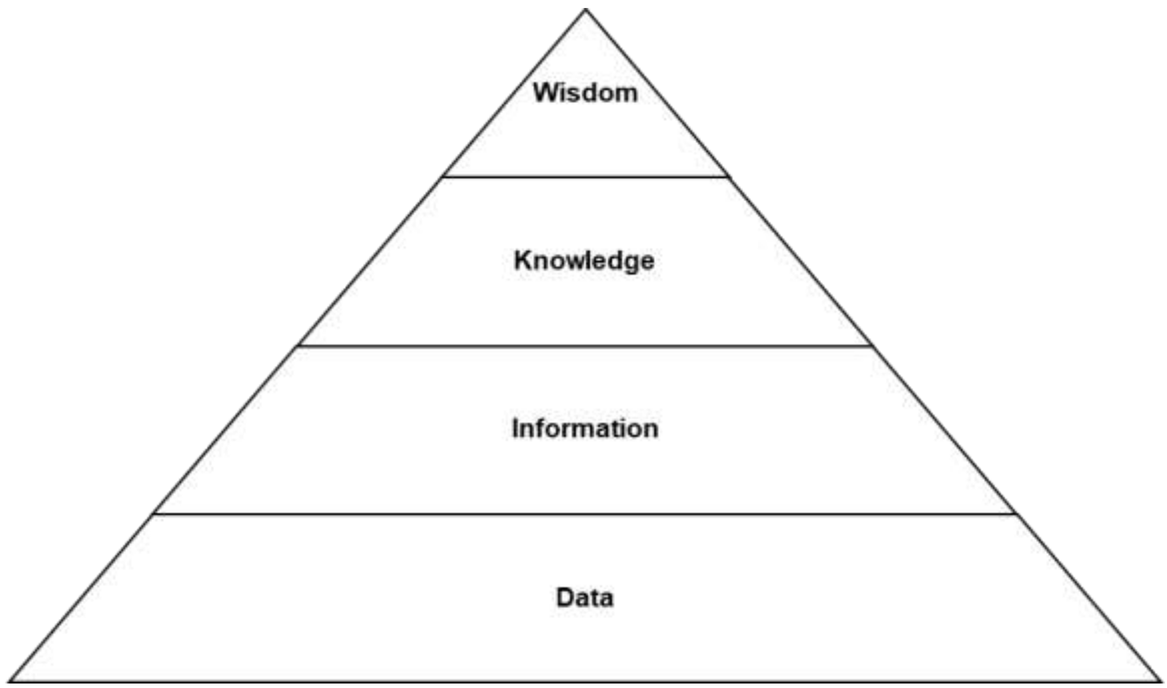

Figure 5 – The DIKW pyramid (Frické, 2009)

At the base of the DIKW pyramid are data that are properties and symbols representing a business object (Frické, 2009). A business object can be either a product or a customer. For example, a customer has a name, location and age; these properties of a customer are data. Organisations capture and store data using forms and RDBMS. In their raw form, data are not usable because they need to be interpreted to make business sense.

The next level after data is information, which results from the processed data. When data are processed they become meaningful (Frické, 2009; Watson, 2016). For example, a marketing director might require information about the location where most of the customers who bought a specific product reside so that campaigns related to that product can be run in that same location. The locations (data) of all customers can be analysed (or processed) to come up with the answer to such a business question. Using this example, the information that is derived from data may be the statement, "… *the majority of customers who bought the product are from Pretoria* …" Information is used to answer questions beginning with "how many", "where", "who", and "what" (Ackoff, 1989). To enhance information delivery to business, organisations normally store information in a database (Winter, 2001). For example, the top ten selling products per week can readily be made available through processing data and storing the results in a database to avoid the cost of back-and-forth data-processing with the aim of fulfilling common or predictable business queries and requirements.

The conversion of information into business knowledge is one of the challenging experiences in the DIKW lifecycle because knowledge results from the iterative use of information with the aim of attaining the "know-how" state (Frické, 2009; Nonaka & Takeuchi, 1995). The trial and error use of information to come up with the skill of knowing how to operate a business effectively constitutes the process of converting information into knowledge. Knowledge is the skill that allows the organisation and its decision makers to know when and how to take action when a specific scenario occurs. For example, the instruction to decommission all products whose annual sales revenue is less than 20% of the actual target is a result of applied knowledge.

The ability that enables a decision maker to know how to make effective decisions in different business situations requires good business knowledge. Knowledge is transformed into wisdom that results from the successful application of knowledge. Intelligence and knowledge are normally used in relation to one another. When comparing these two concepts, Ackoff (1989) considers intelligence to be related to increasing efficiency, while wisdom is related to increased effectiveness.

2.4 Data Analytics and Data-driven Organisation

Organisations' increased interest in using data to drive business decisions has resulted in more organisations ensuring that their data analytics systems are fit for purpose. Many organisations are adopting and shifting their focus towards using data analytics systems to drive their strategic objectives and all data-driven organisations are focusing their interest on ensuring that their BI systems are at the centre of their business strategic planning (Wixom & Watson, 2010). Data-driven organisations use data analytics systems to exploit data that enable them to understand their customers and to communicate the value of their business to different stakeholders, including investors (Ranjan, 2008). Data-driven organisations incorporate data and information management into their day-to-day activities and this ensures that data analytics systems not only add value to customers, but are also used to assess and improve the working conditions of the organisation (Brichni *et al.*, 2015). Data-driven organisations believe that their success is critically determined by the ability of their data analytics systems to add value to their businesses through facilitating and speeding up the organisation's business information and knowledge discovery process (Bhanot, 2017; Smuts *et al.*, 2009).

Data analytic systems deliver knowledge that an organisation can use to understand its market. Such a discovery enables many organisations to perform better than their competitors (Smuts *et al.*, 2009). The determiner of a successful data-driven organisation is its ability to apply analytical techniques to its data and enhance the transformation of data into information (Merendino *et al.*, 2018). Evidence-based decision relies on organisational data being used as the only foundation of business knowledge and it is through capable data analytic systems that such data are converted into knowledge (Balakrishnan & Rahul, 2018; Hedgebeth, 2007). To implement an effective data-driven culture, organisations are defining processes and governance standards that enforce and improve the perceived value of data.

The emergence of big data has introduced new job roles, for example data engineers and data scientists. Data science is regarded as the most in-demand skill in IT. Even though the definition of the role of a data scientist is not universal, many organisations are hiring data scientists to assist in discovering value from organisations' data (Provost & Fawcett, 2013). Data preparation and data-processing activities, which are considered to be the activities on which most data scientists spend their time (Provost & Fawcett, 2013), are fundamental to the maturity of organisations' modern data analytic systems. Data-driven organisations need to improve their data-related processes to ensure on-demand data analytics and knowledge discovery. Data analytics is worthless without a proper data-driven culture being prioritised by the organisation's senior management to ensure that data are valued by all the employees (Lee, 2017).

To ensure the success of a data-driven organisation through effective business decision-making, organisations need to have all the required capabilities to improve data-processing and data quality. The next sections discuss the culture and processes that organisations put in place to strengthen their on-demand data analytics and knowledge discovery.

2.4.1 Data-driven culture skills and competencies

A data-driven culture requires support from the executive management of the organisation, who need to invest in skills and processes that support the conversion of data into information, knowledge and wisdom (Nonaka & Toyama, 2005; Todorova & Hoeben, 2016). A data-driven culture requires changes not only from technical employees, but throughout the entire organisation, because the perception of data has to change throughout the organisation. Data analytics and data-driven decision-making are intertwined and neither of these two concepts can be treated independently from the other (Provost & Fawcett, 2013). With the emergence of big data, data science is considered a fundamental skill requirement for all data-driven organisations. Data science is a data specialisation field that encompasses the use of a set of processes and techniques to extract business value from big data using the emerging technological tools and principles

(Larson & Chang, 2016; Provost & Fawcett, 2013). Data science competencies are critical in supporting the successful adoption of a data-driven culture.

The fundamentals studied in the field of statistics form the core part of data science; therefore, knowledge of statistics is required for data scientists. Data analytic thinking is a skill that not only data scientists need to acquire; the entire organisation has to appreciate and embrace data analytics (Merendino *et al.*, 2018). Senior and executive management can only invest in data-related innovations if they understand how to measure their success and are able to identify other opportunities from such initiatives. Team leaders and project managers need basic data analytic skills because they interact with data scientists and other data specialists in the organisation (Provost & Fawcett, 2013). Although it is reasonable to acknowledge that the level of knowledge and skills among different employee levels and roles cannot be the same, the appreciation of data as a valuable asset for organisational decisions should not be compromised. Therefore, data analytics thinking is key to the culture of a data-driven organisation.

Even though no universally acknowledged personnel skills and knowledge have been accepted for supporting a data-driven culture, Table 3 lists some required common knowledge areas and associated skillsets.

Table 3 – Human knowledge and skills (Mikalef *et al.*, 2018)

Knowledge area	**Skills**
Technical knowledge	Programming, technical infrastructure management, MapReduce, unstructured data management, data collection/integration.
Business knowledge	Business strategy, key performance indicators, business processes, change management.
Relational knowledge	Communication skills, team building.

The success of the adoption of a data-driven culture is highly dependent on the skills of employees and these can be categorised into business, technical and relational skills (Mikalef *et al.*, 2018). Technical skills are required to fulfil roles such as those of data engineers who are required to source, clean and code data from different data sources.

Understanding the relevance of data to a specific business context requires an individual who has a good understanding of business. Project managers need to manage teams where different data-related skills and competencies are utilised (Provost & Fawcett, 2013); in such projects collaboration and communication among different individuals are key to the success of the project (Mikalef *et al.*, 2018). Therefore, relational knowledge is also fundamental to the adoption of a data-driven culture.

2.4.2 Data-driven culture processes

Without proper processes being implemented across the organisation, successful adoption of a data-driven culture is impossible or becomes a challenging experience (Mikalef *et al.*, 2018). Before anything can be done with data, it is important to ensure that data are accurate. A data-driven culture should be accompanied and supported by processes enforcing data accuracy. If data are the basis of every decision, then the importance of data accuracy should be driven by strict data governance and standards (Cai & Zhu, 2015). Data accuracy refers to the extent to which data are truthful and reliable (Todorova & Hoeben, 2016). Organisations' investments in data validation tools should be driven by the ability of these tools to support and improve the accuracy of data.

In addition to data quality processes, data availability and information-sharing processes should be defined to ensure that the relevant stakeholders have access to relevant data at the right time. Free access to data has long been known to be the enabler of effective knowledge creation in an organisation (Nonaka & Takeuchi, 1995), hence data availability processes are core to the successful adoption of a data-driven culture. Data consistency is another factor that determines the quality of data. Data that exist within an organisation have to be consistent so that analytics performed on those data can be of high value (Nasser & Tariq, 2015). Therefore, key supporting processes should be implemented in favour of improved data consistency.

Organisations are striving towards implementing processes ensuring that the definitions of data across different data sources are compatible (Alharthi *et al.*, 2017). Data must be relevant and be used as basis for answering business questions in a specific context. Different organisational units execute different strategic plans that align to the overall vision of the organisation and data relevance plays a major role in ensuring that data acquisition tools only retrieve relevant data required for specific business needs (Cai & Zhu, 2015).

The definition of processes that ensure that data are handled in a manner that supports the requirements of a data-driven culture can improve the adoption of such a culture. Such processes can address some of the common challenges that are linked to the adoption of a data-driven culture. The next section discusses some of the common challenges that data-driven organisations experience in their journey towards implementing a data-driven culture.

2.5 Data-driven Organisational Challenges

It is clear that successful adoption of a data-driven culture increases organisations' business value and business innovation. However, many challenges are linked to the adoption of a data-driven culture and only a few organisations are consequently able to become fully data-driven (Dhar & Mazumdar, 2014). Organisations know the success that results from proper management and use of data but there are still gaps pertaining to how they can realise this value (Bhanot, 2017; Engelbrecht *et al.*, 2016). This suggests that there is a lack of models and frameworks to guide organisations towards becoming fully data-driven. The lack of information about what data are available for organisations to exploit is one of the major challenges hindering organisations' success in becoming data-driven. According to Watson (2016), evidence-based decisions require organisations to be well informed about the available data that can be used and if that is lacking, data might not be utilised to their full potential.

According to Lee (2017) and Merendino *et al.* (2018), lack of support and a strategic transition plan from the executive management is another common challenge experienced by organisations striving towards becoming data-driven. Senior management should invest in processes and tools that support the adoption of a data-driven culture. If they are not playing their part, the adoption of a data-driven culture is more likely to fail.

Big data have introduced even more confusion into the concept of a data-driven organisation because organisations are trying hard to exploit big data and forget about the fundamentals of a data-driven organisation (Nasser & Tariq, 2015; Sun *et al.*, 2015). This has led to lack of knowledge pertaining to the integration of big data systems with traditional or legacy data analytic systems such as BI. Sun *et al.* (2015) suggest that another major challenge is the lack of knowledge regarding whether or not some components of traditional BI systems can be used in collaboration with big data solution components to satisfy the information requirements of a data-driven organisation.

Because of the increased interest in big data and the failure to migrate traditional BI systems successfully into big data analytic systems, organisations end up having to run both traditional BI and big data analytic systems in parallel. The cost of running these systems can be a financial strain for the organisation. Modern job roles such as those of data scientists should play a major role in the improvement of organisations' capability to deal with big data; however, the shortage of qualified data scientists fundamentally contributes to the difficulty of the adoption of relevant data analytics systems that are required to support a data-driven organisation (Surbakti *et al.*, 2019).

The challenges presented in this section suggest that becoming a data-driven organisation is not an easy process. Some scholars have suggested ways through which organisations can increase their chances of successfully becoming data-driven. The following section summarises some of the common factors that can lead to the successful adoption of a data-driven culture.

2.6 Critical Success Factors for a Successful Data-driven Organisation

A number of scholars have identified factors worth considering for organisations that are adopting a data-driven culture. Some studies identify factors in relation to traditional BI systems, while others specifically relate these factors to a data-driven culture that involves big data. A summary of these factors is shown in Table 4.

Table 4 – Data-driven organisations' critical success factors

Success factor	Source
Free access to information for all employees	(Alpar & Schulz, 2016; Nonaka & Takeuchi, 1995)

Success factor	Source
Enhancement of data collection and analysis technologies	(Dhar & Mazumdar, 2014)
Ability to source and access both internal and external data	(Engelbrecht *et al.*, 2016; Tambe *et al.*, 2012)
Executive and senior management support	(Ariyachandra & Watson, 2010; Berndtsson *et al.*, 2018; Hawking & Sellitto, 2010)
A solid understanding of data across the organisation	(Provost & Fawcett, 2013; Taylor-Sakyi, 2016)
Use of IT to support effective decision-making rather than to just produce data	(Hawking & Sellitto, 2010)
Application of measures to ensure data governance, data quality and data consistency	(Berndtsson *et al.*, 2018; Todorova & Hoeben, 2016)
Adoption of suitable tools to visualise and communicate information to decision makers	(Alpar & Schulz, 2016; Hawking & Sellitto, 2010)

Free access to information is a requirement to ensure that the correct people can access the correct data at the right time. It does not necessarily mean everyone should have access to all data, because that would be a data security risk. Most of these factors are directly related to the suggested challenges (section 2.5), because ensuring that measures are put in place to address those challenges can improve the chances of the successful adoption of a data-driven culture. Senior management support is key to ensuring adequate strategic and financial support for data-driven projects (Ariyachandra & Watson, 2010; Berndtsson *et al.*, 2018; Hawking & Sellitto, 2010).

The existing tools that are used to gather and analyse data should be enhanced to ensure that they can cope with organisations' data-processing requirements. Data quality determines the success of a data-driven organisation by enabling the organisation to make effective data-driven decisions. Therefore, data quality standards are essential to the successful adoption of a data-driven culture. Data that are pulled from different sources through the use of advanced data-processing tools should be assessed against data quality standards defined by the organisation (Berndtsson *et al.*, 2018; Todorova & Hoeben, 2016). Data sources can exist either within or outside the organisation; the technological tools used to process data should be able to extract external data sources (Engelbrecht *et al.*, 2016; Tambe *et al.*, 2012).

2.7 Chapter Summary

The increased amount of data generated by digital technologies have led to organisations wanting to use these data to gain value from them. This has forced organisations to change the way they operate. This change involves digitisation of business operations as well as transforming the culture of the organisation towards a data-driven one. To adopt a data-driven culture successfully, organisations embrace the philosophy of data transparency, which demands access to data to be fairly managed and everyone to be aware of what data are available for use in an organisation.

The purpose of this chapter was to explore important aspects of a data-driven organisation and how data analytic systems can enable an effective data-driven culture. The chapter opened with a detailed discussion of a data-driven culture (section 2.2). The relationship between a data-driven culture and data-driven decision-making was explored in section 2.3, where the difference between intuitive and data-based decision-making techniques was described. Many studies have suggested that organisations that use data as a basis for their decision-making are more successful than those that use other decision-making techniques such as intuition-based decision-making.

A data-driven organisation can only succeed if its data analytics systems support its data-driven culture effectively. The details of how data analytics systems assist and improve the success of a data-driven organisation were discussed in section 2.4. This section also discussed personnel aspects of a data-driven organisation, in addition to processes that support a data-driven culture.

Many organisations are interested in becoming data-driven; however, many challenges are experienced by organisations on the journey of becoming data-driven. Section 2.5 discussed some of the common challenges related to the adoption of a data-driven decision-making culture. To address these challenges, a number of scholars have suggested some critical success factors of which organisations need to be aware. A summary of these critical success factors for adopting a data-driven organisation strategy was explored in section 2.6.

Chapter 3: Business Intelligence

3.1 Introduction

In response to the increased generation of data and the proven value of data-driven decision-making, organisations are developing advanced capabilities that can enable them to acquire and process multiple formats of data. Effective data-driven decision-making is a challenging exercise to organisations because it requires proper design and application of data-processing systems. Challenges related to data analytic systems may affect the performance of organisations because their business performance depends heavily on the right decisions being made by the right people at the right time (Gupta *et al.*, 2015; Negash, 2004; Richards *et al.*, 2019).

BI has been widely used in many corporate sectors such as banking, manufacturing and medical institutions to respond to the needs of data-driven decision-making. However, the evolving needs for good quality information has resulted in many organisations experiencing challenges related to ensuring that their BI systems are competent to deliver information of acceptable quality (Brichni *et al.*, 2015). Organisations adopt data analytics solutions with the aim of deriving data-based business insights, which are used as knowledge base for decision-making (Berndtsson *et al.*, 2018). Effective BI systems offer organisations the ability to manage their information and share the information effectively with decision makers.

Different scholars give different definitions of BI. Most definitions relate BI to the collection of software tools and processes whose aim is to transform data into information and knowledge that inform business decision-making. The following are some of the definitions of BI:

- BI is the process of capturing, accessing, understanding, analysing and converting data, which are one of the most valuable assets of an organisation, into actionable information in order to improve business performance (Negash, 2004; Ong *et al.*, 2011). BI is a capability that provides businesses with tools and methods that enable them to easily access and manage information that supports business users in making effective decisions (Brichni *et al.*, 2015).

- According to Gupta *et al.* (2015), BI is used to describe a collection of "applications, technologies, architectures, and processes for gathering, storing, accessing, and analysing operational data to provide business users with timely competitive information to enable better insights for operational and strategic decision making."

These definitions regard BI as a system that supports business decision-making; BI is therefore fundamentally a decision-support system.

The purpose of this chapter is to explore the concept of BI and discuss its components. To understand how different system components contribute to the overall functionality of a BI system, the common architectural components of a BI system are discussed (section 3.2). In section 3.3, the details of a data warehouse and ETL components are also explored to explain why these components are key to the effectiveness of a BI system. The value of BI and its related processes is discussed to enhance understanding of the relationship between BI and data-driven decision-making (section 3.4). The chapter concludes with a discussion of challenges and critical success factors for designing and operating effective BI systems.

A summary of the contents of this chapter is outlined in Table 5.

Table 5 – Chapter 3 outline

Section	Section description	Sub-section	Sub-section description
3.1	Introduction		
3.2	Business Intelligence Architecture		
3.3	Data Warehouse and ETL		
3.4	Value of Business Intelligence in a Data-driven Organisation		
3.5	Typical Business Intelligence Process	3.5.1	Business intelligence measurement
		3.5.2	Business intelligence capabilities

Section	Section description	Sub-section	Sub-section description
3.6	Business Intelligence Challenges		
3.7	Business Intelligence Critical Success Factors		
3.8	Chapter Summary		

3.2 Business Intelligence Architecture

Effective BI is essential to the success of a data-driven organisation and the architectural landscape of different BI system components may support data-driven business objectives. The architecture of a typical BI system is summarised in Figure 6. In many organisations, data are sourced from multiple operational systems and these data sources have different data formats. These heterogeneous data systems are designed to enable business operational activities as defined by the business goals of the organisation. A collection of source systems is represented by a *data source layer* in Figure 6. It is essential that organisations consolidate and analyse all data that reside in these systems (Mukherjee & Kar, 2017; Surbakti *et al.*, 2019). The data coming from these disparate sources need to be stored in a format that is standard and easily accessible to the analytics systems and users. The *ETL layer* is responsible for standardising the formats of data before these are stored in a repository that is optimised for analytics and reporting.

It is difficult to consolidate data analytics results derived from multiple data sources and organisations therefore need to have central storage for the consolidated data that are pulled from multiple data sources (Ariyachandra & Watson, 2010). This central storage of all the available organisational data is referred to as a *data warehouse* (Figure 6). Studies suggest that a data warehouse and ETL are the fundamental components of a BI system because they transform and store data in a format and grain optimised for analytics and reporting (Salinas & Lemus, 2017; Santoso, 2017; Sun *et al.*, 2015).

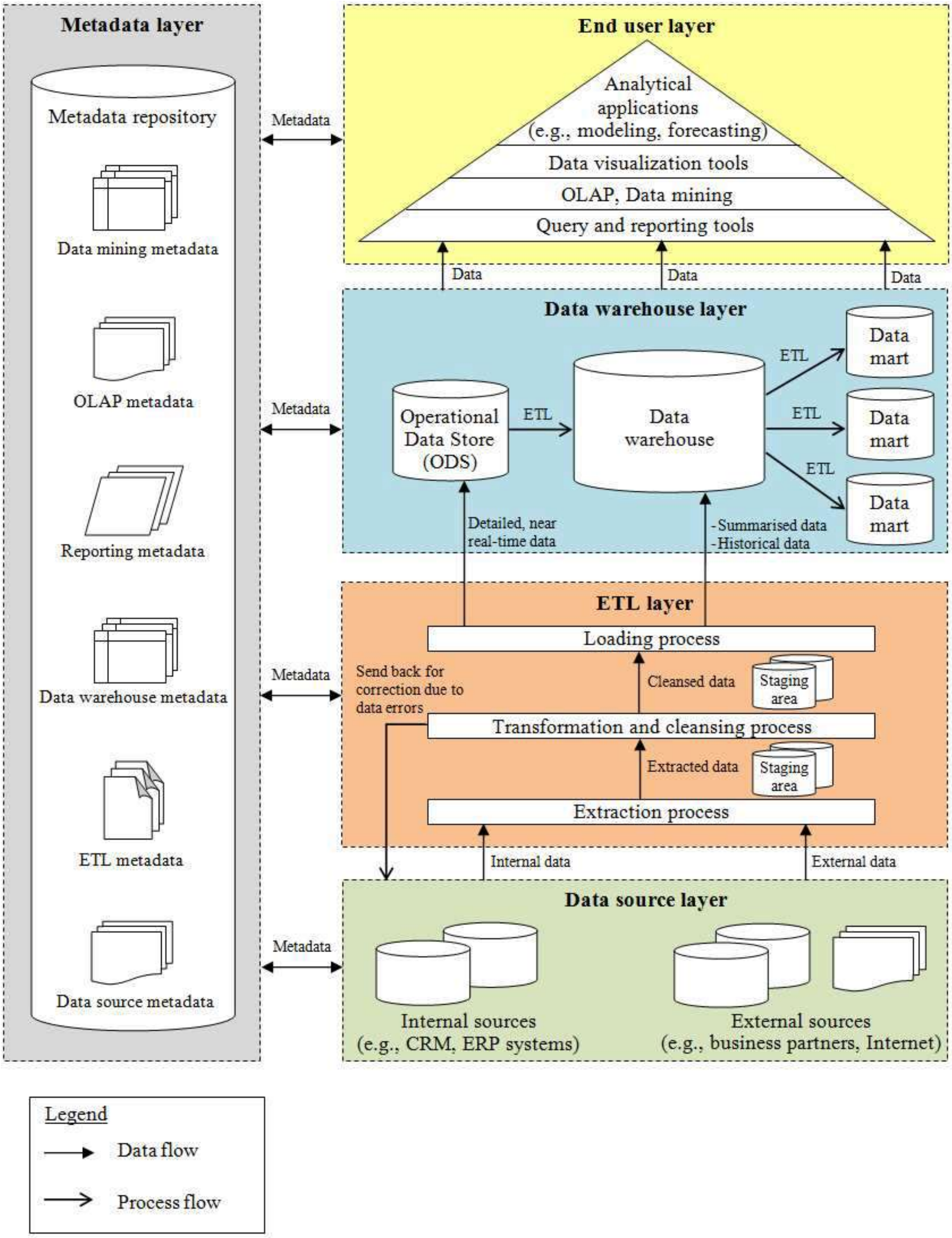

Figure 6 – Typical business intelligence architecture (Ong *et al.*, 2011)

Different data representation tools and techniques are used to visualise information and present it to different information consumers (Ranjan, 2008). Correct and meaningful information visualisation assists a business to plan, define and track the performance of the organisation against its business vision (Ranjan, 2008; Santoso, 2017). Therefore, effective data and information visualisation plays an important role in the interpretation of information. The *end user layer* (Figure 6) of a BI architecture includes a collection of all tools and processes that enable users to access and exploit data. Each layer of a BI architecture must have a *metadata* component that describes how the layer operates and gives guidelines on how each of these layers is to be maintained.

Data insights derived through BI can be used to guide and execute innovative strategies of an organisation and these insights can be used to measure the alignment between business innovations and strategic business goals. Business goals may include plans to improve the working condition of employees, to improve product and customer experience, as well as to strengthen the organisation's data-driven culture (Alpar & Schulz, 2016). The information presented to different information consumers might give knowledge that informs the design of the organisation's operational systems; these systems feed the organisation's BI and its information lifecycle (Brichni *et al.*, 2015).

A data warehouse and its related ETL components form the backbone of a BI system. The next section discusses the details of these two components of a BI system.

3.3 Data Warehouse and ETL

Access to a large amount of data that are stored in operational systems has increasingly become a challenge because of the time spent in trying to explore different data sources with different data layouts (Al-Debei, 2011). To address this problem, organisations have adopted and implemented *data warehouse* systems. A data warehouse is a relational database system that stores all the organisation's data, coming from different systems, in a single data store; these data are stored in a format that is optimised for reporting (Mukherjee & Kar, 2017; Surbakti *et al.*, 2019; Winter, 2001). According to Santoso (2017), the process of pulling differently formatted data from multiple systems and transforming these into a consistent data layout is called *ETL*. It is reasonable to suggest that ETL and the data warehouse are components of a traditional BI system, which work together to transform and optimise data for ease of analysis and reporting. Without effective ETL a data warehouse system cannot be optimal.

If a data warehouse contains a single view of all the organisation's data that are stored in a format that is optimised for analysis and reporting, a data warehouse plays a major role in fulfilling the organisation's data and information-processing needs (Al-Debei, 2011; Devlin, 2010; Olszak & Ziemba, 2012; Santoso, 2017). Several applications can be used to generate reports by pulling and aggregating data that are stored in a data warehouse (McGlothlin *et al.*, 2017). Some of the commonly used data aggregation and reporting applications include Tableau, Microsoft Power BI, and Qlik (Ruzgas & Dabulyte-Bagdonaviviene, 2017). Effective ETL is a fundamental requirement of an optimal data warehouse and an optimal data warehouse is a vital requirement of a BI system (Brichni *et al.*, 2015; McGlothlin *et al.*, 2017; Surbakti *et al.*, 2019).

As suggested by Brichni *et al.* (2015), a data warehouse should have a complete data dictionary. A data dictionary is composed of metadata that describe the details of all data objects and data attributes stored in a data warehouse. Data exploitation and reporting tools can pull data directly from a data warehouse, but the metadata and data dictionary is key to ensuring that these tools exploit good quality data because the metadata contain information about a data warehouse itself. The optimisation of a data warehouse system is highly dependent on the extent to which different components and data objects of a data warehouse and ETL are being documented. Metadata are key to the effectiveness of a data warehouse system.

The exploitation of data from a data warehouse can be optimised through the use of data marts; data marts are slices and subsets of a data warehouse that are designed to service specific information needs (Devlin, 2010). For example, one section of an insurance company may be interested only in claims information, as opposed to those that are interested in sales data. Different data marts, which respectively include claims and sales data, can be created from a data warehouse to serve these distinct information needs. The effectiveness of a traditional BI system is highly dependent on the efficiency of its data warehouse and ETL components. A data warehouse and ETL improve the functionality of the BI system; therefore, these components play a major part in the success of a data-driven organisation. The next section discusses the value that BI adds to the success of a data-driven organisation.

3.4 Value of Business Intelligence in a Data-driven Organisation

For an organisation to be able to service its customers and stay competitive in the market, understanding of the environment in which the organisation operates is key. The success of any organisation depends on how the organisation responds to market changes in a positive manner that improves its service to customers and the community (Engelbrecht *et al.*, 2016; Smuts & Juleka, 2018). Organisations need data to be able to track themselves against their strategic objectives, as well as making decisions about their new strategic developments and amendments. Nowadays, data, information and knowledge are all that organisations need to survive in the market, because information and knowledge derived from data enable organisations to understand their business dynamics and identify improvement opportunities (Dawson & Van Belle, 2013; Olszak & Ziemba, 2012; Smuts *et al.*, 2009).

Organisations normally have multiple operational systems (such as finance, claims, and sales systems), which generate data about organisational products and customers. However, for data to tell a business story and serve as the knowledge base of an organisation, these data need to be put into a system where they are easily accessible. BI systems provide organisations with the capability to source, transform, store and generate reports from data (Negash, 2004; Ranjan, 2008). Organisations not only store and analyse data about their products and customers; they also have access to data about their competitors and such data, when used effectively, can help them to improve their competitive advantage (Brichni *et al.*, 2015). Knowledge management is critical to the success of an organisation (Smuts *et al.*, 2009; Wixom & Watson, 2010) and BI plays a major role in data transformation and the knowledge lifecycle. Therefore, BI enables a data-driven organisation to have access to information that can improve its sustainability.

BI can be used to measure and improve the entire IT landscape of an organisation. According to Brichni *et al.* (2015), an organisation can use BI to assess the ability of its IT landscape applications regularly to support changing market conditions and business requirements. Decisions about decommissioning and adopting new technological solutions to support the organisation can be made based on information and knowledge that are acquired through BI systems. Assessment of IT capability (including BI systems themselves) through data analytics requires an organisation to capture data and analyse information about its IT applications and systems. This is not a common measure in many organisations, because in most cases BI and data analytics are used to make decisions that are not IT-related (Brichni *et al.*, 2015).

IT systems and applications are developed and used by people. Therefore, any change in IT systems is likely to affect the employees of an organisation. Given that every decision of a data-driven organisation is based on data, BI systems can be used to give insight into change management planning and execution (Surbakti *et al.*, 2019). Change management might involve the revision and optimisation of BI processes to ensure that effective BI systems yield quality information for effective decision-making.

The efficiency of a BI system is supported by a number of defined processes whose aim is to ensure that a data-driven organisation can realise the value of its BI systems. The next section describes a typical BI process and how each of the steps in the process can be measured.

3.5 Typical Business Intelligence Process

In order to achieve the vital goal of BI, the organisation should determine the process through which BI projects are operated. The BI process design depends on a number of factors and such factors differ from one organisation to another. These differentiating factors may include the industry or business environment where the organisation operates, the type of data that need to be collected, the infrastructure underlying the storage of data, as well as the number of phases that a BI process has (Lönnqvist & Pirttimäki, 2006). This section uses a BI process model that was suggested by Lönnqvist and Pirttimäki back in 2006 to describe a typical organisational BI process model. This study considered this model because it fairly describes a BI process that includes all the fundamental themes of the data and information lifecycle. The phases of the BI process model are described below.

- ***Phase 1:*** *Identification of data and information needs* – The objective of the first phase in the process is to identify information that is required to solve a specific business problem. It is important that tasks and activities performed in this phase are aligned with specific data and information needs of an organisation (Dawson & Van Belle, 2013). Successful decisions about a specific problem depend on the identification of the correct data and information.

- ***Phase 2:*** *Data and information acquisition* – Once the problem that requires a BI or data analytics solution has been identified and relevant information needs have been defined, the next step is to acquire the required data from the relevant data sources. Data sources may be available within the organisation's database systems or external data might be required to supplement the understanding of the organisation's data (Sun *et al.*, 2015).
- ***Phase 3:*** *Data and information analysis* – Once data have been collected from different sources, this is the phase in which the data are analysed to get insights that can be used by different information consumers to support their distinct analytical needs (Devlin, 2010).
- ***Phase 4:*** *Storage and information utilisation* – The results of data and information analytics need to be stored and presented to people who need to make business decisions. The development of BI products and services is only successful if information outcomes can be stored and communicated to relevant decision makers at the right time (Olszak & Ziemba, 2012). The use of the right technological tools to visualise information essentially improves the understanding of information by information consumers.

BI process models should be assessed and revised appropriately to ensure that they are always up to date. Assessment of each step of the process (section 3.5.1) is critical to the improved value of BI products; that is, information and knowledge artefacts.

3.5.1 Business intelligence measurement

If a business value of BI is to be measured, then processes used to deliver BI products and services should be measured too. BI processes must be assessed and rectified when necessary to ensure that they support effective and successful BI service delivery (Dawson & Van Belle, 2013). To understand why BI and its related processes should be measured, Lönnqvist and Pirttimäki (2006) suggest two reasons. First, organisations measure BI to prove that it is worth investing in BI systems. Decisions on whether or not an organisation should continue investing in BI systems is determined by the value that BI systems add to the organisation. Secondly, BI is measured to ensure that its related processes are effective. For example, if the phase of the BI process that involves data analysis is not effective, then the entire process is likely to be ineffective. BI processes are only effective if the data and information needs of decision makers are satisfied at the right time and the development cost of these services is within budget.

The literature suggests that each phase of a BI process should be measured using a specific measurement metric (Dawson & Van Belle, 2013). In this section, different metrics are summarised and linked to each phase of a BI process model. Table 6 summarises the measures that can be applied in each phase of the process.

Table 6 – BI process measurement (Dawson & Van Belle, 2013)

Phase	**Measure**
Phase 1: Identification of data and information needs.	Efficiency of the identification procedure and the necessity of identified needs.
Phase 2: Data and information acquisition.	Cost and timeliness of data-gathering techniques. Reliability and quality of information acquired.
Phase 3: Data and information analysis.	Accuracy and usefulness of the analysis results. Quality of data analysis outcomes and elimination of irrelevant information.
Phase 4: Storage and information utilisation.	Efficiency and cost of information distribution. Benefits achieved through the use of BI products.

The business necessity of the requirements of data is used to measure data and information requirements. The tools used to extract data from the identified data sources should be effective to ensure that data acquisition is completed within a reasonable time. The accuracy of data analytics results is a suggested measure for the data analytics phase (Yang *et al.*, 2017). Lastly, tools used to store and share information must do so in an effective manner.

3.5.2 Business intelligence capabilities

The operation of BI processes and development of BI products require specific capabilities to exist in the organisation. These capabilities can include personnel competencies, technological tools and other resources that support BI processes. Each phase of a BI process requires different competencies. For example, the identification of information needs requires people who understand business and are able to map business needs to data requirements (Lönnqvist & Pirttimäki, 2006).

Because of increased information demands, it is difficult for BI specialists to cope with all the requests and queries from users. As such, self-service BI is perceived to be the solution to this challenge (Alpar & Schulz, 2016). Self-service BI allows users to analyse data and derive insights without having to consult BI specialists. Nevertheless, users need to acquire essential skills to ensure that they can effectively use an organisation's data analytics and visualisation tools. Systems such as a data warehouse are normally designed to include the capability that allows power users to pull data and create reports themselves (Devlin, 2010). Figure 7 summarises the overview of different levels of self-service BI.

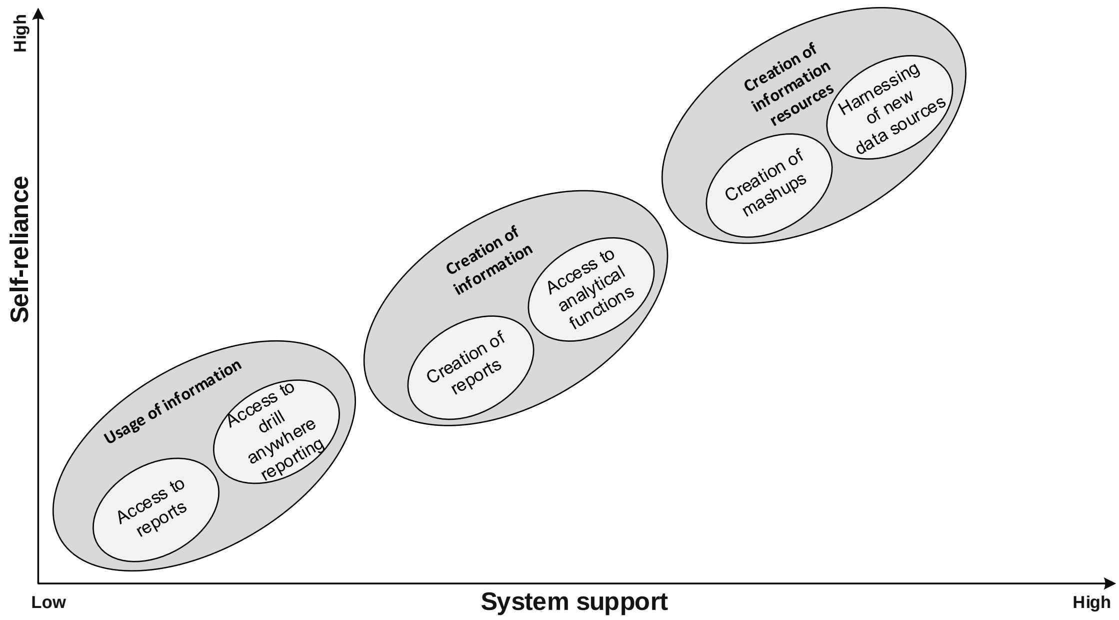

Figure 7 – Levels of self-service BI (Alpar & Schulz, 2016)

It is impossible for users to have the same level of skills and competencies that BI specialists have; however, it is suggested that BI products should be deployed in a state that allows users to perform as many data and information activities as possible. Alpar and Schulz (2016) suggest six levels of activities (Figure 7) that users can perform using a self-service BI system. Each of the levels of self-service BI requires different levels of skills and these skills are dependent on BI technological tools used by the organisation. For the purpose of this paper, a specific toolset used to deliver BI will not be discussed because these differ from one organisation to the other.

BI users should at least be able to access reports containing information they need for business decision-making and they should be skilled enough to view different perspectives and levels of information through drill-downs and drill-ups. The creation of information occurs when users (or power users) are skilled enough to access data from a source (such as a data warehouse) and are able to analyse data and create reports themselves. The emergence of big data has led to an increased demand for self-service BI. New technological tools which allow users to create information from different sources without having to access a data warehouse exist in the market (Salinas & Lemus, 2017). These applications provide users with a simple drag-and-drop interface and the applications themselves create and maintain the data querying and transformation logic in the background.

Traditional BI systems have a limitation when it comes to enabling users to create information. These systems rely on a data warehouse that combines data from multiple sources before such data can then be presented to users for analysis and reporting (Al-Debei, 2011; Winter, 2001). However; traditional BI systems require some level of technical skill that most business users and information consumers do not possess (Alpar & Schulz, 2016). In addition to the challenge of the shortage of skills required to optimise traditional BI systems, section 3.6 discusses some of the common challenges related to designing and maintaining BI systems.

3.6 Business Intelligence Challenges

The increased volume of available data has led to data-driven organisations adopting agile ways of working. Agile business models require BI systems to be able to support the ever-changing information demands of an organisation steadily (Larson & Chang, 2016). BI systems should be able to support long-term decision processes in a non-volatile manner and it should be possible to integrate these systems with emerging data-related innovations. However, the literature suggests organisations face many challenges when implementing and maintaining different components of BI systems to cater for increased information demands.

Besides the lack of skills required to develop effective BI systems, the current BI systems are "not fully capable" to deliver the ever-changing organisational information demands owing to their lack of design flexibility (Knabke & Olbrich, 2011). Lack of flexibility in the design of traditional BI systems hinders the ability to enhance these systems to fit the emerging information and knowledge demands for supporting organisations' decisions. The main reason why traditional BI systems are inflexible is the rigidity of a data warehouse design; the existing data warehouse designs such as star schema and snowflake are not easily changeable to fulfil new requirements (McGlothlin *et al.*, 2017). If a data warehouse is difficult to change, it means such a system is not flexible, lacks leanness and is therefore unable to service agile businesses adequately. Current BI systems rely heavily on a data warehouse that lacks agility and these systems are consequently unable to support the decision support dynamics of today's businesses.

Virtual data warehouses emerged in response to the challenge of the inflexible design of a physical data warehouse that is either modelled using a star schema or a snowflake method. In contrast to a physical data warehouse that stores data and requires physical storage, virtual data warehouses do not store data; they serve as the access point to data that are located in different source systems (McGlothlin *et al.*, 2017). However, virtual data warehouses have not been widely adopted by many BI-based organisations because of a number of limitations associated with this type of data warehouse.

Knabke and Olbrich (2011) highlight three challenges related to virtual data warehouses. First, data-based decision-making sometimes needs to consider historic data and source systems do not store historic data because of performance implications that result from storing such data. Secondly, data analysis queries perform complex aggregation of data and these queries can have a negative impact on the performance of an operational system. Lastly, data stored in operational systems are not cleaned and transformed into a format that is optimised for data analytics and reporting, so cleaning and formatting of the data are required before reports can be generated from a virtual data warehouse. This formatting can hinder the performance of the operational system.

BI systems are able to integrate data from heterogeneous data sources and such data are put into a data warehouse. However, the literature suggests that the exponential volume increase of data from these source systems and the rate at which data layouts change pose major challenges to traditional BI systems (Hawking & Sellitto, 2010; Martins *et al.*, 2015; McGlothlin *et al.*, 2017). The current design of popular data warehouse systems (and ETL) is unable to cope with the increased data volumes that are generated from multiple sources. This problem hinders the effectiveness and reliability of data and information reporting.

The emergence of big data and its perceived value results in many organisations wanting to adopt big data technologies without properly assessing whether big data can be integrated with their existing BI systems. These organisations end up with ineffective BI systems (Sun *et al.*, 2015). There is lack of knowledge regarding whether or not some components of traditional BI systems can be used in collaboration with big data tools to satisfy the information requirements of a data-driven organisation. To address these challenges and increase the likelihood of maintaining effective BI systems, some scholars have identified critical success factors (CSFs) for implementing BI systems. These factors are summarised in the next section.

3.7 Business Intelligence Critical Success Factors

Development and deployment of BI products amount to a complex exercise and projects undertaken to deliver BI products can fail if they are not carefully managed. CSFs identify the dynamics and aspects that need to be considered to ensure the success of a BI project (Dawson & Van Belle, 2013). Factors contributing to the success of a BI project may vary from one organisation to another, depending on dynamics such as industry, the availability of suitable skills, project funding and senior management support (Dawson & Van Belle, 2013; Olszak & Ziemba, 2012). The literature (Table 7) suggests common CSFs across organisations and industries.

Table 7 – Business intelligence critical success factors

Success factor	Source
Senior management support	(Adamala & Cidrin, 2011; Dawson & Van Belle, 2013; Olszak & Ziemba, 2012)
Data quality	(Cai & Zhu, 2015; McAfee *et al.*, 2012)
User involvement	(Alpar & Schulz, 2016)
Clear and realistic objectives	(Larson & Chang, 2016)
Good communication/feedback	(Dawson & Van Belle, 2013; Olszak & Ziemba, 2012)
Team skillset	(Balakrishnan & Rahul, 2018; Negash, 2004)

Executive and senior managers drive the data-driven culture of an organisation and they are the core sponsors of BI projects. Their support is fundamental to the success of a BI system. Data and information quality determines the success of BI, because quality products and services lead to quality decisions. Information consumers or users need to be consulted regularly and be involved in the development of BI services, because these services can only add value to the organisation if users are able to use them effectively.

It is important to ensure that the objectives of each BI product that needs to be developed are both realistic and specific to ensure that the development team is kept focused. A business case for any BI project needs to be defined and aligned to the broader organisation's vision. Collaboration between different BI stakeholders can only be successful if effective communication measures are put in place to ensure that stakeholders are kept up to date and understand their project responsibilities. Lastly, the development and deployment of BI products and services can only be successful if staff with relevant skills are involved in development projects.

3.8 Chapter Summary

Making the right decisions at the right time is a requirement for organisations that not only need to gain a competitive advantage, but also to improve their data analytics culture. Effective decision-making requires proper design and application of data analytic systems.

BI has been widely used in many corporate sectors to source, process and extract value from data. However, the evolving needs for delivering good quality information across all organisational divisions have posed challenges and the effectiveness of traditional BI systems is being questioned. The purpose of this chapter was to explore BI components and ways in which BI adds value to the organisations' decision-making process.

It was suggested that BI is the collection of tools, processes and capabilities whose aim is to convert data into information and knowledge that informs business decisions. A data warehouse is the fundamental component of a traditional BI system because it stores all organisational data in a format that is optimised for information reporting (section 3.3). Another important component of a BI system is a collection of data visualisation and reporting tools.

The value that BI adds to the organisation was discussed in section 3.4. It was suggested that BI offers a data analytics capability that drives organisational success by enabling the organisation's understanding of factors that affect its competitive advantage. BI systems should be assessed regularly to ensure that they satisfy data and information requirements effectively (section 3.5). The four phases describing the process of developing a BI product were discussed in section 3.5 and the suggested metrics to measure each phase of the process were described in section 3.5.1. An optimal BI system is one that is maintained and operated by employees who have the required skills and competencies. A number of common BI capabilities that the BI development team should strive to design were discussed in section 3.5.2. The self-service capability was emphasised as an important competency that BI systems should possess.

Common BI system challenges were discussed in section 3.6. The most common challenges faced when designing BI systems include lack of adequate professional skills that are required to develop BI products. The inability of a data warehouse to cope with increased volumes of data is also a crucial challenge for existing BI systems. The last section of the chapter (section 3.7) listed CFSs for successfully developing and operating BI systems. The identified CFSs include senior management support, data quality, user involvement, clear and realistic objectives, good communication/feedback and an adequate team skillset.

Chapter 4: Big Data in Support of a Data-driven Organisation

4.1 Introduction

The emergence of innovations such as the IoT, cloud computing and social networking has caused the generation of data to increase at an exceptional speed and these innovations have introduced the concept of big data (Atzori *et al.*, 2010). In 2018 Merendino *et al.* suggested that the required analysis of big data had introduced new technological tools and that the perception of traditional BI systems (also referred to as traditional data analytics) had changed. The emergence of big data analytics resulted from the continued effort to overcome the challenge of speedy processing and transformation of big data into business knowledge. Speedy processing of big data can improve the business performance of an organisation through up-to-date reporting of business information as well as automated decision-making (Provost & Fawcett, 2013).

The purpose of this chapter is to explore the definition and the evolution of big data (sections 4.2 and 4.3). An overview of different definitions of big data is outlined to compare the perceptions and interpretations of the concept of big data by different scholars. The dimensions of big data are described to emphasise what differentiates big data from traditional data; this is discussed to give insight into why traditional BI systems struggle to cope with big data information requirements.

As opposed to traditional data analytics that use processes and software tools to convert structured data into information, big data analytics entails the use of software tools and processes to convert big data into business information. The details of how big data analytics improves the decision-making process of an organisation are discussed in section 4.5. Data are complex and every data-related innovation introduces challenges with which organisations have to deal. Some of the most common challenges that organisations experience when adopting big data are described in section 4.6.

A summary of the contents of this chapter is outlined in Table 8.

Table 8 – Chapter 4 outline

Section	Section description	Sub-section	Sub-section description
4.1	Introduction		
4.2	Big Data Definition		
4.3	Big Data Evolution		
4.4	Big Data Analytics	4.4.1	Data science and big data analytics
		4.4.2	Big data visualisation
		4.4.3	Self-service data analytics
		4.4.4	Need to transition from BI to big data analytics
4.5	Value of Big Data in a Data-driven Organisation	4.5.1	Personalised marketing and improved profitability
		4.5.2	Automated decision-making
		4.5.3	Improved customer product offerings
4.6	Big Data Challenges		
4.8	Chapter Summary		

4.2 Big Data Definition

The term big data is used to refer to data sets that are large in volume to the extent that it becomes difficult to process them using a traditional database management system, or to visualise them using traditional data visualisation tools (Elgendy & Elragal, 2016; Madden, 2012; Taylor-Sakyi, 2016). Using more natural language, Madden (2012) suggests that big data refers to data sets that are too big, too fast and too hard. *Too big* means organisations must source and process petabytes of data from multiple data sources. *Too fast* means these data change very quickly and speedy processing is required to extract information from these data sets. *Too hard* means the use of traditional data analytics tools to analyse these data is challenging and as a result, extraction of value from these data at an acceptable timeline becomes impossible. Figure 8 depicts the summary of different dimensions of big data, also referred to as the 5Vs.

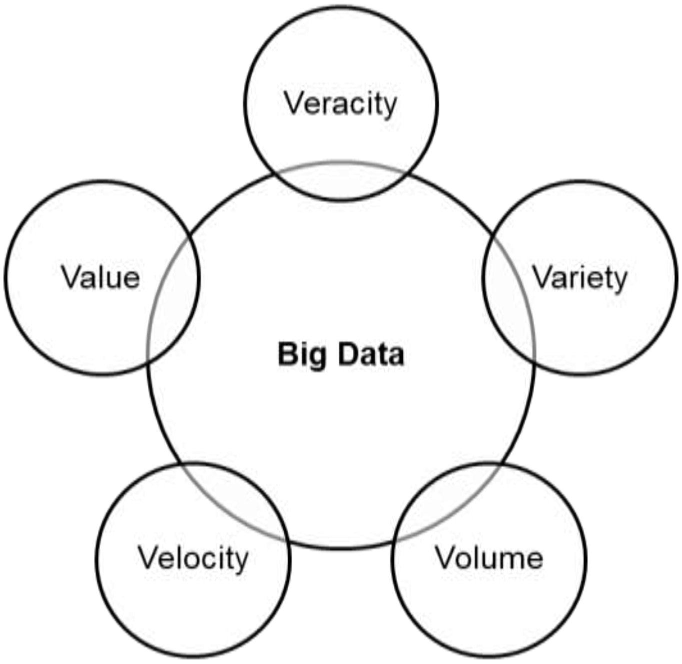

Figure 8 – Big data dimensions (Lee, 2017)

To demonstrate how big data differs from traditional data, the 5Vs are commonly used to describe big data (Alharthi *et al.*, 2017; Cai & Zhu, 2015; Taylor-Sakyi, 2016). The 5Vs are *volume*, *velocity*, *variety*, *veracity* and *value* (Figure 8). Volume refers to the amount (in storage size) of data collected by the organisation with the aim of converting these into business insights. Velocity refers to the speed at which data are collected and processed to service the information demands of a data-driven organisation. Variety refers to distinct multiple data types and formats in which these data come. For example, some data come in an unstructured form whereas others are semi-structured and structured data formats. Veracity refers to the unreliability and uncertainty of the sources where big data reside. Unreliability and uncertainty result from the quality challenges linked to big data, which are mostly related to completeness, accuracy, consistency and latency (Cai & Zhu, 2015). The ultimate aim of using data is to extract business insights that can be used to make effective business decisions. If data produce good information and knowledge that guides effective business decisions, then such data are of high value to the organisation (Merendino *et al.*, 2018).

Traditional data have been exploited by data-driven organisations for a long time. The use of traditional BI systems has been fundamental to the transformation of data to information. To explain how big data emerged and how data complexity has increased to the extent that organisations are unsure about the value of traditional BI systems, the next section discusses the overview of how big data has evolved over the years.

4.3 Big Data Evolution

Even though big data as a concept is still emerging, the collection of large volumes of data goes far back to the times of mainframe systems in the 1990s. Lee (2017) suggests that the World Wide Web (WWW) contributed significantly to the increased generation of data that mainframe systems had to collect to feed analytics. Between 1994 and 2004 e-commerce commenced and online firms formed the core part of the increased quantity of data because of the ease of capturing data through web content. Web usage mining techniques were used to analyse users' web page usage patterns by analysing their mouse clicks and searches. Web structure mining was used to analyse different structural components of websites to gain insight into the popularity of different website components. Through the use of text-mining techniques, web content mining emerged and was used to collect and analyse information that came from web pages. The evolution of big data is summarised in Table 9.

Table 9 – Big data evolution (Lee, 2017)

Years	Era	Enablers (Techniques and tools)
1994 – 2004	Big Data 1.0	Web usage mining, web structure mining, and web content mining.
2005 – 2014	Big data 2.0	Social media content mining, social media usage mining, and social media structure mining.
2015 –	Big data 3.0	IoT applications and streaming analytics.

Between the years 2005 and 2014 big data 2.0 emerged and was extensively driven by Web 2.0. Social media enhanced the usage of web 2.0 and organisations started to capitalise social media platforms for business operations such as marketing. Social media forced organisations to shift towards the analysis of social media content to improve their customer understanding. During that time the emergence of social media content mining, usage mining and structure mining activities underpinned big data (He *et al.*, 2013). Social media users shared information about products and services and organisations accessed and analysed such information to improve their products and services (Merendino *et al.*, 2018).

According to McAfee *et al.* (2012), social media analytics is the dominating analytics field that is rapidly growing every day in response to the perceived value of social media information for organisations. Sentiment analysis and social network analysis are two popular types of social media analysis. User sentiments and opinions are extracted from social media content through sentiment analysis. Social network dynamics and a connection structure for each social media user are measured through social network analysis. Facebook, Twitter and LinkedIn are the commonly used social networking sites that organisations peruse to gather their customers' social networking data (Jamali & Abolhassani, 2006; Ruzgas & Dabulyte-Bagdonaviviene, 2017).

Big data 3.0 emerged in 2015 and is the current era of data and information revolution. IoT devices that generate data through audio, videos and images are central to big data 3.0. IoT is primarily a result of the emergence of sensors and other devices that share data through the use of the internet without any human interaction (Atzori *et al.*, 2010). Streaming analytics has gained more interest from data-driven organisations owing to the increased demands that require fast information generation from streamed data. Streaming analytics is the process of analysing data while it is captured by IoT devices such as sensors (Katal *et al.*, 2013). Therefore, streaming analysis addresses the challenge of real-time data analytics. In general, organisations are perusing big data to enhance their data analytics competencies. Section 4.4 discusses big data analytics.

4.4 Big Data Analytics

In the current era, data generation is exponentially increasing and organisations are demanding speedy transformation of these data to information. Many scholars have suggested that mobile devices, social network platforms and other IoT enablers are the major contributors to the increased generation of data (Atzori *et al.*, 2010; Katal *et al.*, 2013; Ruzgas & Dabulyte-Bagdonaviviene, 2017; Taylor-Sakyi, 2016). As highlighted in section 4.2, these data come from varying data sources and different data sources may have varying data layouts. Big data analytics is a collection of innovative software tools and processes whose aim is to apply advanced analytic techniques to extract business knowledge from big data. To describe big data analytics, Alharthi *et al.* (2017) use three characteristics (volume, velocity, and variety) to compare big data and BI (Table 10).

Table 10 – Big data versus business intelligence (Alharthi *et al.*, 2017)

Characteristic	**Big data**	**Business intelligence**
Volume	Infinite	Finite
Velocity	Real time	Offline
Variety	Unstructured	Structured

Increased data volumes and varying formats require data analytic systems that are capable of converting such data into information and communicate that information through the use of effective information visualisation tools. Traditional data analytics or BI systems are being disrupted by the emergence of big data (Balakrishnan & Rahul, 2018; Madden, 2012; Salinas & Lemus, 2017; Sun *et al.*, 2015). The challenges experienced from the use of traditional data analytic systems have forced many organisations to shift their focus towards big data analytics. In the past data analytics was the responsibility of IT departments; however, information consumers are now demanding access to data to perform analytics themselves with minimal reliance on their IT departments (Alpar & Schulz, 2016). Modern data analytic systems are designed to offer functionalities that minimise the reliance of users on IT (Provost & Fawcett, 2013). Therefore, big data analytic systems improve data exploitation by enabling information consumers to be part of the transformation of data into information.

Data consumers not only need access to internal organisational data; they also need data from external data sources such as social media platforms. Social media information is perceived to be useful in understanding society and such data can be used by organisations to understand their customers better (Taylor-Sakyi, 2016). Most data that are part of big data are unstructured (e.g. images, videos and free text) and traditional data analytics systems are unable to process these data formats (Madden, 2012).

Data storage becomes an issue when multitudes of data are extracted from external data platforms. In response to this, organisations are adopting cloud computing and use cloud computing storage services to address this challenge because these services are perceived to be cheaper that traditional storage infrastructure (Yang *et al.*, 2017). When big data are stored in the cloud, different types of data analytics methods are used to transform these data into information. Location analytics, sentiment and graph analytics are some of the emerging and common big data analytics techniques used by data scientists and data engineers (Ruzgas & Dabulyte-Bagdonaviviene, 2017). Data science is a data analytics specialisation that has become popular in organisations that are adopting big data analytics. Section 4.4.1 explores the relationship between data science and big data analytics.

4.4.1 Data science and big data analytics

Data science emerged in response to the requirement for fast processing of heterogeneous big data sets located in multiple sources. Ruzgas and Dabulyte-Bagdonaviviene (2017) suggest that data science is a field of data analytics where scientific methods are used to analyse large volumes of data sets that are both structured and unstructured. Even though the tools used to support data science operations differ from one organisation to the other, some common tools and personnel skills are common across many organisations.

Data-mining tools and skills are some of the fundamental competencies required to strengthen the maturity of organisations' data science capabilities. Data mining involves the identification of structures and connections in big data and applying statistical models to discover patterns or trends from different data objects (Aggarwal, 2015). Some scholars suggest that data mining involves the exploration of big data sets to uncover patterns and rules through the use of automated and semi-automated technological tools and processes (Hand, 2007; Koh & Tan, 2011; Peral *et al.*, 2017). Data mining, therefore, plays a major role in converting big data into business information and knowledge that inform the decision-making process of a data-driven organisation. Data science is key to big data analytics and data mining is the backbone of data science.

4.4.2 Big data visualisation

The existing traditional reporting tools struggle to deliver big data reporting and data visualisation requirements. Effective reporting of big data information determines the success of the decision-making that is supported by big data analytics. Solutions to data visualisation and reporting problems should take priority in big data analytics systems because these problems can easily undermine the purpose of big data analytics (Surbakti *et al.*, 2019). By its nature, big data changes very often and therefore, reporting of big data information should be done as quickly as possible to ensure that decisions are aligned with and are based on relevant and up-to-date data.

Reporting and visualisation of big data analytics results play a major role in the lifecycle of big data. While big data analytics assist organisations in uncovering insights that may be hidden in big data, visualisation and reporting help business users and decision makers to understand big data insights. Decision makers and data consumers need a capability that enables them to access big data visuals from multiple kinds of devices. For effective big data analytics systems, on-demand access to big data information through the use of relevant reporting and visualisation tools is a must. Therefore, visualisation tools used in big data analytic systems must be chosen wisely to satisfy these requirements (Keim *et al.*, 2013).

A complete big data analytics system should include tools for reporting, dashboard creation, and advanced data analytics. In 2016, Taylor-Sakyi suggested that most big data analytic systems that existed in the market included all these capabilities. The core requirement of big data analytics is speedy processing and accuracy of big data. Large volumes of data must be analysed speedily and the quality of analytics is determined by the accuracy of the results (Provost & Fawcett, 2013). In-memory analysis is embedded in most big data analytics tools. It improves the determination of relationships between big data objects, given hundreds of parameters (Han & Chang, 2002). In-memory analysis becomes handy when free-text format data from social media such as Twitter and Facebook are being analysed. Data from the social media ecosystem are becoming more attractive to data-driven organisations. Hence, the tools used to analyse and visualise big data should be able to effectively process and present information that is derived from text mining and other social media data analytic techniques.

4.4.3 Self-service data analytics

Big data analytics is useless without proper capabilities to offer self-service data analytics to information consumers. Decision makers and information consumers should be able to access, analyse and report big data insights through proper visualisation tools without having to rely on IT and without having to acquire advanced technical skills (Alpar & Schulz, 2016). The choice of big data analytics tools should be made cautiously because analysing exponentially growing and varying formats of data can become a challenge and the creation of self-service data analytics capabilities for such data can be difficult. Traditional data analytic systems have achieved positive developments in offering self-service reporting capabilities such as online analytical processing (OLAP) and tabular analysis services, Tableau, QlikView, and other tools available in the market (Knabke & Olbrich, 2011; Sun *et al.*, 2015). However, these systems are unable to process big data.

Systems and tools that support effective self-service data analytics need a flexible IT infrastructure that can handle complex data operations. Workload balancing, job prioritisation, high availability, parallel processing, resource assignment and monitoring should form part of the IT infrastructure that supports big data analytics to enhance the ability of big data analytic systems to offer self-service reporting (Ruzgas & Dabulyte-Bagdonaviviene, 2017).

4.4.4 Need to transition from BI to big data analytics

Understanding of traditional data analytics (or BI) processes and techniques is very important in understanding the need for big data analytics. RDBMS play a major role in traditional data analytics because they provide a user-friendly querying ability to data consumers and developers (Al-Debei, 2011). RDBMS store data in tables, which have rows and columns. The way RDBMS store data does not cater for storing and analysing big data (Elgendy & Elragal, 2016; Lee, 2017). The retrieval of data from traditional systems such as data warehouses becomes a challenge as the volumes of data increase.

Poor performance of a data warehouse results in a less effective data analytics system as a whole. Organisations are pursuing big data with the intention of understanding their businesses better and making informed decisions based on big data information. The lack of speedy transformation of large volumes of data by RDMS leaves organisations with no choice but to adopt big data analytics.

Organisations need to keep up with data changes and make sure that decisions are based on up-to-date information. This requires data analytic systems that are able to cope with processing big volumes of data at speed. ETL and data warehouse systems form the backbone of traditional data analytic systems, but these system components are unable to cope with speedy transformation and storage of big data sets (Knabke & Olbrich, 2011). An effective data warehouse system relies on the cleaning and formatting of data by ETL tools. The ETL process is resource-intensive and this can delay information generation (Elgendy & Elragal, 2014).

Organisations are looking for analytical solutions that are flexible enough to source differently formatted data and analyse these without problems (McGlothlin *et al.*, 2017). Emerging big data analytics technologies are considered to address challenges that traditional BI systems experience with processing big data. Traditional BI systems normally have different levels of data-processing, which require different supporting tools. For example, they include one software tool to extract and transform data, another one for storing the data, and another tool for data analysis and reporting (Devlin, 2010). This makes the system complex and optimisation of the entire data-to-information conversion process becomes a challenge. The emerging big data analytics systems include all the data-to-information conversion capabilities in one system (Elgendy & Elragal, 2014).

The ability to create analytics on live data is another determiner of which systems survive the challenges related to extracting business value from big data. According to Taylor-Sakyi (2016), the implementation of live big data analytics allows organisations to remodel their processes continually to ensure that effective business operations are in place to improve the organisation’s performance. As suggested by Han and Chang (2002), data mining and live data streaming assist in ensuring that decision-making is improved through information reporting that is generated from live data.

Big data analytics is not just an ideology, it is a collection of tools and processes that are used to derive business insights from big data (Chan, 2013; Shoro & Soomro, 2015). Section 4.5 gives an overview of some of the key offers of big data analytics to a data-driven organisation.

4.5 Value of Big Data in a Data-driven Organisation

Data-driven organisations use data as a basis of their decision-making and their ability to source and process more data without challenges therefore allows them to improve their decision-making. Big data contain larger volumes of data that are stored in multiple formats and types. Access to big data means data-driven organisations can exploit any type of data that they can imagine. According to Provost and Fawcett (2013), data science is the most important field that enables organisations to realise the value of big data, because data scientists ensure that big data tools and techniques are used to extract value from big data. Effective use of big data tools and competent application of data science skills improve an organisation's decision-making. Details of use cases where big data, when utilised effectively, can add value to a data-driven organisation are explored next.

4.5.1 Personalised marketing and improved profitability

Data science and data mining are competencies that fundamentally assist data-driven organisations to extract value from big data. Big data analytics allows organisations to collect and store detailed data in varying layouts to allow organisations to perform low-level data analysis. For example, data scientists can analyse the reaction of each customer to different advertisements. Such detailed analysis of data allows a data-driven organisation to enhance its marketing strategies and provide personalised marketing to its customers. Some studies suggest that such a level of marketing improves the productivity of a data-driven organisation (Provost & Fawcett, 2013; Puglisi *et al.*, 2017). Consequently, it can be suggested that big data improves the productivity of a data-driven organisation.

Customer-level analysis of big data allows an organisation to determine the profitability of each of its customers. Data-driven organisations can leverage big data solutions, which allow them to know exactly which customers need to be retained based on their individual profitability. The customer-level profitability analysis ensures that the organisation invests in ensuring that it retains profitable customers, as opposed to blindly utilising resources on less profitable customers (Niraj *et al.*, 2001). According to Lee (2017), banks are leveraging big data initiatives with the notion of increasing their revenue and improve customer retention. The ability to assess customer-level profitability also guides and can improve the organisation's product pricing strategies.

4.5.2 Automated decision-making

Effective big data analytic systems are suggested to automate decision-making through the use of data science models. Data science models can be fed with different data and be trained to perform different operations based on data parameters that are passed to the models (Van Der Aalst, 2016). This suggests that data science models can be trained to make data-based decisions and inform business about the next plan of action based on specific data events. The results of big data analytics inform decision-making and such results tell a story to the extent that decision makers can share the results of a data science model without even translating them.

A good example of the application of data science to automate decisions occurs in customer-level online advertisement. The decisions on which online advertisements should be presented to which user are all automated and such split-second decisions are made by the system itself. The ability to automate decision-making thus ensures that organisations are more effective. Automated decision-making releases some effort from personnel resources and enables the organisation to spend more time on activities that require personnel intervention. Therefore, automated decision-making also reduces cost.

4.5.3 Improved customer product offerings

The ability to collect more data enables organisations to identify data-based opportunities that they could not identify before, when they were limited to only processing structured data (Elgendy & Elragal, 2016). For example, the ability to analyse videos that are posted by those customers whom the organisation considers to be more profitable than others can add more insights to understanding their opinions and lifestyle. Organisations are exploring the option of analysing social media data to identify their customer needs (He *et al.*, 2013). Decisions can then be taken based on the assessment of these data to ensure that products are tailored to suit customer needs. Therefore, big data improve the performance of a data-driven organisation by increasing the potential to create new products and improving the existing customer product offerings.

4.6 Big Data Challenges

The value that big data add to data-driven organisations has been demonstrated in many studies. However, big data adoption has also proven to introduce some challenges to those organisations that are pursuing this data innovation. An overview of the most common big data challenges is summarised in Table 11.

Table 11 – Big data challenges

Challenge	Sources
Data quality concerns	(Cai & Zhu, 2015; Katal *et al.*, 2013; Lee, 2017; Taylor-Sakyi, 2016)
Lack of skillset	(Alharthi *et al.*, 2017; Lee, 2017; McAfee *et al.*, 2012; Nasser & Tariq, 2015)
Data security and governance risks	(Alharthi *et al.*, 2017; Taylor-Sakyi, 2016)
Expensive toolset	(Nasser & Tariq, 2015)
Lack of frameworks for well-architected architectures	(Nasser & Tariq, 2015)
Unsupported investment justification	(Lee, 2017; Surbakti *et al.*, 2019)
Inability to integrate with legacy systems	(Alharthi *et al.*, 2017)

The value of data fundamentally depends on the quality of the data. Owing to a variety of sources and formats of big data, it is difficult to define measures that ensure that the quality of big data is maintained (Alharthi *et al.*, 2017; Lee, 2017; McAfee *et al.*, 2012;

Nasser & Tariq, 2015). Analysis of poor quality data yields poor analytics results, which can mislead the decision-making of a data-driven organisation. The challenges pertaining to the quality of big data are included in many big data literature studies that seek to investigate big data adoption by organisations. This is a major challenge, considering that big data are emerging and there are no existing data quality assessment techniques for big data.

There is a perception that big data analytic systems are expensive to adopt. This is due to the computing power required by the supporting operations such as data mining and text mining, in addition to storage requirements for storing large volumes of data (Nasser & Tariq, 2015). Organisations are determined to optimise their IT budget and it is difficult to obtain capital for big data projects. However, if an organisation is unable to provide the required resources for processing big data in a reasonable timeframe, it is likely to use outdated data for its decision-making.

As organisations gain access to collect more data, concerns about data security need to be considered. The variety of data sources and data formats of ever-changing big data introduce major security concerns. Existing data governance guidelines for big data are still immature (Alharthi *et al.*, 2017; Taylor-Sakyi, 2016). While most big data tools are considered to work effectively when integrated with cloud computing services, the emergence of cloud computing has raised security concerns in many organisations (Surbakti *et al.*, 2019). Hence, big data privacy and security management concerns have been important talking points in many organisations (Yang *et al.*, 2017).

It is also suggested that many organisations are finding it hard to justify investment in big data (Surbakti *et al.*, 2019). Big data uses emerging technologies that many executive managers regard as risky investments. Modern technology tools are expensive and product or project sponsors require justification for such expensive investments. Many organisations struggle to come up with problem definitions that clearly require big data innovations. Moreover, given that organisations are not obligated to use big data, senior managers need justifiable reasons for big data adoption before they can approve the required budget.

Studies report that big data-related skills are increasingly being demanded in many organisations across all industries (Provost & Fawcett, 2013). Academic institutions have introduced big data and data science degrees to assist in bridging the gap between the market for data scientists and organisations' demands (Mills *et al.*, 2016). Some studies have revealed that there is a lack of frameworks that guide the design and development of effective big data analytic systems.

To position the application of the selected research methodology and tools, section 4.7 discusses theories that were used to guide the selection of the methodologies and interpretation of research findings.

4.7 Research Theory

Typically, a research study is conducted to explore essential information about a phenomenon, not to describe everything that must be known about the phenomenon. A research theory underpins the interpretation of the events that are observed during the research (Palvia *et al.*, 2006). Theory supports a research paradigm because the way a researcher understands a research problem depends to a significant degree on the researcher's background information. A researcher's background information is determined and supported by the theories to which the researcher has been exposed. According to Krauss (2005), a researcher's belief system affects the choice of research methodology, because of the researcher's ontological assumptions. Based on the purpose of the study, the researcher considered theories that would guide information-processing and interpretation of the findings. The organisational information-processing theory (OIPT) and a fit-viability model (FVM) are the two theories that were applied in the current study because they are concerned with information processing and the adoption of technological innovations.

The relevance of OIPT is based on the ultimate goal of data analytics, which is making sure that organisational data are processed to produce information that can be used to guide all business decisions of a data-driven organisation. OIPT is concerned about information processing, hence it is relevant to this study.

FVM suggests that organisations need to assess their abilities to cope with the new technological innovations that they consider adopting. The appropriateness of the technological innovations should also be checked against the organisation’s culture to ensure successful adoption. Therefore, FVM concerns the adoption of new technological innovations by organisations. The adoption of big data analytics requires the organisation to have the necessary capabilities to support big data analytic systems. This suggests that organisations first need to assess their readiness to adopt big data-related innovations. Hence, the FVM was also considered relevant to this study.

The following sections describe the details of these two theories, as well as their relevance to this study.

4.7.1 Fit-viability model

The FVM is the theory that provides guidelines related to the adoption of new technological systems by organisations. The theory states that the decision on whether or not an organisation should adopt a technological system it considers is based on two factors: the *fitness* and *viability* factors (Kwanya, 2014; Liang *et al.*, 2007; Ossai & Wickramasinghe, 2021). According to Kwanya (2014), fitness refers to the extent to which the system is able to deliver its expected value to the organisation. This means that fitness assessment should be done timeously to ensure that the system remains valuable as the organisation emerges. Figure 9 summarises the relationship between the different dimensions of an FVM.

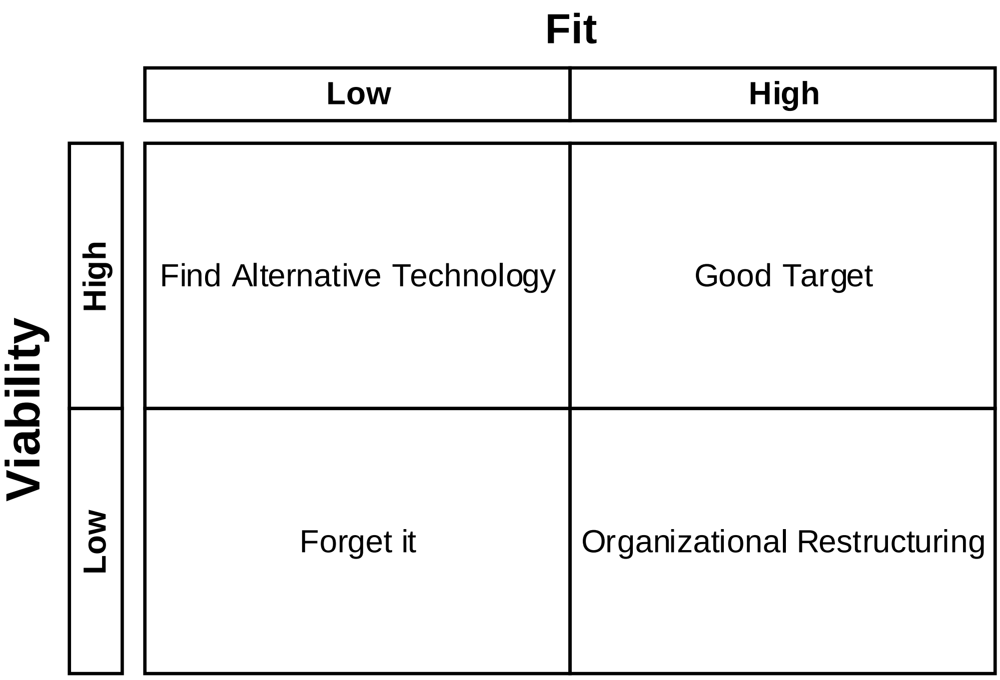

Figure 9 – Fit-viability model (Liang *et al.*, 2007)

Viability refers to the extent to which the organisation is ready to cope with the requirements of the new system (Liang *et al.*, 2007). The infrastructure that will host the system is assessed according to the demands of the effective use of the system. As shown in Figure 9, if the viability and fit are both unfavourable, the suggestion is that the organisation must not adopt the system. If both of these measures or factors are favourable, then it is considered reasonable for an organisation to adopt the new system. The high variability and low fit (or the other way round) might result in either organisational restructuring or finding an alternative system. An example of the application of an FVM in big data adoption is the study that was conducted by Kwanya in 2014. A summary of the application is depicted in Figure 10.

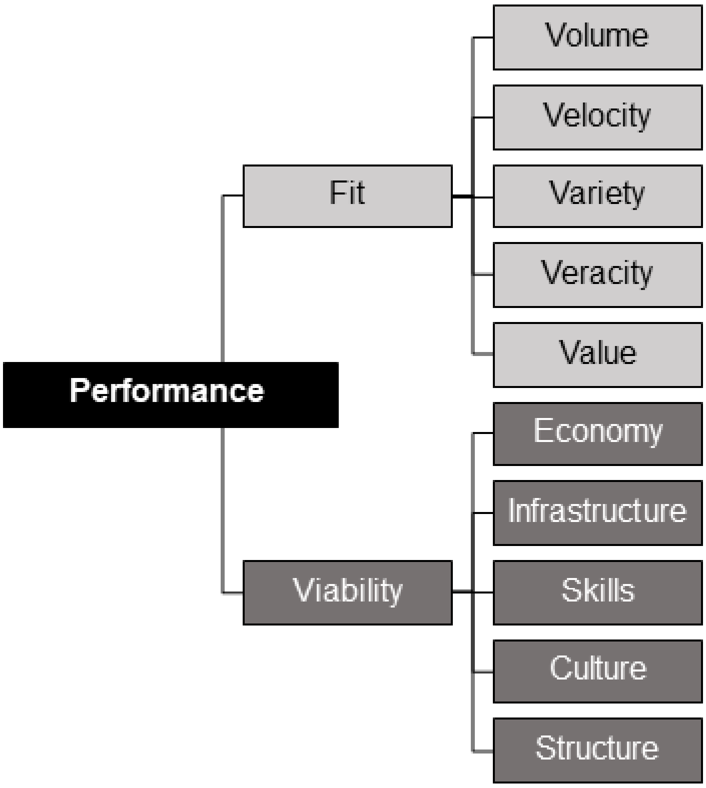

Figure 10 – FVM application in big data (Kwanya, 2014)

When the theory is applied in big data adoption projects, the fit element can be related to the ability of a big data analytics system to satisfy the requirements of a data-driven organisation. The 5Vs (volume, velocity, variety, veracity and value) of big data have to satisfy the expectations of the organisation collaboratively. Therefore, the characteristics of big data have to add value to the organisation's decision-making and improve its performance.

Viability is the measure that can be used by the organisation to assess its readiness to operate big data analytics effectively (Figure 10). The budget required to fund big data projects and upskilling of employees to ensure that they can work with big data analytics are some of these factors that the organisation has to assess. The infrastructure to host big data systems, including increased data storage for larger volumes of data, has to be in place. Big data analytics requires the organisational culture to change and everyone has to be data-centric, not only the analysts and IT employees (Provost & Fawcett, 2013). The application of the FVM in this research study follows the same process that Ossai and Wickramasinghe (2021) applied to analyse users' opinions on the efficiency of diabetes mobile applications.

4.7.2 Organisational information processing theory

The proposition of the OIPT is that the more the organisation finds itself experiencing uncertainties, the more data it requires to understand its environment (Galbraith, 1974; Lai *et al.*, 2020; Premkumar *et al.*, 2005). To describe the notion behind OIPT, Galbraith (1974:28) suggests the following:

> *"… the greater the task uncertainty, the greater the amount of information that must be processed among decision makers during task execution in order to achieve a given level of performance ..."*

According to Premkumar *et al.* (2005), the balance between uncertainties and information is hypothesised to improve the performance of the organisation because it allows the organisation to: (1) improve its ability to plan proactively for the upcoming situations, (2) enhance its ability to adapt to situations and re-plan, and (3) decrease the effort required to improve its performance.

In the context of this study, OIPT can be applied to ensure that the uncertainties related to the challenges that big data pose to the organisation are compared to the organisation's abilities to address them. The capabilities that the organisation implements to address big data challenges can be viewed as the enhancer of the information-processing need. The balance between uncertainties and information processed can then result in the successful adoption of big data analytic systems and the integration of these systems with BI. Furthermore, data-driven organisations need to access more data in view of the value that can be extracted from these data. That is an information-processing requirement. The issues linked to big data adoption and traditional data analytic systems are regarded as the uncertainties that data-driven organisations experience. A framework for integrating traditional BI and big data analytic systems can then improve the balance (fit) between information-processing needs and organisations' uncertainties. This follows the process that Lai *et al.* (2020) applied to investigate a supply chain collaborative decision-making model.

4.8 Chapter Summary

The purpose of this chapter was to explore the details of big data and big data analytics. The concepts of big data and big data analytics were defined and described (sections 4.2 and 4.4). Big data is the term used to describe data that are large in volume, change very often, and whose format varies from time to time depending on distinct sources from which these data are being extracted. Big data are described using five dimensions (5Vs): *volume*, *velocity*, *variety*, *veracity* and *value*. Big data analytics enables organisations to gain value from big data by extracting business insights from big data. Data science practices allow organisations to extract and communicate big data insights through information visualisation.

A data analytics capability is only effective if it provides enhanced data visualisation tools that serve as a communication medium between data scientists, data miners, and business decision makers (section 4.4.2). Self-service data analytics is one of the growing demands in both big data and traditional BI systems. Information consumers need to rely less on IT staff and tools allowing them to exploit data must therefore be adopted (section 4.4.3). While exploring the reasons why organisations are adopting big data analytics (section 4.4.4), it was suggested that the demand to access more structured and unstructured data is one of the leading factors forcing organisations to adopt big data analytics.

Three use cases where big data analytics can benefit a data-driven organisation were discussed in section 4.5. Firstly, analysis of customer-level data that come in any format allows the organisation to understand its customers better. This can result in a focused and improved customer-level profitability analysis. Secondly, the ability to analyse large volumes of data, including unstructured formatted data, allows organisations to access information that enables them to automate decision-making through the use of data science models. Lastly, an improved customer product offering requires an organisation to understand its customers better. Access to both internal and external data can offer the organisation improved customer understanding.

Like other IT innovations, big data come with a number of challenges that organisations need to address (section 4.6). Some of the common challenges include lack of formal data quality assessment frameworks, costly tools, lack of justifiable investment motivations, as well as lack of a personnel skillset required to extract value effectively from the analysis of big data. The details of the theories applied in guiding the research process and the interpretation of the research findings were discussed in section 4.7.

Part III – Research Design and Methodology

Part III (Figure 11) contains Chapter 5, which describes the methodology and research design techniques of the study.

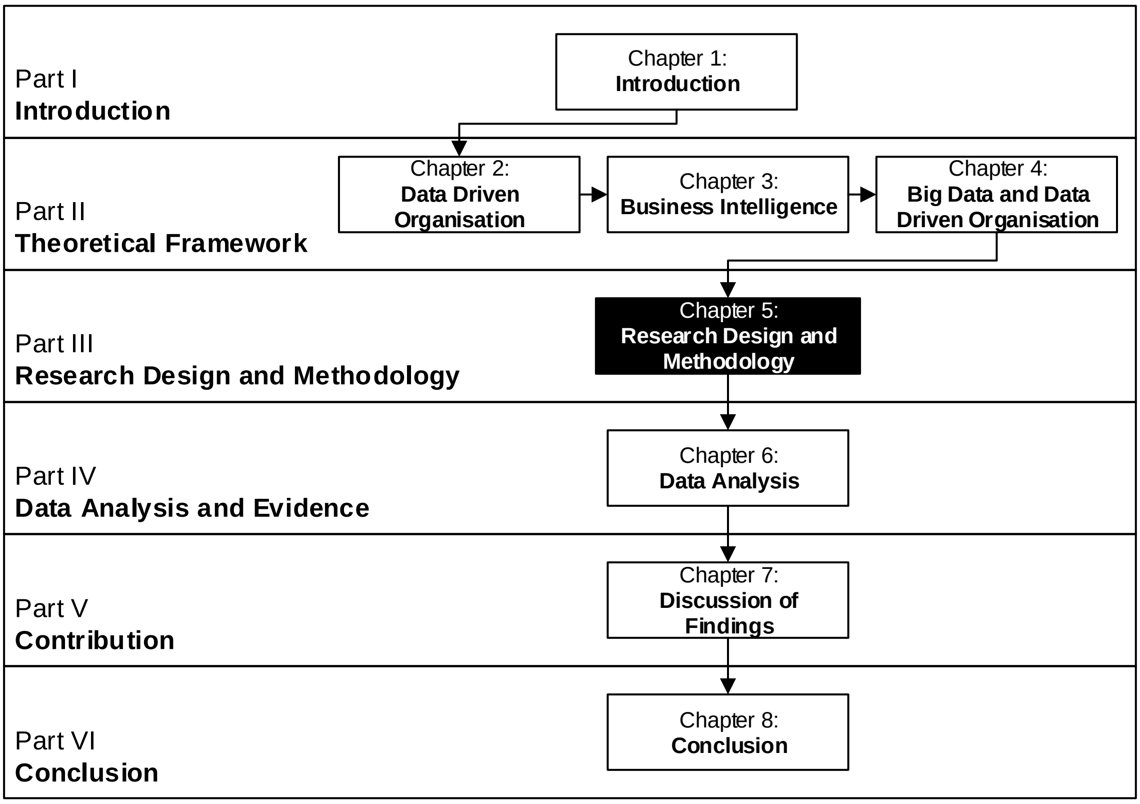

Figure 11 – Part III outline

Chapter 5: Research Design and Methodology

5.1 Introduction

Scholars have shown increased interest in how IS research has emerged from both a design and a methodology perspective (Chen & Hirschheim, 2004; McVilly *et al.*, 2008). A methodological approach to conducting research in IS plays a major role in the quality of a research study (Alharahsheh & Pius, 2020). It is important that the researcher understands the limitations of different research methodologies before he or she makes a decision on which methodology to follow when conducting a specific research study. Therefore, careful attention to selecting the research methodology to use in the research should not be overlooked.

Philosophical assumptions guide the selection of the appropriate research methodology. Philosophical assumptions play a major role in the validity of the study, because they give a basis to interpretation of the study outcomes (Orlikowski & Baroudi, 1991). Positivist and interpretivist philosophical assumptions are the two widely used research philosophical assumptions in IS (Cavaye, 1996). The details of these comparative research paradigms (or philosophical assumptions) and research methodologies, which are widely used in IS, are described in sections 5.2.1 and 5.2.2. Quantitative and qualitative methods are the commonly used research methods to conduct research in IS. An overview of both qualitative and quantitative research methodologies is summarised in section 5.2.

Section 5.3 describes the details of the research methodology and design specifically for the current study. The applied sampling strategy and data collection methods are described in sections 5.3.4 and 5.3.5 respectively. The overview of a data analysis technique that was applied to extract value from research participants' feedback is described in section 5.3.6. Towards the end of the chapter, ethical considerations applicable to this research study are summarised in section 5.3.7.

A summary of the contents of this chapter is outlined in Table 12.

Table 12 – Chapter 5 outline

Section	Section description	Sub-section	Sub-section Description
5.1	Introduction		
5.2	Information Systems Research	5.2.1	Research paradigms
		5.2.2	Research methodologies
5.3	Research Methodology and Design	5.3.1	Research questions
		5.3.2	Research paradigm
		5.3.3	Research methodology
		5.3.4	Sampling
		5.3.5	Data collection
		5.3.6	Data analysis
		5.3.7	Ethics and anonymity
5.4	Chapter Summary		

5.2 Information Systems Research

The purpose of this section is to give an overview of research paradigms and methodologies that are widely used and applied in IS, where after the choices made for this study are presented.

5.2.1 Research paradigms

A research paradigm is a common viewpoint that groups and binds all theories used in research (Goles & Hirschheim, 2000). Such theories determine the processes and methodologies that will be used in the research study. Positivism and interpretivism are the two widely used research paradigms in the field of IS (Goldkuhl, 2008; Tsang, 2014). The comparative attributes of positivism and interpretivism research paradigms are summarised in Table 13.

Table 13 – Positivism and interpretivism beliefs (Alharahsheh & Pius, 2020)

Belief	Positivism	Interpretivism
Ontology	Single reality	Multiple realities
Epistemology	Objective	Subjective
Methodology	Quantitative	Qualitative
Axiology	Truth	Understanding

The ontology element of a research paradigm is concerned with the nature of the reality of the world (Alharahsheh & Pius, 2020). It determines the knowledge acquisition based on what is known about the phenomenon that is being investigated. Epistemology is concerned with what a researcher perceives to be valid knowledge, based on how a researcher views the world (Kroeze, 2012). Methodology refers to the research strategy followed to conduct research (Palvia *et al.*, 2006). Axiology refers to the primary aim and the value that the research aims to uncover (Kroeze, 2012). The following sections discuss the details of these two widely used research paradigms in IS.

5.2.1.1 Positivism

The ontological belief of positivists assumes that the world exists independently of human experiences and therefore that reality exists objectively. Epistemologically, Orlikowski and Baroudi (1991) suggest that positivists are generally concerned with the assessment of theories where deductive reasoning forms the core part of the research study results. Generally, positivists believe that study outcomes are independent of the researcher and anyone else involved in the research (Chen & Hirschheim, 2004). This suggests that if the same study is repeated by different researchers, there is an expectation that the same outcomes will be obtained.

Positivists separate themselves from the phenomenon that is being investigated and the use of deductive reasoning is fundamental to proposing and verifying positivist theories (Orlikowski & Baroudi, 1991). In most cases, positivists attempt to understand the predictive nature of the concept being studied (Myers, 2002). Positivists believe that everything can be measured and verified through mathematics (Krauss, 2005). Quantitative research methodologies are typically used by positivists to conduct research where quantitative data collection methods, such as experiments, are used (Queirós *et al.*, 2017).

5.2.1.2 Interpretivism

Ontologically, interpretivists believe that the world is shaped by the existence of human beings and their interaction with socially connected processes (Chen & Hirschheim, 2004). Human understandings and experiences are fundamental to the reality of the world. Human feelings and understanding fundamentally contribute to the validity of knowledge and interpretive researchers are therefore generally interested in research participants' different experiences and viewpoints (Myers, 2002).

Knowledge is not obtained through deductive reasoning, which generally seeks to prove the hypothesis; rather, social knowledge constructed in consideration of human understanding, culture and feelings contributes to knowledge creation and validity. From the epistemological point of view, interpretivists assume that knowledge should be obtained through the understanding of humans and subjective meaning of reality should be considered (Kroeze, 2012). Methodically, qualitative research methods are typically used in interpretive studies and qualitative data collection techniques such as unstructured and semi-structured interviews are used to acquire research data (Krauss, 2005).

5.2.2 Research methodologies

The selection of the appropriate research method for application in a particular study is very important, as it affects the quality of the research study. Epistemological and ontological beliefs of a researcher form the basis of the selection of the appropriate research method. Moreover, the researcher's professional background and the phenomena being investigated guide the research method selection (Myers, 2002). Different research methods can be associated with either a positivist or interpretivist research paradigm.

The most commonly used methodologies in IS are quantitative and qualitative research methodologies. It is also possible to use a mixed method that combines both qualitative and quantitative methodologies to conduct research in IS (Venkatesh *et al.*, 2013). Mixed methods are used where the tension between the two dominant methodologies is not important but the focus is on applying a design that best answers the research question (McVilly *et al.*, 2008).

Quantitative research uses numerical procedures and methods to identify relationships between factors related to the phenomena being investigated. Qualitative research is normally focused on describing and understanding the phenomenon studied, while considering cultural factors of the setting where the study is being conducted (Myers, 1997). In most cases, qualitative studies assume an interpretivist paradigm, while quantitative studies assume a positivist paradigm (Mackenzie & Knipe, 2006). However, according to Myers and Avison (2002), it is possible for a qualitative study to assume either an interpretive or positivist research paradigm.

A number of strategies exist that can be used in both qualitative and quantitative studies. To give some context to different research strategies that are commonly used in IS, sections 5.2.2.1 and 5.2.2.2 give a summary of both quantitative and qualitative research strategies.

5.2.2.1 Qualitative research strategies

According to Myers and Avison (2002), the commonly used qualitative research strategies in IS are action research, case study research, ethnographic research and grounded theory. Because of their relevance to most IS qualitative studies, high-level descriptions of these research strategies are summarised next.

Action research

Action research is one of the well-established research strategies in the social sciences and medical science fields and has been in existence since the mid-twentieth century (Baskerville, 1999). This research strategy has been widely adopted by IS researchers to investigate IS-related problems. Action research adopts a viewpoint of the interpretivist paradigm and the use of qualitative data collection and analysis techniques is common to it. The core assumption of action researchers is that organisations and their IT systems need to be considered as whole entities because isolating these two may lead to a poor quality research study and unjustifiable knowledge (Baskerville & Wood-Harper, 1996).

Action researchers seek to develop social understanding of organisations and their IS. Collaboration of different actors involved in action research is one of the fundamental values in gaining understanding of social systems (De Villiers, 2005).

Case study research

The use of the case study research method has received increased attention in the field of IS. Case study research is a collection of methods that emphasise qualitative data collection and analysis. Qualitative data collection methods such as interviews are used with the aim of promoting in-depth understanding of the phenomena investigated (Annansingh & Howell, 2016). Generally, the aim of case study research is to understand the organisation where the research is being conducted and the outcomes of the study cannot be generalised (Benbasat *et al.*, 1987; Gable, 1994; Yin, 1981). This allows the researcher to investigate organisational IS in their natural context. The knowledge acquired through case study research can enhance theories that can be applied practically in the organisation (Fidel, 1984).

Case study research can examine either a single or multiple cases. A single-case study method is more suitable for investigating concepts that have never been investigated before, whereas confirmation or providing proof of the existing model, framework or hypothesis can employ multiple case studies (Gustafsson, 2017). It is also possible to use case studies to describe or explain why certain events occur. Organisational decision-making and innovative projects are some of the typical research areas where a case study research strategy can suitably be applied (Benbasat *et al.*, 1987; Yin, 1981).

Ethnographic research

Ethnographic research is underpinned by the belief that the researcher has to spend time in the context and setting where the research is being conducted to understand the experiences and culture of that particular environment (Myers, 1997). Depending on the complexity of the phenomenon studied, it is possible for an ethnographic research study to take more time than when other research strategies are applied. This is due to the time the researcher has to spend trying to understand the lives and experiences of people who are involved in the study (Myers, 1999). As opposed to case study research where data are normally collected through interviews and may be supplemented with some documents, ethnographers literally have to be involved in studying the social dynamics of the setting where the study is being conducted. This means that observations are a requirement of ethnographic studies.

Grounded theory

Grounded theory is a research strategy that is best suited to studies whose aim is to develop a theory or a model (Urquhart *et al.*, 2010; Wiesche *et al.*, 2017). Wiesche *et al.* (2017) suggest that the application of grounded theory in IS is limited because most IS research studies do not seek to develop a theory or model. However, another opinion from Urquhart *et al.* (2010) suggests that over the last decade there has been an increase in interest in the application of grounded theory in IS research studies. Even though grounded theory studies are typically used to develop theory, in IS grounded theory is also used to study technological changes. Grounded theory strategy emphasises the interplay between data collection and analysis, which allows the researcher to extract the emerging data insights that may have not been considered in the initial stages of the study (De Villiers, 2005). Urquhart *et al.* (2010) suggest that when grounded theory strategy is used in collaboration with other research strategies, it is mainly used for data coding, as opposed to developing a model or theory.

5.2.2.2 Quantitative research strategies

Qualitative research involves the collection and analysis of data that can be quantified and manipulated using mathematical and statistical formulae. Numerical data are collected and analysed through statistical analytic tools to discover insights for the quantitative research study. Field experiments, simulations and surveys are mentioned as the commonly used quantitative research strategies in IS (Queirós *et al.*, 2017).

Field experiments

Field experiments are conducted in a natural setting and are underpinned by the act of variable manipulation. The researcher tries to manipulate at least one independent variable and observe how the outcomes of the manipulation affect a dependent variable (Chatterji *et al.*, 2016). Experiments are conducted in the laboratory to ensure that the researcher can manipulate constant or independent variables and record observed outcomes. It is, however, notable that laboratories often do not resemble the real world (Queirós *et al.*, 2017).

Simulation

Simulation research strategy is generally used to analyse complex real-world problems where a computer model is used to simulate as many real-life processes as possible (Kothari, 2013). The analysed real-life processes have to be related to a specific phenomenon studied. Even though simulation can be used to solve complex real-life problems, Queirós *et al.* (2017) suggest that the use of this research strategy can be complicated in view of the time and knowledge required to create a computer model that can simulate as many real-life processes as possible.

Surveys

A survey is a research technique that allows a researcher to collect data from the research participants through a set of organised questions. The formulation and organisation of survey questions play a major role in the quality of the feedback received from the research participants. In survey research, participants are referred to as a population (Pinsonneault & Kraemer, 1993). Surveys are considered to be cheaper when compared to other quantitative research techniques because of their ability to reach a wider audience with less effort and cost.

5.3 Research Methodology and Design for the Current Study

This section describes the research methodology and design of the current research study.

5.3.1 Research questions

The aim of the current study was to investigate a framework for hybrid DD-DSS that includes components from both traditional data and big data analytics to enhance the information processing of data-driven organisations. The following main research question was formulated to guide the investigation of a hybrid DD-DSS framework:

What are the core elements of a conceptual framework that can be used to design a hybrid data-driven decision support system that combines traditional business intelligence and big data analytics?

In attempting to answer the main research question of the study, the following sub-questions were investigated:

- *What is meant by a data-driven organisation and data-driven decision-making?*
- *What constitutes traditional business intelligence and big data analytics?*
- *What are the big data and business intelligence adoption and implementation challenges encountered by organisations?*

5.3.2 Research paradigm

One of the fundamental information requirements of the study was to understand the role of data in a data-driven organisation and explore the limitations of traditional BI systems in supporting a data-driven culture. These are the challenges that force data-driven organisations towards big data adoption. Big data adoption also has its own challenges that organisations have to address. During the investigation the researcher had to engage affected stakeholders to understand their perceptions and experiences, as suggested by multiple scholars (Alharahsheh & Pius, 2020; Chen & Hirschheim, 2004; Kroeze, 2012; Myers, 1997).

Big data adoption introduces emerging technological tools into the organisation and the experiences of the affected employees (research participants) regarding these tools needed to be considered because these were regarded as of significant value to the study. New data-driven technological innovations such as big data affect the culture of how data is managed and utilised in the organisation. The acceptance of such innovations can be determined by employees' perceptions and organisational culture as a whole. Research participants' background experiences, perceptions and feelings regarding the advantages and disadvantages of big data adoption over traditional BI had to be considered during the investigation.

Positivism can become a challenge when applied in a study where research participants' understanding of individual's experiences, feelings, and organisational culture plays a major role in the study (Alharahsheh & Pius, 2020; Kroeze, 2012). Therefore, the current study employed an interpretive research paradigm because the nature of the study considers research participants' background experiences, perceptions and feelings as important in generating useful knowledge. The research participants' perceptions and feelings about how big data adoption affects the culture of a data-driven organisation form a core part of the study.

5.3.3 Research methodology

The aim of the study was to understand issues and opportunities related to the use of data to inform decision-making in a data-driven organisation. This was done by investigating the challenges of both big data and traditional BI systems. The keyword understanding is very important to this study; qualitative research methods are considered relevant to studies where contextual descriptive understanding is of importance (Haddadi *et al.*, 2017; Orlikowski & Baroudi, 1991).

The study employed an interpretive qualitative research methodology. Research participants who were working with data and related systems were asked to describe the meanings they assign to the concepts related to big data, data-driven organisations, BI, and other data analytics concepts. A single-case qualitative research strategy was used to conduct the study because of the importance of understanding the organisation or the setting where the study was being conducted, in addition to paying careful attention to not generalising the study outcomes outside the study context. Many authors (Annansingh & Howell, 2016; Benbasat *et al.*, 1987; Gable, 1994; Yin, 1981) suggest the use of a case study for such research studies. Furthermore, the aim of the study was to develop a framework and case study research is suggested to be suitable for application in studies whose aim is to develop a framework or a model (Gustafsson, 2017).

5.3.4 Sampling

In the following sections aspects of the case study participants and sampling methods are discussed.

5.3.4.1 Study setting

The study was conducted in one of the financial service companies in South Africa. This company is in the process of assessing big data innovations to understand how these emerging data-driven tools can enhance the company's decision-making. The company has multiple BI departments where each department supports a specific business division. Each company division has its own BI systems. The chief information officer (CIO) of the company has suggested that each BI department should assess its systems and come up with a strategy to enhance the division's utilisation of data to improve the company's data-driven culture. The adoption of big data has been considered by most BI departments as the next innovation that will assist the company in improving its data-driven decision-making.

5.3.4.2 Research participant selection

The investigation of a framework for creating a hybrid DD-DSS requires the researcher to gather in-depth information and understanding of both big data and traditional BI systems from the research participants who work in the environments where these systems are used. A purposeful criterion-based sampling strategy was used to identify research participants for the current study. This sampling strategy was chosen because it suggests the selection of participants based on their knowledge and experiences about the investigated phenomenon (Coyne, 1997; Suri, 2011). A criterion that was used to select research participants was based on whether or not the employee worked in a BI department; and only employees who worked in BI departments and assumed roles relevant to the study were invited to participate in the study.

5.3.5 Data collection

Several data collection methods can be used in a qualitative research study. A list of commonly used methods in IS includes individual or group interviews, focus groups and observations (Kaplan & Maxwell, 2005; Lethbridge *et al.*, 2005). Semi-structured interviews were considered more relevant to the current study and were used to collect research data. A semi-structured interview is a data collection strategy that involves an interchange of words between the interviewer and the interviewee where the interviewer formulates questions to be answered in an interview, while allowing the interviewee to highlight issues he or she believes may be of importance to the conversation (Hove & Anda, 2005). The reason why semi-structured interviews were considered more applicable to this study is their ability to manageably allow new ideas and concepts to emerge from both the interviewer and the interviewee during the interview discussion (Fusch & Ness, 2015). Such flexibility fits well into a qualitative interpretive study (Horton *et al.*, 2004).

5.3.5.1 Semi-structured interview guide testing

An interpretive qualitative study cannot guarantee that the success of pilot testing will ensure the success of the actual data collection process for the research study because, by its nature, interpretive research seeks to understand individuals' contextual experiences (Myers, 2002). For this study, pilot testing was done to validate the understanding and the structure of the interview questions. To validate and ensure that the language used in the interview questions was understandable, one research participant was asked to go through the interview questions and provide feedback on the structure and whether or not the questions were easy to understand. The reviewer confirmed that all questions and the flow of the semi-structured interview guide were in order.

5.3.5.2 Research participants

Based on the sampling method used in this study, 35 research participants were identified as eligible for participation and a request for participation was sent via email to all these prospective research participants. Twenty-one prospective participants indicated that they would like to be part of the study. The details of the participants, including the participant's identification code, job title, departmental code and the date of the interview, are depicted in Table 14. To maintain anonymity, each research participant was allocated an identification code (for example RP1) and the department's name was allocated a code as well (for example OML1).

Table 14 – Research participants

Participant Identification Code	Job Title	Department Identification Code	Interview Date
RP1	Business Analyst	OML1	27 July 2020
RP2	Lead Business Analyst	OML2	27 July 2020
RP3	Senior Analyst Programmer	OML2	27 July 2020
RP4	Data Scientist	OML2	29 July 2020
RP5	Data Analyst	OML2	29 July 2020
RP6	Senior Manager	OML2	31 July 2020
RP7	Business Specialist Manager	OML3	31 July 2020
RP8	Database Administrator	OML2	03 August 2020
RP9	Lead Analyst Programmer	OML2	03 August 2020
RP10	Lead Business Analyst	OML2	03 August 2020
RP11	Senior Manager	OML1	03 August 2020
RP12	Senior Business Analyst	OML2	03 August 2020
RP13	Analyst Programmer	OML2	04 August 2020
RP14	Analyst Programmer	OML2	05 August 2020
RP15	Senior Business Analyst	OML3	07 August 2020
RP16	Senior Business Analyst	OML4	07 August 2020
RP17	Manager	OML2	07 August 2020
RP18	Analyst Programmer	OML2	11 August 2020

Of the 21 prospective participants who had shown interest in participating in the study, 18 research participants took part. This yielded a semi-structured interview response rate and participation of 86%.

5.3.5.3 Interview questions

The semi-structured interview questions were grouped according to their supporting research study objective. The semi-structured interview included 10 questions, of which 9 were formulated to gain insights into the information required to investigate the specific objectives of the study and answer research questions. The interview questions and their objectives are depicted in Table 15.

Table 15 – Interview questions

Objectives	Interview question
• Establish the definition of a data-driven organisation. • Determine the research participant's understanding of the role of data in decision-making.	1. In your opinion, what is a data-driven organisation? 2. What is the role of data in your organisation's decision-making?
• Establish the contextual definitions of BI and big data analytics. • Understand the components of a typical BI system.	3. In your opinion, what is business intelligence (BI)? 4. What are popular components of a traditional business intelligence (BI) system? 5. How would you define big data? 6. What is meant by big data analytics?
• Determine the research participant's perception of the challenges and shortcomings of both traditional BI and big data analytics. • Establish the research participant's thoughts on the hybrid data-driven decision support system.	7. What are the known challenges of traditional business intelligence (BI) systems? 8. What are the challenges of adopting and using big data analytics systems? 9. Given an opportunity to integrate traditional business intelligence (BI) with big data analytics, which components would you consider combining from both systems, and why?
• Gather any additional information that the research participant would like to share.	10. Are there any general comments that you would like to add?

The aim of 1 additional question was to obtain any general information that the research participants wanted to share about the study or anything related to big data and BI.

5.3.5.4 Interview process

Because of the regulations to deal with the COVID-19 pandemic, interviews could not take place in person and the semi-structured interview questions were created and distributed electronically via Microsoft Forms application. This enabled research participants to capture their feedback electronically in a free text format. The electronic semi-structured interview questions included a research participant consent section that required all research participants to signal their agreement to participate in the study, given the study background and participation conditions. The semi-structured interview questions allowed participants to express their own understanding in their own terms when answering the questions. This made it easy for participants to elaborate on answers and could also indicate if some questions were misunderstood, so that the researcher could follow up to clarify these if necessary.

The semi-structured interview questions were sent to all prospective research participants who had indicated their interest in being part of the research and they were asked to provide answers to the interview questions within three weeks, between 17 July 2020 and 7 August 2020. One research participant indicated difficulty in completing the interview questions and requested a one-week extension. Since the interview feedback was already in a text format, no transcribing was required; the data were simply exported into Microsoft Excel and were analysed.

5.3.6 Data analysis

A thematic data analysis technique was used to identify themes from the collected qualitative data. Thematic analysis is a process of organising, examining and identifying the patterns within a data set to identify common themes that can be used to interpret the phenomenon studied (Alhojailan, 2012; Braun & Clarke, 2006). This data analysis technique was chosen because of its emphasis on data interpretation and understanding through the application of data coding and theme identification. Alhojailan (2012) suggests that thematic analysis is useful in understanding the research participants' explanations of the events based on their behaviour and experiences. Thematic analysis offers the flexibility of adopting either inductive or deductive approaches to data analysis. In an inductive approach, data codes and themes are derived from the collected data, whereas in a deductive approach, the researcher defines themes before data collection and analysis. It is possible to use a hybrid approach where the two methods are applied complementarily (Fereday & Muir-Cochrane, 2006).

A hybrid approach was used to analyse the data collected from semi-structured interviews. Some themes were defined from the literature study of BI and big data. Themes that emerged from the collected data were also considered to enhance the contextual interpretation of different aspects of the study as per their applicability in the study setting. An iterative approach that Braun and Clarke (2006) used in their study was used to analyse data. A detailed data analysis method for this study is described next.

5.3.6.1 Inductive data analysis

Qualitative data analysis was done to interpret, understand and explain qualitative data that were collected for the study. A qualitative data analysis technique typically involves uncovering the relationships between themes (Pope *et al.*, 2000; Sant, 2019). Semi-structured interview questions were formulated and Microsoft Forms application was used to collect feedback electronically from the selected research participants.

The inductive data analysis approach was used to interpret and understand research participants' feedback on interview questions. This approach was chosen because it is flexible and allows the researcher to address ideas emerging from the research during data collection. An iterative process (Figure 12) that includes five steps was used to conduct data analysis. The five steps of the process are *data preparation*, *thorough reading*, *data categorisation and coding*, *data filtering and rearrangement*, and *data categorisation and coding revision*.

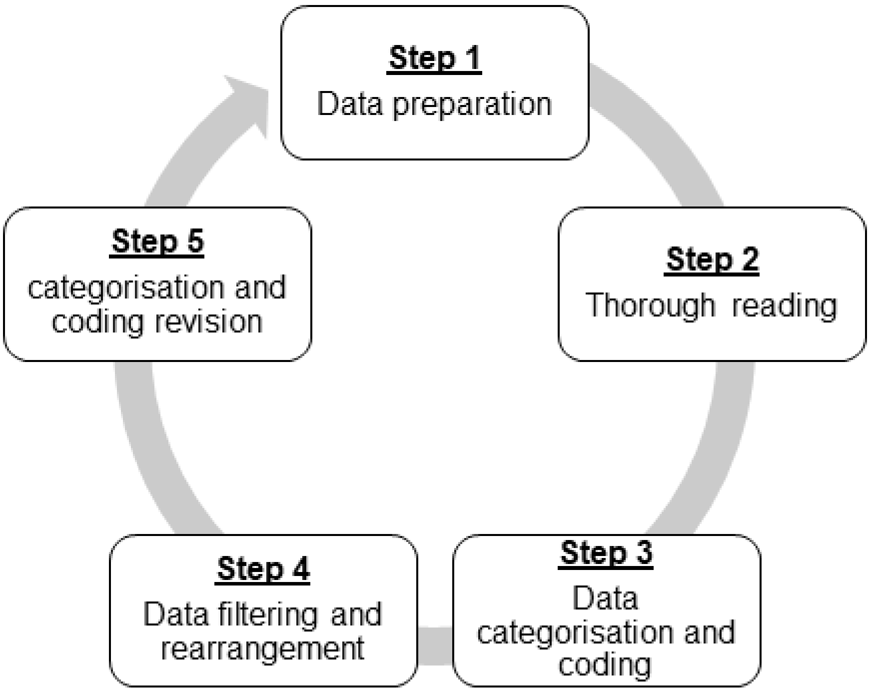

Figure 12 – Data analysis process (Braun & Clarke, 2006)

***Step 1:* Data preparation.** This step involved cleaning and formatting the data into a format that could be read easily. Research participants' feedback was exported from Microsoft Forms into Microsoft Excel. Answers to interview questions were then split into multiple worksheets, each worksheet containing answers to one question. The fonts were changed accordingly to improve the readability of the feedback. A copy of the original Microsoft Excel file that was exported from Microsoft Forms was backed up into Microsoft Office OneDrive and Google Drive to ensure these data were safely stored for future reference.

***Step 2:* Thorough reading.** Following data preparation, all research participants' answers to each interview question were read thoroughly to understand research participants' opinions and insights into interview questions. This is the step that Pope *et al*. (2000) calls data familiarisation because it entails iterative reading of responses to interview questions. Generally, this step enables researchers to familiarise themselves with the collected data. While answers to interview questions were read, some spelling mistakes were identified and corrected. Different abbreviations were also written in full to enhance the readability of the feedback.

***<u>Step 3:</u>* Data categorisation and coding.** The categorisation of data was done according to primary and emerging themes, which were then linked to one or more research questions and objectives. Manual data coding was done to identify commonalities between data aspects. Manual coding was considered the best choice because it is flexible and does not require any technical knowledge of the coding software application.

***<u>Step 4:</u>* Data filtering and rearrangement.** Some lines of text were not coded because they were not considered valuable to the investigation of the research questions. These included answers that were perceived not to be related to any of the research objectives. In addition to data that needed to be filtered out, some lines of text were rearranged, where the researcher felt that an answer to another question was given as part of a different question. For those lines, text was then moved into the relevant worksheet where the rest of the answers to that specific question were listed.

***<u>Step 5:</u>* Data categorisation and coding revision.** After thorough reading of text and code identification, some codes and categories needed to be combined, in addition to those that needed to be redefined. As a result, some codes were combined and new codes were defined.

5.3.7 Ethics and anonymity

Ethics and morality are important in a qualitative research study, especially where qualitative data are being collected, because the discussion that occurs during data collection might involve opinions that need to be strictly protected (Mingers & Walsham, 2010). The current research study attempted to address all the possible ethical considerations required to protect participants' views and ensure that the study would not cause any harm to the research participants.

To ensure that the time spent while answering the interview questions did not negatively affect research participants' activities, research participants were given a fair amount of time (three weeks) to answer the interview questions electronically. The research participants were also informed that participation in the study was purely voluntary. A prolonged period of time for answering the interview questions was aimed at ensuring that participants would be able to answer the interview questions in their own time and at their preferred speed. The background and context of the research study were included in the invitation to participate, to make it easy for research participants to decide whether or not they were interested in being part of the study.

The first section of the interview question form included the context of the research study as well as the consent agreement that every participant had to sign before proceeding to answering the questions. The names of participants are not disclosed in the report on the study findings. Codes are used to anonymise participants' names and their employment details, such as the company department where the research participant works (see Table 14). All research participants were treated in a fair and respectful manner and no group communications were held, as group communications could disclose the names of people who participated in the study. Every communication was personalised and occurred directly between the researcher and individual research participant.

5.4 Chapter Summary

The purpose of this chapter was to describe the research methodology and design of the study. An overview of IS research paradigms and methodologies was given in the initial sections of the chapter. The two commonly used paradigms in IS research are positivist and interpretivist worldviews (section 5.2.1). The IS research literature suggests that there is consistent domination of positivist studies in IS even though the application of interpretivism is increasing. The research paradigm gives insight into the choice of research methodologies that can be followed in conducting IS research. The current study is qualitative and high-level descriptions of the common qualitative research methodologies were summarised in section 5.2.2.

The details of a research methodology and the design of the current study were discussed in section 5.3. The current study employed an interpretive research paradigm because the nature of the study required the researcher to interact with participants to understand their opinions on the concepts involved in addressing the research question. A case study qualitative research methodology was used to conduct the study because every aspect of the investigation needed to be considered in its original setting – the organisation where the study was being conducted.

A purposeful criterion-based sampling strategy was used to identify research participants for the current research study (section 5.3.4). Semi-structured interviews were used as a data collection method in the study. Microsoft Forms was used as a data collection instrument; research participants were asked to answer interview questions in Microsoft Forms. The chapter was concluded with an overview of ethical considerations (section 5.3.7), which were addressed to enforce the morality of the study, in addition to protecting research participants' opinions and data.

Part IV – Data Analysis and Evidence

Part IV (Figure 13) contains Chapter 6 whose aim is to present research data findings.

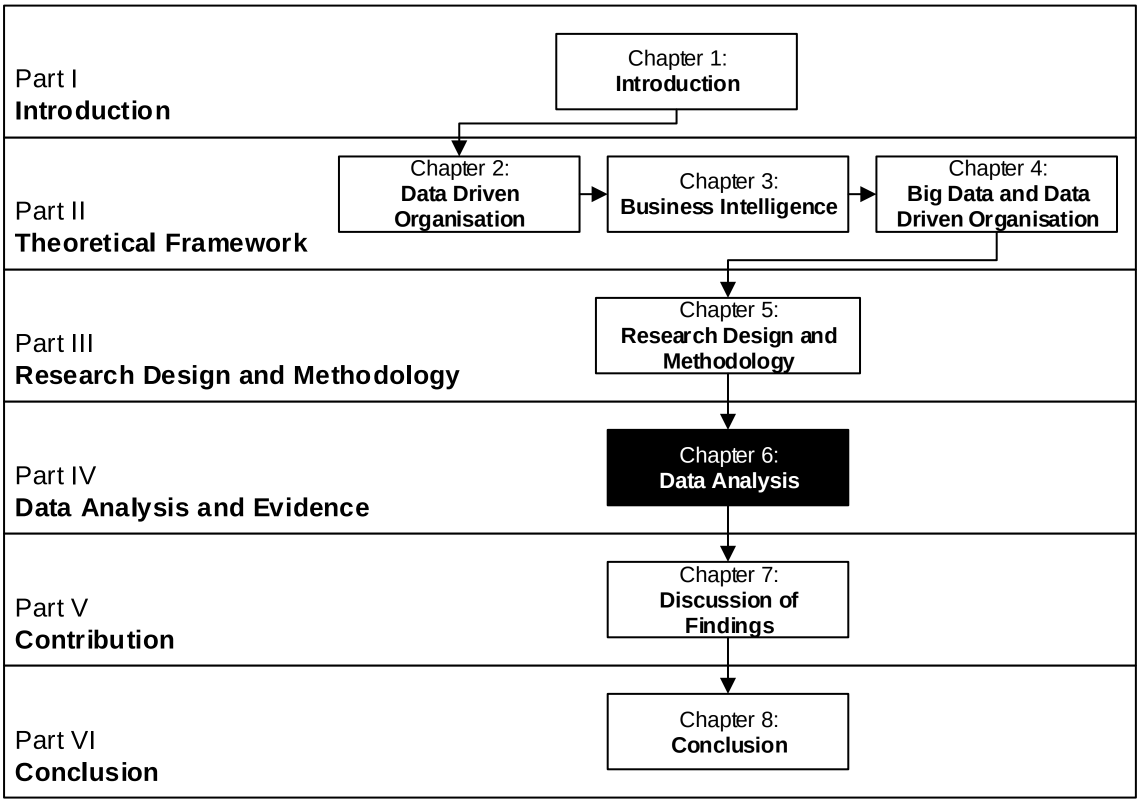

Figure 13 – Part IV outline

Chapter 6: Data Analysis

6.1 Introduction

The purpose of this chapter is to discuss data findings resulting from the analysis of research data. A thematic data analysis technique was used to identify themes from the collected qualitative data. A hybrid thematic approach (section 5.3.6) involving both inductive and deductive thematic techniques was used to analyse data that were collected from semi-structured interviews. Primary and emerging themes were identified and are presented in this chapter. Primary themes are constructs of investigation the researcher identified before collecting and analysing the research participants' answers to interview questions. Emerging themes are those constructs that were identified during data analysis.

Based on research participants' feedback and the literature review of this study, research acquisition pillars were identified as a guide to the interpretation and presentation of data analysis findings. The context in which each of the identified pillars gives insight into the research questions is summarised in section 6.2.

A summary of the contents of this chapter is outlined in Table 16.

Table 16 – Chapter 6 outline

Section	**Section description**	**Sub-section**	**Sub-section description**
6.1	Introduction		
6.2	Research Data Acquisition Pillars		
6.3		6.3.1	Definitions
		6.3.2	Value
		6.3.3	Challenges
		6.3.4	Architecture
		6.3.5	General comments
6.4	Summary of Findings	6.4.1	Definition

Section	Section description	Sub-section	Sub-section description
		6.4.2	Value
		6.4.3	Challenges
		6.4.4	Architecture
6.5	Chapter Summary		

6.2 Research Data Acquisition Pillars

The considerations and suggestions in the literature were used to guide the acquisition of data for the study. The purpose of this section is to describe how the literature study findings were used to inform the collection of the study data. The adoption of a data-driven culture and related technologies can have a significant influence on the way employees do their work. The suggestion is that employees need to be educated about the reason why data-driven innovations are being considered by the organisation and what organisational benefits are expected to result from these innovations (Provost & Fawcett, 2013; Watson, 2016). Therefore, it was considered critical to assess the research participants' understanding of the core concepts that are related to big data analytics and BI. The *definition* pillar was then used to guide the investigation of research participants' understanding and opinions about what the BI and big data analytics concepts mean in the context of a data-driven organisation.

BI and big data analytics studies suggest that there are many benefits linked to the adoption of DD-DSS in support of an organisational data-driven culture (Brynjolfsson & McElheran, 2016; Sun *et al.*, 2015; Wixom & Watson, 2010). However; these benefits vary from one organisation to another, depending on different organisational factors. To understand the perceived benefit of data and related technological systems in the natural setting of this study, the *value* pillar was used to guide data acquisition and interpretation of the benefits linked to BI and big data analytic systems.

Generally, the adoption of technological innovations involves challenges that organisations need to address. Big data adoption and the attempt to integrate traditional BI and big data analytics are no exception. Data analytics studies have reported that the adoption of big data and the attempt to integrate traditional BI and big data analytics involve some challenges (Engelbrecht *et al.*, 2016; Knabke & Olbrich, 2011; Surbakti *et al.*, 2019). Details of these challenges are described in sections 2.5, 3.6, and 4.6. The integration of BI and big data analytic systems requires good assessment of both the strengths and weaknesses of these systems. The *challenges* pillar was used to determine, understand and interpret the contextual challenges reported by the research participants.

The study aim was to propose an integrated framework for hybrid DD-DSS. The integration of BI and big data analytic systems seeks to address the challenges experienced when using both systems. The research participants' feedback on how they think a hybrid DD-DSS can be created was of the utmost importance to the creation of an integrated framework. The literature also suggests that the quality and efficiency of any DD-DSS depends on the architectural setup of different elements of the system (Chan, 2013; Ong *et al.*, 2011; Winter, 2001). The *architecture* pillar was then used to guide the interpretation of the participants' feedback regarding the architectural components of hybrid DD-DSS.

The interview questions were categorised into four pillars, which guided how research data should be acquired and presented. As described above, these four pillars are *definition*, *value*, *challenges*, and *architecture*. The relevance of these pillars (represented by squares) to the study and the way in which they link to the research objectives (represented by ovals) are depicted in Figure 14.

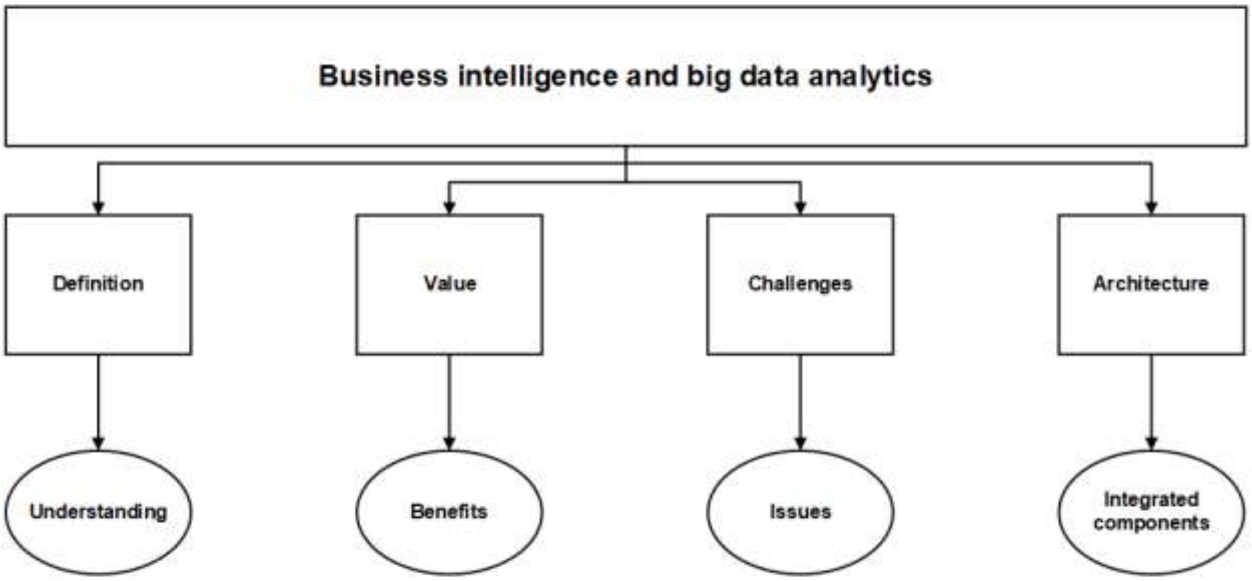

Figure 14 – Pillars of research data acquisition

In order to report the findings from this study, the primary and emerging themes extracted from data were allocated to these four pillars. *Primary themes* were predefined from the BI and big data literature. *Emerging themes* were identified from research participants' feedback during data analysis.

The findings of the primary and emerging themes for each interview question are discussed in subsequent sections of this chapter. The interview questions are grouped into the four pillars and reported according to the pillar into which they give insight. The exact order of the interview questions is shown in Table 18. Therefore, the order of the interview questions will be re-arranged based on the pillar that each question supports, but the numbering of questions will be kept unchanged for clarity.

6.3 Research Questions and Data Acquisition Pillars

As described in section 6.2, the four pillars were used to guide the investigation of the research objectives during data analysis. An overview of research questions, data acquisition pillars and interview questions is summarised in Table 17.

Table 17 – Research questions and data acquisition pillars

Research question	Pillar	Interview question
What is meant by data-driven organisation and data-driven decision-making?	*Definition:* Data-driven organisation.	1. In your opinion, what is a data-driven organisation?
	Value: The value of data in an organisation's decision-making.	2. What is the role of data in your organisation's decision-making?
What constitutes traditional business intelligence and big data analytics?	*Definition:* BI.	3. In your opinion, what is business intelligence (BI)?
	Architecture: A traditional BI system.	4. What are popular components of a traditional business intelligence (BI) system?
	Definition: Big data.	5. How would you define big data?
	Definition: Big data analytics.	6. What is meant by big data analytics?
What are the big data and business intelligence adoption and implementation challenges encountered by organisations?	*Challenges:* Traditional BI.	7. What are the known challenges of traditional business intelligence (BI) systems?
	Challenges: Big data analytics.	8. What are the challenges of adopting and using big data analytic systems?
	Architecture: Hybrid data-driven decision-support system.	9. Given an opportunity to integrate traditional business intelligence (BI) with big data analytics, which components would you consider combining from both systems, and why?

Table 17 shows the research questions, the related data acquisition pillar as well as the interview questions that were formulated to gather data that were used to investigate the research questions. The following sections present data findings as per the identified data acquisition pillars.

6.3.1 Definitions

The aim of the definition pillar was to determine the research participants' understanding of the concepts related to BI and big data analytics in the context of a data-driven organisation.

6.3.1.1 Data-driven organisation

To determine the participants' opinion on the definition of a data-driven organisation, the following question was asked:

Question 1: In your opinion, what is a data-driven organisation?

This question was answered by 18 participants, of whom 14 indicated that an organisation is a data-driven one only if its decision-making is underpinned by data analytics. While some participants did not specify the types of decisions that need to be underpinned by data analytics, 10 participants specifically mentioned that all strategic and operational decisions need to be based on data. Some participants (for example RP4) mentioned that all kinds of decisions must be grounded by data-based evidence, while other participants (for example RP1) specifically mentioned that strategic and operational decisions of a data-driven organisation are derived from data.

> RP4: *"… a data-driven organisation is the one that has embedded data collection, wrangling, analysis and decision-making into every aspect of its business, and uses the insights provided by the analysis of the data to perform evidence based decision-making at all stages …"*

> RP1: *"… a data-driven organisation is one that uses data by deriving useful insights to aid strategic and operational decision-making …"*

RP3: *"… an organisation that leverages the knowledge gained by analysing its available data store to correctly determine past behaviour and predict future behaviour …"*

RP5: *"… has a strong data governance …"*

RP10: *"… enabling these organisations to make informed decisions, and validating a course of action before committing to it … which will lead to commercial growth, stability and obtaining their objectives and goals."*

The primary and emerging themes for a definition of a data-driven organisation are summarised in Table 18.

Table 18 – Data-driven organisation themes

Primary themes	Emerging themes
1. An organisation where data drives decision-making. 2. Strategic decisions are guided by data. 3. Operational decisions are guided by data. 4. Predictive data analytics is essential. 5. Historical data analytics is essential.	1. Data governance is essential. 2. Data assist in tracking strategic objectives.

Five participants mentioned the importance of predictive data analytics as being fundamental to the decision-making process of a data-driven organisation, while the necessity of historical data analytics was also mentioned by 2 participants. Two emerging themes were identified, as depicted in Table 18. Three participants mentioned strict data governance as one of the fundamental values of a data-driven organisation. The ability to measure the success of an organisation through the use of data was also mentioned by 3 participants.

6.3.1.2 Business intelligence

To determine the research participants' perception of BI definition, the following question was asked:

Question 3: In your opinion, what is business intelligence (BI)?

With regard to the primary themes, most participants suggested that BI is the process of transforming data into information. Of the responses of 18 participants who answered this interview question, 13 suggested that the definition of BI is linked to the transformation of data into information. Eleven participants' responses mentioned that BI entails supporting business decisions through the use of data. Data analytics and reporting formed another theme mentioned in 10 participants' responses. Seven participants suggested that BI involves both processes and software tools.

> RP1: *"… Business intelligence is the use of analytical software to change data into useful insights which can assist an organisation with business decisions …"*

> RP18: *"… the software and services that transform data into actionable insights for business decisions …"*

> RP11: *"… the 'art' of using software to create information out of data …"*

From the emerging themes, 4 participants related the definition of BI to data and information quality; they mentioned that data and information quality are core to BI.

> RP12: *"… correct data to derive the right information at just the perfect time in the right format …"*

> RP3: *"… having the correct data available at the correct time and at the correct grain for interpretation to guide business decisions …"*

> RP9: *"… provide insights in an easy to understand and simple manner …"*

RP18: *"… collection of relevant data …"*

Primary themes and the emerging theme for a definition of BI are summarised in Table 19.

Table 19 – Business intelligence themes

Primary themes	Emerging themes
1. Means of transforming data into information. 2. Use of data to support decision-making. 3. Data analytics and reporting. 4. Software tools and processes used to transform data to information.	1. Data and information quality.

It was also suggested that ensuring information presentation quality standards helps in improving understanding of information (see quotation from RP9, for example).

6.3.1.3 Big data

To determine the participants' opinion on the definition of big data, the following question was asked:

Question 5: How would you define big data?

All 18 participants answered this question. The primary and secondary themes from their responses are summarised in Table 20. Fifteen participants suggested that big data are fundamentally measured and determined by their storage size (volume). Variety is another theme that 12 participants regarded as the determiner of whether or not data are big or just traditional. Eight participants indicated that velocity is an important factor or attribute of big data. Each of the keywords value (for example, RP14 quotation) and veracity (for example, RP16 quotation) was mentioned by 2 participants.

RP14: *"… that is very valuable due to insights that can be gathered from it …"*

RP15: *"… consolidate this data and use it to obtain insights across the value chain …"*

RP16: *"… data can be structured, unstructured or semi-structured …"*

RP18: *"… different sources inconsistencies uncertainty in data …"*

RP5: *"… Data which is too big to open or look at using traditional means …"*

RP4: *"… difficult to access and wrangle with existing BI tools …"*

RP6: *"… need special software or systems to be able to handle it …"*

RP11: *"… too large to be analysed through traditional means …"*

Primary themes for the definition of big data are summarised in Table 20.

Table 20 – Big data themes

Primary themes	Emerging themes
1. Volume. 2. Variety. 3. Velocity. 4. Veracity. 5. Value. 6. Beyond traditional BI tools' processing capacity.	

No emerging themes were identified from the definition of big data.

6.3.1.4 Big data analytics

To determine the participants' perception of the definition of big data analytics, the following question was asked:

Question 6: What is meant by big data analytics?

In relation to the primary themes, the majority of participants mentioned that the term big data analytics is used to describe the means through which big data are analysed and transformed into information. Out of 18 participants who answered this interview question, 15 suggested that big data analytics refers to the use of modern data analytic tools to analyse big data. The following phrases were extracted from some of the research participants' answers.

> RP1: *"… advanced analytical techniques to analyse large volumes of data from a range of sources which could be structured, semi-structured and unstructured …"*

> RP4: *"… using the relevant tools to perform data wrangling or analysis or modelling on these large data sets that you would not be able to do with current technology …"*

> RP9: *"… tools and methods … both structured and unstructured formats including traditional data and big data …"*

Answers from 8 participants suggested that there is no clear difference between traditional BI and big data analytics. Parts of the definitions given by these participants (examples below) did not indicate any clear distinction between traditional BI and big data analytics. Machine learning and Hadoop key terms also emerged from the definitions that the participants gave.

> RP8: *"… to build enterprise trends, customer behaviour and correlation …"*

> RP3: *"… manipulating data to gain business related insights …"*

The primary and emerging themes for the definition of big data analytics are summarised in Table 21.

Table 21 – Big data analytics themes

Primary themes	Emerging themes
1. Modern big data analytics tools. 2. Same as traditional data analytics. 3. Predictive modelling and analytics. 4. Data mining.	1. Machine learning. 2. Hadoop.

Predictive modelling and predictive analytics were mentioned by 5 participants when they defined big data analytics. Three participants mentioned data mining in their answers.

> RP9: *"… the use of machine learning, data mining predictive analysis and other advanced tools/methods to gain insight and make decisions …"*

> RP4: *"… using the relevant tools to perform data wrangling or analysis or modelling on these large data sets that you would not be able to with current technology. E.g. Apache Spark and Hadoop …"*

6.3.2 Value

The aim of the value pillar was to guide the investigation into the benefits that data add to the decision-making process of a data-driven organisation.

6.3.2.1 Value of data in an organisation's decision-making

To determine the participants' opinion on the role that data play in their organisation's decision-making, the following interview question was asked:

Question 2: What is the role of data in your organisation's decision-making?

All 18 participants answered this question and 11 of them suggested that data were used for decision-making in their organisation. Eight participants also suggested that the organisation used data to identify issues in addition to optimising its business operations and internal processes. Some of the phrases extracted from their responses are quoted below.

> RP4: *"… understanding of certain issues or segments to make tactical decisions …"*

> RP5: *"… optimise current processes …"*

> RP9: *"… improve and optimize the sales process and provide insight into areas that need attention …"*

The use of data to improve customer offerings and services was highlighted in 6 of the participants' answers. The use of data to perform predictive modelling also appeared in 5 of the participants' answers.

> RP5: *"… to create new opportunities …"*

> RP16: *"… making strategic business decisions that will grow the business and give it a competitive edge …"*

> RP3: *"… to manage operational business processes and to perform basic predictive modelling"*

> RP4: *"… we also use data to perform predictive modelling and A/B testing to decide which models to scale to a wider base …"*

The keywords for the primary themes and emerging theme for the value of data in a data-driven organisation are summarised in Table 22.

Table 22 – Themes for the value of data in a data-driven organisation

Primary themes	Emerging themes
1. Decision-making. 2. Business process and business operations optimisation. 3. Product offering and customer service improvement. 4. Predictive modelling.	1. Business opportunity identification.

Answers from 4 participants indicated that identification of business opportunities was another result of the use of data in the organisation.

6.3.3 Challenges

The challenges pillar was used to guide the investigation of the participants' perceptions of the limitations of both traditional BI and big data analytic systems.

6.3.3.1 Business intelligence

To determine the research participants' suggestions on the limitations of traditional BI, the following question was asked:

Question 7: What are the known challenges of traditional business intelligence (BI) systems?

The question was answered by all 18 research participants and all answers indicated that traditional BI systems are expensive. Seven participants indicated that traditional BI systems lack the self-service functionality.

RP10: *"… too expensive and hard to justify ROI of BI …"*

RP17: *"… expensive exploration tools …"*

RP15: *"… providing true self-service analytics …"*

RP13: *"… difficult for traditional BI to create self-service analysis"*

The primary and emerging themes for the challenges and limitations of traditional BI systems are summarised in Table 23.

Table 23 – Themes for traditional business intelligence challenges

Primary themes	**Emerging themes**
1. Expensiveness. 2. Lack of self-service functionality. 3. Poor performance due to processing of large data volumes. 4. Inability to process unstructured data. 5. Lack of flexibility.	1. Hinder business innovation. 2. Inability to analyse desperate data sources. 3. Lack of mobile devices support capability.

Six participants indicated that performance degradation resulting from increased data volumes is a shortcoming of traditional BI systems; this is in addition to these systems being unable to process unstructured data, as suggested by four participants.

RP16: *"… traditional business analysis systems struggle to process very large volumes of data; they often cause performance issues ..."*

RP14: *"… cannot process semi-structured and unstructured data …"*

It was also mentioned that users of traditional BI systems become attached to these systems and tend not to want to explore new systems. Four participants suggested that this was a challenge when it came to organisations' innovation.

RP4: *"… companies do not like to change from what they have …"*

RP12: *"… makes it almost impossible to shift strategies at the same pace the world is changing …"*

> RP3: *"… business tends to run 5 years behind the technology that is current …"*

Traditional BI systems are also perceived to have shortcomings when analysis of data has to be performed from disparate data sources, in addition to the lack of flexibility of these systems. Keywords for difficulty in analysing disparate data sources appeared in 3 answers. Another 3 answers also indicated that BI systems lack flexibility. Two participants suggested that traditional BI systems' lack of mobile device support is another notable limitation.

> RP13: *"… it is not flexible, not easy to make changes …"*

> RP16: *"… they are also not dynamic enough …"*

> RP9: *"… has not yet adapted to the mobile world and is not readily available on mobile devices …"*

6.3.3.2 Big data analytics

To determine the research participants' opinions about the challenges of big data analytics, the following question was asked:

Question 8: What are the challenges of adopting and using big data analytic systems?

From the primary themes (Table 24), it looks as if there is a key concern about the lack of skills required to support big data analytics in the organisation. The question was answered by 18 participants and 13 answers indicated that the lack of the required skillset was a challenge. Six participants indicated that big data analytic systems were expensive.

> RP16: *"… the lack of skill sets might also be a big challenge; there is not a lot of skilled workers for big data analytics …"*

RP14: *"… big data analytics skills are still high in demand so organisations lack some of those skills …"*

RP17: *"… Cost, it takes time to get all the required sources …"*

The primary and emerging themes for the reported challenges of big data analytics are summarised in Table 24.

Table 24 – Themes for big data analytics challenges

Primary themes	**Emerging themes**
1. Lack of skillset.	1. Validation of fast growing data.
2. Expensiveness.	2. Data security risk.
3. Computer power requirements.	3. Difficult to prove its ROI.
4. Legacy system integration.	
5. Undefined processes.	

It has also been indicated that big data analytics tools and systems are difficult to integrate with legacy systems. The perception is that there are no properly defined processes to guide the adoption of big data analytic systems.

RP2: *"… tools and process are not clearly identified or perhaps lacking …"*

RP11: *"… no use cases defined …"*

From the emerging themes (Table 24), 4 participants highlighted that the speed at which big data grow makes it difficult to validate its quality. Security risks were also identified as a considerable challenge by 4 participants. Three participants indicated that it was difficult to prove a return on investment (ROI) for the adoption of big data analytics.

6.3.4 Architecture

The architecture pillar was used to guide and investigate participants' opinions about the essential architectural components of a traditional BI system and how these components can be configured to work complementarily with big data analytic systems.

6.3.4.1 Typical business intelligence system

To determine the research participants' understanding of the components of a typical BI system, the following question was asked:

Question 4: What are popular components of a traditional business intelligence (BI) system?

All 18 participants answered this interview question and the majority (16) of them indicated that a data warehouse is one of the essential components of a traditional BI system.

> RP16: *"… the major component of a business intelligence system is a data warehouse, this is critical because that is where data is stored and organised before it is presented to the end user …"*

> RP3: *"… Data warehouse with data pipelines leading to and from for data collection, aggregation and export in varying forms …"*

> RP6: *"… the visualisation or front end layer which the business users interface with like your Power BI or Tableau as examples …"*

> RP2: *"… real time and online analytical processing of data …"*

Reporting tools were also identified by 15 participants, followed by analytical tools that were mentioned by 14 participants. See the phrases quoted from the feedback from RP6 and RP2 above, for example. The ETL keyword was mentioned by 10 participants.

> RP16: *"… ETL is also very important because this is the layer of the system, which extracts data from source systems, transforms and conforms data before it is stored …"*

A summary of primary and emerging themes related to the popular components of a typical BI system is given in Table 25.

Table 25 – Traditional business intelligence system themes

Primary themes	Emerging themes
1. Data warehouse. 2. Reporting. 3. Analytical tools. 4. ETL.	1. Data sources.

Eight participants mentioned that data sources are part of a traditional BI system.

> RP9: *"… multiple sources of data …"*

> RP10: *"… gathering data from various sources …"*

6.3.4.2 Hybrid data-driven decision-support system

The following interview question was asked to gather the research participants' suggestions on the way in which traditional BI and big data analytics system components can be complementarily integrated:

Question 9: Given an opportunity to integrate traditional business intelligence (BI) with big data analytics, which components would you consider combining from both systems, and why?

Seventeen of the 18 participants who were involved in the study answered this question. The responses show positive interest in the combination of components from the two DD-DSS types; 10 participants indicated that they believed that components from the two systems types could be combined. Seven participants recommended that ETL and a data warehouse remain part of the hybrid DD-DSS, but said that these should be used primarily for the processing of structured data.

> RP4: *"… traditional BI is the base, and appropriate use of big data and the related tools sit on top of that (and does not work as a replacement) …"*

> RP1: *"… data warehouse for structured data, big data capabilities for unstructured data … components as they will complement each other where both unstructured and structured data can be analysed efficiently …"*

An overview of themes for a hybrid DD-DSS is summarised in Table 26.

Table 26 – Themes for a hybrid data-driven decision-support system

Primary themes	Emerging themes
1. The two systems' architectural components can be combined. 2. Use ETL and data warehouse for structured data. 3. Use big data tools for unstructured data. 4. Use BI for operational reporting. 5. Use big data analytics for advanced analytics.	1. Explore cloud computing. 2. The two systems' architectural components should not be combined.

Four participants indicated that big data tools should be used where processing of unstructured data is required. The use of traditional BI for operational reporting was also mentioned by 3 participants, while the use of big data analytics tools for advanced analytics was suggested by 2 participants.

> RP5: *"… structured data that is used for day to day reporting and putting in the warehouse …"*

RP6: *"… you could dig a little deeper into that data looking for insights and then if needed even result in some advanced analytics being undertaken like a forecasting model, or a propensity model or even some machine learning being kicked off …"*

RP10: *"… structured data stored in the warehouse and the unstructured data stored on image processing …"*

It is worth noting that cloud computing is another keyword 2 participants mentioned in their answers. Two participants also suggested that traditional BI and big data analytic tools should not be combined; instead they can run as separate systems. One of them suggested that traditional BI systems can process big data.

RP15: *"… why should you have the same type of tools … both should have platforms where data are stored …"*

RP17: *"… to be honest, I do not see why traditional BI cannot be used for big data. With the right infrastructure and the right toolset, I believe traditional BI can accommodate big data …"*

6.3.5 General comments

The research participants were given an opportunity to share any general information or comments about the study. To gather such information, the following question was asked:

Question 10: Are there any general comments that you would like to add?

Most participants who answered this question emphasised the importance of data in an organisation and how using data can improve decision-making when handled appropriately and with the use of relevant tools. Few participants had an opinion about the universal definition of big data and its general perception by organisations.

RP3: *"… big data means different things to different people, depending on the business requirement; a one size fits all approach is not always helpful …"*

RP4: *"… big data is a big buzz word at the moment, but it needs to be carefully evaluated if it is really what you need to get the job done as it is not always an appropriate tool …"*

RP9: *"… big data is a dream, buzz word or goal used by organisations ..."*

RP17: *"… large volumes of data though are not necessarily better than smaller volumes of data. Data quality is very important and should not be overlooked …"*

The quotations above indicate the concerns and opinions related to how participants perceive the definition and expectations of big data in their organisation.

6.4 Summary of Findings

The findings of data analysis agree to a fair degree with BI and big data literature that is discussed in the literature study. The purpose of this section is to summarise the findings and identify their connection to the BI and big data literature. The findings are summarised using the four data acquisition pillars (definition, value, challenges, and architecture).

6.4.1 Definition

The literature suggests that a data-driven organisation is an organisation that uses data to derive insights that are used to inform the organisation's business decisions (Todorova & Hoeben, 2016; Wang *et al.*, 2019; Watson, 2016). The themes (Table 18) obtained from the definition of a data-driven organisation for this study also show that participants agree with this definition and this definition is therefore applicable to the organisation where the study was conducted.

BI may be defined differently by different people, depending on what it means in their context; there is no universal definition of BI (Olszak & Ziemba, 2012). It is for that reason that research participants were asked to provide their definition of BI. The themes (Table 19) of the research participants' definition suggest that BI refers to the use of tools and processes to transform data into information that is then used to support all the decisions of a data-driven organisation.

Big data can be defined by using 5Vs: *velocity*, *volume*, *variety*, *veracity*, and *value* (Cai & Zhu, 2015; Taylor-Sakyi, 2016). The big data definition themes (Table 20) also suggest that 5Vs are used to describe big data, from the research participants' point of view. Participants also indicated that big data cannot be handled by the traditional data-processing and analytics tools. Big data analytics was defined as a process of transforming big data into information that can be used to inform business decisions. The difference between BI and big data analytics is determined by the type of data that each of these systems can transform into information. For example, big data analytics can process all types of data, as opposed to BI systems that can only process structured data of limited volume.

6.4.2 Value

To determine the participants' understanding of how data add value to their organisation, they were asked to give their opinion on the value that data add to the organisation. The feedback received (Table 22) suggests that data are used to support the organisation's decision-making. Predictive modelling was suggested to be core to the use of data to identify business improvement opportunities and also to improve customer service offerings. This is in agreement with the literature studies (section 2.3), which suggest that data-driven organisations are likely to perform better than other organisations.

6.4.3 Challenges

The participants gave their opinions on the challenges and limitations of both traditional BI and big data analytic systems. The feedback received (Table 23) suggests that lack of flexibility, lack of mobile device support and limited support for self-service data analytics are among the core challenges that the organisation experiences in the use of traditional

BI systems. Other challenges identified include poor performance and high cost. Some studies have suggested the same challenges in relation to traditional BI systems (section 3.6).

On the other hand, big data analytic systems are perceived to pose their own challenges as well (Table 11). The lack of a skillset required to develop effective big data analytic systems is among the suggested challenges. Research participants also suggested that processing of big data requires more computer power than that of traditional structured data, which makes the adoption of big data more expensive. These challenges have also been mentioned in multiple literature studies, as discussed in section 4.6.

6.4.4 Architecture

To gain some insight into what participants regard as part of a traditional BI system, the participants were asked to identify the components of a typical tradition BI system. Their feedback suggested that source systems generate data that are then processed through ETL and stored in a data warehouse (Table 25). Analytics and reporting tools pull data from a data warehouse and users derive business insights from data through the use of these tools. The feedback from research participants agrees to the components of a typical traditional BI system architecture that many studies have suggested (section 3.2).

Research participants were also asked to give their opinion on how traditional BI systems could be integrated with big data analytics. Their feedback (summarised in Table 26) suggested that traditional BI components can still be used for basic data analytics and reporting, whereas advanced analytics, which requires unstructured data and large volumes of data, can be handled through the use of big data analytic tools.

6.5 Chapter Summary

The purpose of this chapter was to discuss and present the finding of the study's data analysis results. Four pillars (described in section 6.2) were used to guide the acquisition and analysis of data: definition, value, challenges and architecture.

The definition pillar was used to guide the investigation of the definition of BI and big data analytics concepts in the context of a data-driven organisation. The value pillar was used to gather information about the benefits of effective use of data in a data-driven organisation. The challenges pillar guided the investigation of the issues related to both BI and big data analytic systems. To understand the research participants' understanding of the essential components of a BI system, the architecture pillar was used.

The architecture pillar was also used to gather research participants' opinions on what components would constitute a hybrid DD-DSS. The interview questions were constructed to acquire data to investigate the research questions. The connection between research questions, data acquisition pillars and interview questions was summarised in section 6.3.

The findings of the definitions of a data-driven organisation, BI, big data, and big data analytics were discussed in section 6.3.1. The findings on the value that data add to the success of a data-driven organisation were detailed in section 6.3.2. BI and big data analytic systems challenges were summarised in section 6.3.3, after which findings related to architectural components of both BI and hybrid DD-DSS were discussed (section 6.3.4). An overview of the relationship between the data analysis findings and the study's literature survey was summarised in section 6.4.

Part V – Contribution

Part V (Figure 15) of the study contains Chapter 7, which discusses the details of the study findings.

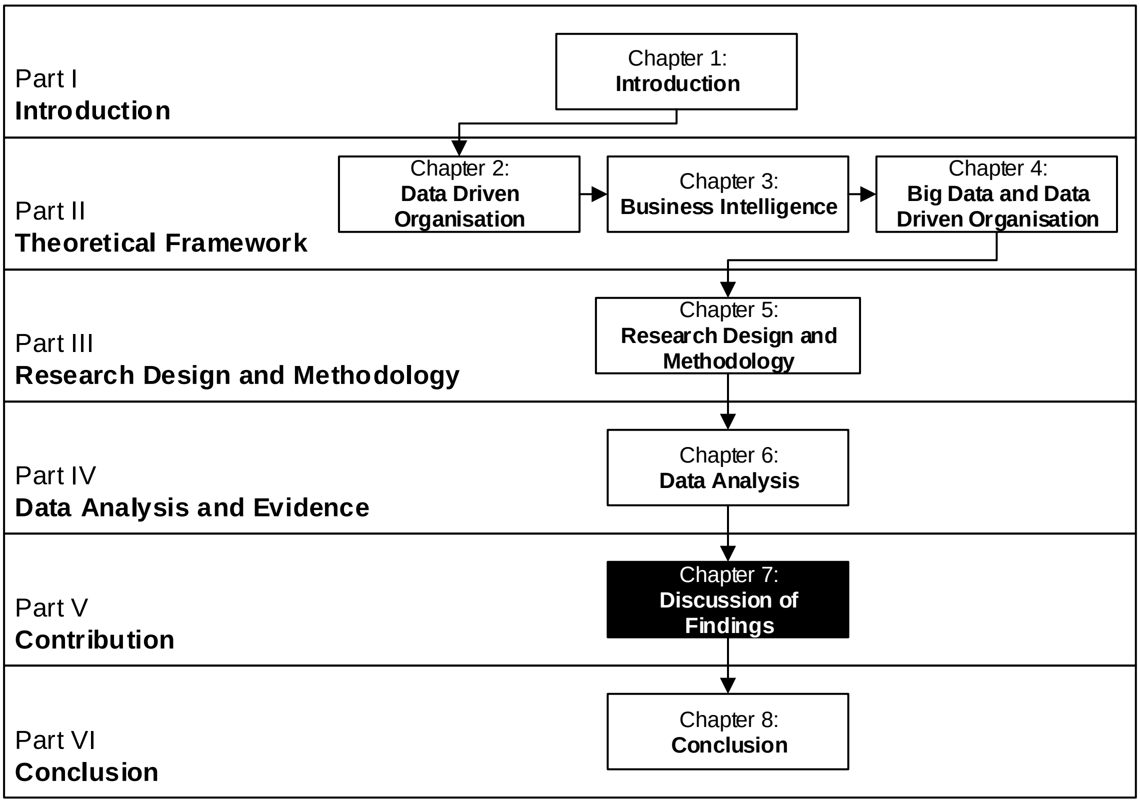

Figure 15 – Part V outline

Chapter 7: Discussion of Findings

7.1 Introduction

An organisation is perceived to be data-driven if it has fully adopted a data-driven decision-making culture (Wixom & Watson, 2010). The concept of data-driven decision-making and that of a data-driven organisation are closely related. BI and big data analytics use processes, tools and software services to convert data into actionable insights, which drive all business decisions of a data-driven organisation (Hedgebeth, 2007; Merendino *et al.*, 2018). The purpose of this chapter is to discuss the findings of the current study by combining the results from research participants' feedback and the findings from the study's literature review.

The first section (section 7.2) of this chapter describes the relationship between a data-driven organisation and data-driven decision-making in the context of the research participants. Findings pertaining to the correlation and variances between BI and big data analytics are discussed in section 7.3. Research participants listed the challenges of BI and big data analytics and the details of these identified challenges are also discussed in this chapter (section 7.4).

The closing section of this chapter is focussed on detailing a proposed framework for the adoption and creation of a hybrid DD-DSS. The details of each of the framework components are described in section 7.5.

A summary of the contents of this chapter is outlined in Table 27.

Table 27 – Chapter 7 layout

Section	Section description	Sub-section	Sub-section description
7.1	Introduction		
7.2	Data-driven Organisation and Data-driven Decision-making		

Section	Section description	Sub-section	Sub-section description
7.3	Business Intelligence and Big Data Analytics		
7.4	Business Intelligence and Big Data Analytics Challenges		
7.5	Framework for a Hybrid Data-driven Decision Support System	7.5.1	Theory application
		7.5.2	Business motivation
		7.5.3	Information requirements
		7.5.4	Supporting mechanisms
		7.5.5	Data attributes
		7.5.6	Supporting processes
		7.5.7	Hybrid DD-DSS architecture
7.6	Chapter Summary		

7.2 Data-driven Organisation and Data-driven Decision-making

Even though multiple scholars give different definitions of the terms and correlation between *data-driven organisation* and *data-driven decision-making*, there is a common theme that advocates that an organisation whose decision-making is underpinned by facts derived from data is considered to be a data-driven organisation (Lee, 2017; Merendino *et al.*, 2018; Olszak & Ziemba, 2012). Research participants gave feedback on their definitions and understanding of these two concepts from which a number of characteristics were identified. Convincingly, it was noted that the definitions given by the participants were aligned with those proposed by the literature.

While defining a data-driven organisation, one research participant suggested that

> *"A data-driven organisation is the one that has embedded data collection, wrangling, analysis and decision-making into every aspect of its business, and uses the insights provided by the analysis of the data to perform evidence based decision-making at all stages."*

This definition suggests that all stages or levels of decisions must be based on data. Another definition that was given by one research participant suggested that everyone in the organisation needed data to make decisions, not just managers and senior or executive managers. This definition suggested that

> *"A data-driven organisation would be an organisation that gives priority to having the correct data available when required to assist in making decisions whenever possible. This would not only be at a strategic level but also for the staff on the ground. Everyone needs the correct data at the right time to assist with decisions."*

Different levels of data-processing and usage may support different kinds of organisational decisions. While all research participants generally mentioned that *"all decisions"* of a data-driven organisation must be based on data, some participants mentioned both strategic and operational decisions as two types of decisions about which they are concerned. For example, one research participant mentioned that

> *"A data-driven organization is one that uses data by deriving useful insights to aid strategic and operational decision-making."*

Improved data quality and governance were also mentioned as the critical attributes of a data-driven organisation. A data-driven organisation definition included phrases such as:

> *"… is an organization that has a strong data governance … gives priority to having the correct data … ensures processes are in place for accurate data …"*

It is important to note that data quality processes are influenced by people.

While motivating the importance of senior management support for the adoption of a data-driven culture, Watson (2016) specifies that senior management must implement strategies that ensure that data-driven business processes are defined. Therefore, based on the participants' feedback, it can be recommended that such processes should include data governance and data quality processes. The motivation for the perceived value of a data-driven culture can improve the adoption of and adherence to data-driven business processes (Surbakti *et al.*, 2019). There should, therefore, be transparency about the perceived value of data and its related processes in the organisation.

To consolidate findings pertaining to a data-driven organisation and related concepts, Figure 16 shows the relationship between different attributes of a data-driven organisation. Figure 16 was constructed using the information provided by the research participants and the study's literature review. The literature emphasises the importance of organisations' senior management in communicating (transparency) and enforcing the culture of data analytics in the organisation. Employee positivity about data analytics is determined by their understanding of the business value of data. Both the literature and research participants' feedback suggested that data governance and data quality determine the effectiveness of data-driven decisions.

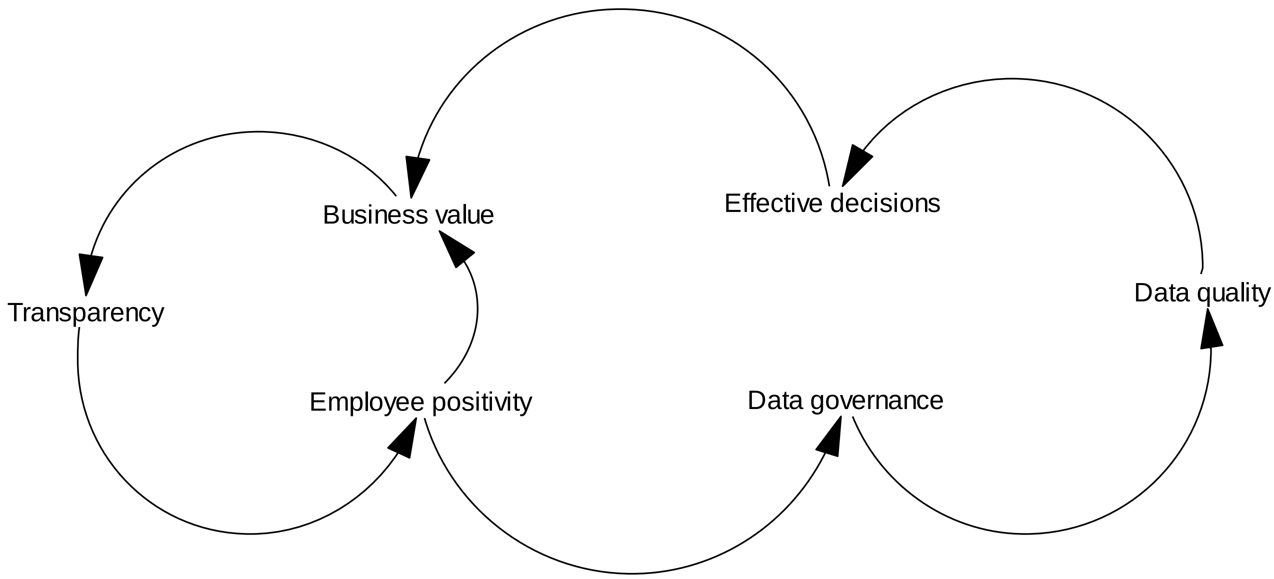

Figure 16 – Data-driven organisational attributes

As shown in Figure 16, data governance processes and standards influence data quality. Effective data-driven decisions are determined by the quality of data used to inform those decisions. Effective decisions are linked to improved business value. Transparency about the perceived value of data-driven decision-making is linked to improved employee motivation and positivity. Employees' positivity is a business value on its own because it enhances their ability to define and follow data governance processes. To summarise both research participants' feedback and the literature study, it can be stated that: *business decisions that are informed by data are referred to as data-driven decisions and an organisation that has entirely adopted a data-driven decision-making culture is referred to as a data-driven organisation*.

7.3 Business Intelligence and Big Data Analytics

BI and big data analytics both play a major role in a data-driven organisation because these systems are used to convert data into information that drives the business decisions of a data-driven organisation. In addition to the literature definitions of these important aspects of a data-driven organisation, research participants were asked to share their understanding of these two concepts.

When defining BI, some research participants stated that BI is the conversion of data into information, yet the following definitions (given by research participants) do not mention an organisation's intention to convert data into information:

> RP9: *"BI is about turning data into information..*
> RP11: *"Business intelligence is the 'art' of using software to create information out of data"*

On the other hand, some definitions (examples below) specified a reason why organisations strive to convert data into information.

> RP14: *"Business intelligence is the process of collecting data and storing it in a uniform manner in order to gain insights from the data by analysing and reporting on it. This enables the business to make better decisions."*

> RP2: *"The use of software and processes to turn raw data into meaningful information which can be used to make informed decisions…"*

It can be mentioned that research participants' feedback suggests that *BI is a collection of tools and processes that are used to convert data into information that give insights into the business decisions of a data-driven organisation.*

The definition of big data analytics was also investigated from the descriptions of research participants. Some participants generally suggested that big data analytics is the process of converting big data into information. Two examples of such definitions are given below:

> RP17: *"Doing analysis on large volumes of data to identify trends and to convert data to information."*

> P18: *"Collecting, organizing, and analysing big data."*

Similarly, concerning BI, some big data analytics definitions were detailed and some even indicated the tools and techniques involved in the conversion of big data into information.

> P10: *"Analysis of big data allows analysts, researchers and business users to make better and faster decisions using data that was previously inaccessible or unusable. Businesses can use advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics and natural language processing to gain new insights from previously untapped data sources independently or together with existing enterprise data."*

> P16: *"Big data analytics is the process of collecting, organising and analysing large volumes of data from different sources with the intention of discovering patterns or trends in the data. The information generated from big data is useful in making sound business decisions that will potentially grow the organisation."*

In summary, research participants' feedback suggested that *big data analytics is the transformation of big data into actionable insights, which are then used to inform the business decisions of a data-driven organisation.* It is worth noting that traditional data are also considered to be part of big data analytics. The literature definitions of big data analytic systems suggest that it entails the processing of big data to support the decision-making of the organisation (Balakrishnan & Rahul, 2018; Madden, 2012; Salinas & Lemus, 2017; Sun *et al.*, 2015). However, these definitions do not explicitly state that processing of traditional structured data is also part of big data.

7.4 Business Intelligence and Big Data Analytics Challenges

Many challenges are related to the transformation of data into information and knowledge for use in a data-driven organisation (Sun *et al.*, 2015). As part of this study, the research participants were asked to list the challenges that are specific to their setting and these are summarised in Table 28; the number of research participants who suggested a specific challenge is put in brackets.

It is important to note that the research participants' feedback shows that both BI and big data analytics systems are believed to be expensive. From the BI systems point of view, the lack of flexibility can be related to the inability of these systems to cope with increased volume and varying data formats, because inflexible systems may be difficult to scale.

Table 28 – Business intelligence and big data analytics challenges

Business intelligence	Big data analytics
• Expensive (8). • Lack of self-service functionality (7). • Poor performance due to processing of large data volumes (6). • Inability to process unstructured data (4). • Personal attachment to legacy systems that hinders innovation (4). • Limited analysis of desperate data sources (3). • Lack of flexibility (3). • Lack of mobile devices support capability (2).	• Lack of skillset (13). • Expensiveness (6). • Increased computer power requirements (4). • Limited legacy system integration (3). • Undefined processes (2). • Data validation difficulty (4). • Increased data security risk (4). • Undefined and unjustifiable ROI (3).

The lack of adequate skills for developing big data analytics systems forces organisations to spend money on reskilling and upskilling their employees to ensure that they are capable of assuming the necessary big data analytics roles. Taking the upskilling capital into consideration and given that the emerging big data technological tools are expensive, the entire big data analytics system workload may be even more expensive to develop and maintain (McAfee *et al.*, 2012; Nasser & Tariq, 2015).

7.5 Framework for a Hybrid Data-driven Decision Support System

The findings of this study, which include both feedback from research participants and the literature study, were used to create an integrated framework for designing a hybrid DD-DSS that includes components from both traditional BI and big data analytics. The application of the OIPT and FVM also guided the development of the framework. The application of the two theories is summarised next.

7.5.1 Theory application

The primary concern of the OIPT is underpinned by the belief that an organisation has to acquire and process as much data as possible to ensure that it can survive its environmental dynamics. According to the literature review findings and research participants' feedback, the core reason why data-driven organisations adopt big data is to ensure that they are able to exploit as much data as possible to assist them in making effective business decisions, as per the OIPT suggestion. The difference between big data and traditional data is determined by the attributes of data; therefore, the framework for hybrid DD-DSS should address the *data attributes* concepts to ensure that organisations are able to exploit the required data effectively to support their business decisions. Research participants' feedback indicated that the difficulty of data validation, which leads to big data quality problems, is one of the reasons why organisations are unable to derive value from big data. Data will be of good quality if the analytics results solve or answer specific decision-based questions (Cai & Zhu, 2015; Dawson & Van Belle, 2013). There should be alignment between data that the organisation acquires for specific decision-making requirements. Therefore, the framework for hybrid DD-DSS has to include the correlation between *data attributes*, *information requirements* and *business motivation* for both the required information and the data formats that underpin the required information.

Improved data quality also requires an organisation to define data quality processes that will be adhered to by both data-processing systems and information consumers. Without data quality supporting processes, organisational data can easily lead to poor decisions. The *supporting processes* were therefore considered important to the suggested framework. Lack of skills to develop and support big data analytic systems was highlighted as a major challenge by both research participants (section 6.3.3.2) and the literature review findings (section 4.6). Expensive tools and senior managers' lack of interest in funding big data projects were added to the list of factors hindering the success of big data projects. Hence the framework includes the supporting mechanisms element, which suggests how these issues should be addressed. The FVM suggests that a balance between the business problems that the adoption of big data is expected to solve and the readiness of the organisation to support big data adoption determines the success of the organisation's transformation towards big data analytics. Hence the framework includes the *supporting mechanisms* element as well as the supporting processes to ensure that the organisation is able to operate big data systems effectively.

The research participants' feedback indicated that a hybrid DD-DSS should be configured in such a way that traditional data and big data-processing elements can be loosely coupled. The research participants also listed architectural components they believe should be integrated from both traditional BI and big data analytics systems. The literature review specified that the quality and efficiency of any DD-DSS depends on the architectural setup of different elements of the system (Chan, 2013; Ong *et al.*, 2011; Winter, 2001). Therefore, the framework also includes a *hybrid DD-DSS architecture* element, which depicts how the integrated components should function, according to the research participants' feedback (section 6.3.4.2).

The identification of different elements of the framework was based on the importance of ensuring that a DD-DSS is aligned with the information requirements of an organisation, ensuring that proper skills required to support the creation of a hybrid DD-DSS are considered, essential data governance processes guiding the processing of different types of data are put in place and the required architectural components ensuring effective data processing are provisioned. Therefore, the proposed framework is divided into these elements: *business motivation*, *information requirements*, *supporting mechanisms*, *data* *attributes*, *supporting processes* and *hybrid DD-DSS architecture.* The details of the framework are graphically summarised in Figure 17.

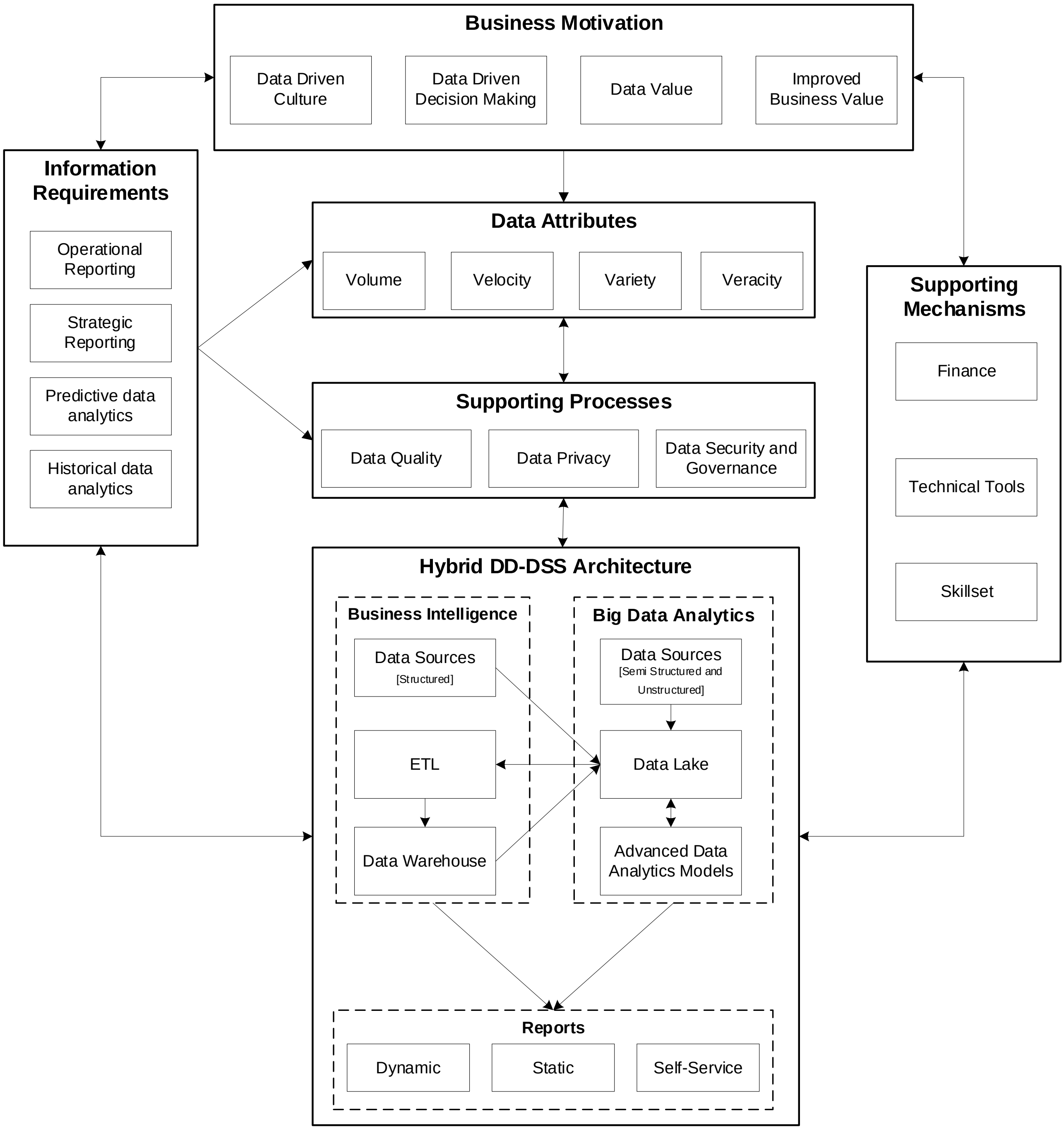

Figure 17 – Proposed hybrid DD-DSS framework

The descriptions of the different framework elements are presented in the following sections.

7.5.2 Business motivation

A business motivation element is focused on ensuring that every aspect of a data-driven organisation has a business value proposition. There should be a clear and communicable reason for the perceived business value of each of the attributes of a data-driven organisation. The business value of the effective use of data to improve a data-driven culture of an organisation that uses data-based facts in all of its decisions must be motivated. As suggested by a number of studies, senior and executive management commitment to ensuring that a data-driven culture is fully adopted is key (Adamala & Cidrin, 2011; Katal *et al.*, 2013; Wixom & Watson, 2010). The motivation element of the framework requires active involvement of the senior management to encourage the perceived value of a data-driven organisation. Correlation between different features of a data-driven organisation (section 7.2) is fundamental to communicating the business value of a data-driven culture.

7.5.3 Information requirements

The information requirements element is focused on ensuring that all decisions about data manipulation and processing are determined by the requirements of the end product – information. People who are working on a DD-DSS project need guidance on what the business problem is that the project seeks to solve. Each information requirement should have a business motivation for the business value that the information requirement seeks to realise. Typical information requirements that research participants specified in their feedback are operational and strategic reporting (section 6.3.2). Operational reporting is normally focused on day-to-day and weekly reporting of business operations at a detailed level. Strategic reporting is focused on high-level reporting of business outcomes and long-term planning is generally based on strategic reports. Predictive and historical data analytics were also suggested as some of the reasons why the organisation might require more data and information.

7.5.4 Supporting mechanisms

Supporting mechanisms include the assessment of budgets, instruments, appliances, tools and skills against the required information. The supporting mechanisms should also be motivated from a business value point of view. It was suggested that one of the reasons why senior management or sponsors hesitate to fund big data initiatives is the limited motivation of the perceived business value of big data (section 6.3.3.2). Each of the identified supporting mechanisms should have a sufficient business motivation that links the mechanism to the information requirements at a specific time. For example, if predictive data analytics requires data mining and machine learning skills, the business value motivation of predictive data analytics must be linked to employee upskilling in data mining and machine learning competencies.

7.5.5 Data attributes

The information requirements element is key to data attributes. All data being processed must have a business information requirement that has a qualified business value motivation. It is possible that different formats of data can be used to fulfil the same requirement. In that case the choice of processing one type of data over the other should be motivated, hence the data attributes element is linked to a business motivation element. The data attributes element is focussed on determining what forms of data are required to fulfil information requirements. The volume, velocity, variety, and veracity attributes of big data are considered in this element of the framework. The value, which represents the fifth V of the big data 5Vs, is dealt with at a business motivation element because it is concerned with the business value of data. The details of the 5Vs of big data and how big data differ from traditional data are discussed in section 4.2.

7.5.6 Supporting processes

Data have to be protected and data quality have to be monitored and assessed. The supporting processes element of a framework is focused on the practices, methods and guidelines that need to be defined to ensure that data quality and security are considered in every step of a data lifecycle. In most cases, these processes will vary depending on the type of data being processed to satisfy a specific business information requirement. Data quality concerns have been identified by a number of studies (Cai & Zhu, 2015; Nasser & Tariq, 2015) and the feedback from the research participants also indicated concern about data quality processes (see Table 24). Hence this element of a framework is key to the success of a data-driven organisation.

7.5.7 Hybrid DD-DSS architecture

The architectural components element of the framework proposes the configuration of technical elements of a hybrid DD-DSS. The architecture of a hybrid DD-DSS includes BI and big data analytics compartments. From a BI point of view, the components included are structured data sources, ETL and a data warehouse. A data warehouse is composed of data marts. The big data analytics compartment includes semi-structured and unstructured data sources, a data lake and advanced data analytics models.

A data lake is a common landing area for all the organisation's data, which arrive in a data lake in their raw form. These data include data forms that are structured (e.g. database and structured text files), semi-structured (e.g. documents and spreadsheets), and unstructured (e.g. images and videos). Importantly, when structured data are pulled from their data sources, these are not fed directly into a data warehouse through ETL, but are fed into a data lake. This is handled differently in a traditional BI system setting. Traditionally, structured data move from data sources straight to a data warehouse. The ETL part of the data warehouse transforms and formats the data into a grain that is suitable for a data mart where the data will be loaded. Data marts are different perspectives of data that reside in a data warehouse. The design of data marts can either be star schema or snowflake (Ong *et al.*, 2011; Salinas & Lemus, 2017).

Components of a big data analytics compartment involve the flow of all data formats into a data lake. Semi-structured and unstructured data are generally associated with big data (sections 4.2 and 6.3.1.3). Advanced analytics such as machine learning, text mining, natural language processing and predictive modelling can be used to process and analyse unstructured data. These examples of unstructured data-processing techniques were highlighted by research participants while defining big data analytics (section 6.3.1.4). A collection of tools and techniques that are used to process big data is represented by the advanced data analytics component of a big data analytics architectural section. In some cases, results of advanced analytics operations are pushed back to the data lake so that other system components such as data warehouse can have on-demand access to these results.

A decision on whether advanced analytics or ETL and a data warehouse should be used to identify and create relationships between data objects is dependent on the information requirements (section 7.5.3) and business motivation (section 7.5.2) framework elements. Such a decision should consider factors suggested in the supporting mechanisms framework element (section 7.5.4). There is no direct connection between advanced analytics models and the data warehouse except through a data lake. This is done to decouple a data warehouse and advanced analytics models so that there is no dependency between the two system components, because such dependency may increase the cost and the complexity of a hybrid DD-DSS.

Finally, at the far end of the system architecture are the reporting tools and processes. The reporting layer can source information from either a data warehouse or from the advanced data analytics models. Three types of reports were highlighted in the research participants' feedback, namely dynamic, static and self-service reports. Dynamic reports allow users to view information through an interface enabling them to change the structure and the presentation of the reports themselves. Self-service reports are similar to dynamic reports, except that they also allow users to access more data from data sources as opposed to being limited only to data that are pre-loaded into the report. Static reports have a rigid structure and presentation layout; users can only view reports' information and static reports are very limited in terms of allowing users to interact with the information.

7.6 Chapter Summary

In this chapter, the findings of the study were discussed. The relationship between a data-driven organisation and data-driven decision-making was described, based on the literature study as well as the insights obtained from the research participants' feedback (section 7.2). The importance of setting out guidelines that improve the quality of data at all levels of a data lifecycle is important, because the effectiveness of data-driven decisions is determined by the quality of data.

The difference between BI and big data analytics was explored in section 7.3. The key difference between BI and big data analytics is based on the type of data that these two systems ideally process. BI processes structured data to convert these into actionable insights that can be used as a base for the organisation's business decisions. Big data analytics processes both traditional (structured) and non-traditional (semi-structured and unstructured) data and converts these into information and knowledge on which business decisions can be based. The findings concerning the key challenges of both BI and big data analytic systems were summarised in section 7.4.

A discussion of the proposed framework for designing a hybrid DD-DSS was explored in section 7.5. The relevance of the OIPT and FVM theories in the creation of the proposed hybrid DD-DSS framework was described. The proposed framework is divided into these elements: business motivation, information requirements, supporting mechanisms, data attributes, supporting processes and hybrid DD-DSS architecture. A business motivation element (section 7.5.2) is focused on ensuring that every aspect of a data-driven organisation has a business value proposition. The information requirements element (section 7.5.3) is focused on ensuring that all decisions about data manipulation and processing are determined by the information and knowledge requirements.

The supporting mechanisms element (section 7.5.4) includes the assessment of budgets, instruments, appliances, tools and skills against the required information. The data attributes element (section 7.5.5) is focussed on determining the types of data that are required to satisfy business information requirements at any specific time. The supporting processes element (section 7.5.6) of a framework is focused on the practices, methods and guidelines that need to be defined to ensure that data quality and security are considered at every step of the data lifecycle. The last element of the framework is the architectural components element, which proposes the configuration of technical elements of a hybrid DD-DSS.

Part VI – Conclusion

Part IV contains Chapter 8, which provides the conclusion of the study.

Chapter 8: Conclusion

8.1 Introduction

The purpose of this chapter is to present the conclusion of the study and summarise the research question findings. The aim of the study was to investigate the elements of a framework that can be used to guide the development of a hybrid DD-DSS. The following main research question was formulated to guide the investigation:

What are the core elements of a conceptual framework for designing a hybrid data-driven decision-support system that combines big data analytics and traditional business intelligence?

The following sub-questions were formulated to guide the investigation of the main research question:

- *What is meant by a data-driven organisation and data-driven decision-making?*
- *What constitutes traditional business intelligence and big data analytics?*
- *What are the big data and business intelligence adoption and implementation challenges encountered by organisations?*

The study was broken down into six parts. Part I provided the introduction by describing the background of the study and the problem statement. The research questions and the overview of the research strategy were described in Chapter 1. Part II contained three chapters. This part of the study described a literature review for a data-driven organisation (Chapter 2), BI (Chapter 3), and big data analytics (Chapter 4). Part III contained Chapter 5, which described the research methodology and design of the study. Part IV contained Chapter 6, whose aim was to explore the analysis of research data and present the findings. The contribution of the study was explored in Part V. Chapter 7 discussed study findings forming part of the study contribution. Two appendices containing the interview questions (Appendix A) and research participants' feedback (Appendix B) have been attached to the study.

A summary of the contents of this chapter is outlined in Table 29.

Table 29 – Chapter 8 outline

Section	Section description	Sub-section	Sub-section description
8.1	Introduction		
8.2	Study Summary	8.2.1	Research questions summary
8.3	Summary of Contributions	8.3.1	Theoretical contribution
		8.3.2	Practical contribution
8.4	Suggestions for Future Research		

8.2 Study Summary

The purpose of this interpretive study was to investigate the framework for adopting a hybrid DD-DSS. The study was conducted in one of the leading insurance companies in South Africa. The company is in the process of assessing big data innovations to understand how big data can enhance its decision-making. The data-driven culture of the company is currently being serviced by multiple traditional business BI systems. Each company division has its own BI systems, which service the information requirements of the division. The CIO of the company has proposed that each division should assess its systems and come up with a strategy to enhance the division's utilisation of data to improve the company's data-driven culture. Most of the divisions have considered the adoption of big data analytics as their next innovation that will assist the company in improving its data-driven decision-making.

The literature review suggested that traditional BI systems have considerable shortcomings and are unable to fulfil all the data-processing needs in many data-driven organisations. A data-driven organisation is an organisation whose culture entails that all its business decisions are driven by data insights. The challenges that organisations face owing to the use of traditional BI systems are perceived to be addressed by big data analytics. Data-driven organisations are adopting big data systems and tools with the aim of gaining access to as much data as possible. Big data analytics also pose some challenges and many organisations are hesitant to adopt big data analytics fully because of the high failure rate of big data adoption projects. The literature suggested that one of the reasons for failure is limited understanding of how BI and big data analytics can be combined and there is therefore a need to investigate how the two systems can be integrated.

To investigate the elements of the framework that can be used to guide the creation of a hybrid DD-DSS, the interview data collection technique was adopted and used to collect study data from the research participants. The formulation of the interview questions was aligned to the following study objectives:

- *To explore what constitutes a data-driven organisation and how data improves the decision-making of a data-driven organisation.*
- *To understand the architectural components of both traditional BI and big data analytic systems.*
- *To propose a framework for creating a hybrid DD-DSS through understanding the common challenges and organisations' perceptions of the integration of traditional BI and big data analytic systems.*

8.2.1 Research questions summary

The primary research question for this research study was:

What are the core elements of a conceptual framework that can be used to design a hybrid data-driven decision support system that combines traditional business intelligence and big data analytics?

The literature study and feedback from research participants were used to investigate the framework elements. The OIPT and FVM theories guided the interpretation of the findings from these two data sources and these theories were used to guide the identification of the framework components. The proposed framework includes these elements: business motivation, information requirements, supporting mechanisms, data attributes, supporting processes and the hybrid DD-DSS architecture. These elements of the framework were discussed in section 7.5.

The rest of this section summarises the study findings that answer the research questions. The following sub-questions were investigated in an attempt to answer the primary research question of this study:

- *What is meant by a data-driven organisation and data-driven decision-making?*
- *What constitutes traditional business intelligence and big data analytics?*
- *What are the big data and business intelligence adoption and implementation challenges encountered by organisations?*

8.2.1.1 Research sub-question 1

The first question of the study was aimed at defining and describing the concepts of a data-driven organisation and data-driven decision-making. To gather this information from both the literature and research participants, the following question was formulated to guide the investigation:

What is meant by a data-driven organisation and data-driven decision-making?

According to the feedback given by the research participants, data-driven decision-making can be described through its relationship to a data-driven organisation. Data-driven organisations collect data from different data sources and process the data to transform these into information with the goal of using the information as a base for their decisions. These organisations are said to be data-driven because they use only data to drive their decision-making. Data-driven decision-making is consequently a process of using data to inform an organisation's decision-making. This is the opposite of the use of gut feeling and intuition as input to business decision-making.

Different levels and types of decisions can be informed by data, but there should never be an exception involving the use of intuition or gut feelings as a base for some decisions in a data-driven organisation. A data-driven organisation employs a data-driven decision-making culture at all levels and this culture applies to all types of decisions. Improved data quality is essential to data-driven decisions because high-quality data result in effective decision-making. Therefore, data governance and standards that guide the processing of data to ensure improved information quality must be central to all data processes of a data-driven organisation.

The importance of senior management support in enforcing a data-driven decision-making culture cannot be over-emphasised. The role of executive and senior management in enforcing and creating an environment where data are perceived as key to the organisation's success is vital. If all employees believe in the value of data, they will have a positive attitude to enforced data quality standards, which will result in effective decisions that would then determine the success of the organisation. Data-driven decision-making is core to understanding business dynamics and making informed decisions about the future of an organisation.

8.2.1.2 Research sub-question 2

Understanding of both traditional BI and big data analytics is fundamental to making the right decision on which components of these systems can be integrated to satisfy the organisation's information and knowledge requirements effectively. To gather this information, the following research question was used to guide the investigation of BI and big data analytic systems components:

What constitutes traditional business intelligence and big data analytics?

To gather information about BI components, research participants were asked first to define BI (section 6.3.1.2) and then to list the components of a typical BI system (section 6.3.4.1). The suggested definition of BI states that BI is a collection of software tools and processes that are used to convert structured data into information and knowledge that is used as a base for all the decisions of a data-driven organisation. The essential constituents of a BI system were perceived to be the structured data sources, ETL tools and processes and a data warehouse. Analytical and reporting tools were suggested to be part of a traditional BI system, but were suggested to be part of a typical big data analytics system as well, so these are not part of the differentiating factor between traditional BI and big data analytic systems.

Source systems, such as human resources, customer relationship management and enterprise resource planning systems, generate data that are in structured format. ETL systems extract data from the source systems, transform the data into a unified format and grain, and load these data into a data warehouse. A data warehouse is a central storage area for all structured data that the organisation needs for analysis and reporting. Analysis and reporting tools are then used to exploit data that are stored in a data warehouse.

To gather information about big data analytics components, research participants were asked to first describe the concept of big data analytic systems (section 6.3.1.4). Their feedback suggested that big data analytic systems use advanced analytics tools and techniques to process data that can be described using the 5Vs (section 4.4). Big data sources, a data lake and advanced data analytics models were suggested to be the essential components of a typical big data analytic system. Organisations are transforming their BI systems into big data analytics because they need the ability to process varying types of data at an acceptable speed, underpinned by the use of efficient tools.

8.2.1.3 Research sub-question 3

The creation of a hybrid DD-DSS should address the limitations of both traditional BI and big data analytics systems, in addition to leveraging the advantages of these two types of systems. The limitations of big data analytics systems are directly linked to the attributes of big data; hence, big data challenges were investigated. To explore the shortcomings of big data and traditional BI systems, the following question was formulated to guide the investigation:

What are the big data and business intelligence adoption and implementation challenges encountered by organisations?

Interestingly, when the two systems were compared to identify their shortcomings, it was recognised that both systems were perceived to be expensive. It was suggested that BI systems are expensive. The costliness of a traditional BI can be linked to its inability to scale effectively. The lack of flexibility of BI systems imposes an increased cost linked to the maintenance of these systems. Traditional BI systems were also suggested to be unable to cope with processing large volumes of data because their performance becomes poor when these systems have to process large data sets. The inability of a BI system to process semi-structured and unstructured data was reflected as a challenge that hinders organisations' ability to explore as much data as possible.

It was also revealed that BI systems lack support for self-service functionalities. The increased information demands of different organisations' stakeholders require systems that allow information consumers to exploit data themselves with minimal assistance from IT. Therefore, in view of the increased information demands, self-service data exploitation and reporting become big advantages. Traditional BI systems are losing out on this advantage. Organisations' increased use of mobile devices to do business presents a challenge to organisations where traditional BI systems are used, because these systems are suggested to lack support for mobile device functionalities.

From the big data analytics setting, the costliness of the related systems results from the capital required to invest in upskilling employees who need to develop and maintain big data analytics systems. The lack of personnel skills results in an organisation either having to hire expensive contractors or to spend money in upskilling its employees to acquire big data analytics skills. Additionally, big data analytics systems were suggested to require increased computer processing power so that they can perform optimally. The required capital to invest in implementing an optimal big data analytic system was also perceived as a challenge. Executive managers are hesitant to sponsor big data projects because of undefined ROI expectations linked to these projects.

A belief that big data analytics systems are difficult to integrate with legacy systems was also observed. In some cases, traditional BI systems are perceived to be the legacy systems and research participants therefore believe that big data analytics are difficult to integrate with BI systems. The emergence and adoption of big data require organisations to have a capability that assists their systems to cope with large volumes of data coming from varying sources that change frequently. As such, defining processes to validate the quality of these data were suggested to be a major challenge. Finally, the emergence of big data was said to elevate data security risks. The increased data volumes and formats present a challenge in securing such data because traditional data security standards are not fully applicable to big data.

8.3 Summary of Contributions

Theory significantly affects the interpretation and explanation of the research findings. An FVM and OIPT are the two theories that were used to interpret and explain the results of this study. The contribution of the study depends on its findings. The contribution of this study is twofold – a theoretical and a practical contribution.

8.3.1 Theoretical contribution

The literature suggests that big data adoption has proven to add value to data-driven organisations and organisations that are able to exploit as much data as possible perform better than others. However, organisations experience challenges when they transform their BI systems into big data analytics and these challenges are causing some organisations to lose interest in big data adoption (Katal *et al.*, 2013; Lee & Kang, 2015; Nasser & Tariq, 2015). Santoso (2017) suggests that one of the major challenges of big data is its inability to integrate with traditional BI systems.

Some studies have been conducted in an attempt to combine big data and traditional BI (Salinas & Lemus, 2017; Santoso, 2017; Sun *et al.*, 2015), but the list of studies of this nature is still limited (Santoso, 2017). This study proposed a framework that can be used to guide the integration of BI and big data analytics. The framework assists in assessing the components of both BI and big data analytics systems and making a case-by-case decision on which components can be used to satisfy the specific data requirements of an organisation. The study, therefore, contributes to expanding the existing literature in an attempt to integrate BI and big data analytics.

8.3.2 Practical contribution

The study was conducted in a company that is still in the process of assessing big data adoption with the intention of implementing big data analytics to enhance organisational information and knowledge creation. The benefits of a data-driven organisation, data-driven culture and data-driven decision-making, which were explored in this study, are based on feedback that employees of the company provided. The company can thus use the findings of this study to understand the opinions of its employees about its intentions to adopt big data solutions, as well as integrating big data analytics and traditional BI.

Senior management of the company have defined the company's data-driven strategy. They can compare the company's data-driven strategy with the findings of this study to assess if the employees' understanding of a data-driven culture is aligned with the company's data-driven culture. The company can use the framework that is proposed in the study to guide the integration of BI and big data analytics components. The suggested hybrid DD-DSS framework can also guide project managers in securing and motivating the resources required to deliver big data and BI projects. Research participants suggested that their inclusion in the study had assisted them in enhancing their perception of big data and BI. This feedback was captured under the general comments interview question.

8.4 Suggestions for Future Research

The adoption of big data has gained interest in many organisations across different industries (McAfee *et al.*, 2012). This study was conducted as a single-case study and further research is needed to conduct the study in multiple cases in the South African insurance industry so that insights from other companies can be obtained to improve the success rate of the adoption of big data solutions in the industry. The assumption is that there are more companies that might benefit from the use of the suggested framework; however, the framework might be more useful if it is modified to include factors from multiple cases. Moreover, the adoption of cloud services by the company where this study was conducted suggests a gap that requires further investigation to assess whether there are existing cloud services that can improve the integration of BI and big data analytics.

References

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis, 16*(1), 3-9.

Adamala, S., & Cidrin, L. (2011). Key success factors in business intelligence. *Journal of Intelligence Studies in Business, 1*(1), 107-127.

Aggarwal, C. C. (2015). *Data mining: the textbook*: Springer.

Al-Debei, M. M. (2011). Data warehouse as a backbone for business intelligence: issues and challenges. *European Journal of Economics, Finance and Administrative Sciences, 33*(1), 153-166.

Alharahsheh, H., & Pius, A. (2020). A review of key paradigms: Positivism VS interpretivism. *Global Academic Journal of Humanities and Social Sciences, 2*(3), 39-43.

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons, 60*(3), 285-292.

Alhojailan, M. I. (2012). Thematic analysis: A critical review of its process and evaluation. *West East Journal of Social Sciences, 1*(1), 39-47.

Alpar, P., & Schulz, M. (2016). Self-service business intelligence. *Business & Information Systems Engineering, 58*(2), 151-155.

Amankwah-Amoah, J., & Adomako, S. (2019). Big data analytics and business failures in data-rich environments: An organizing framework. *Computers in Industry, 105*, 204-212.

Annansingh, F., & Howell, K. (2016). Using Phenomenological Constructivism (PC) to Discuss a Mixed Method Approach in Information Systems Research. *Electronic Journal of Business Research Methods, 14*(1), 39-49.

Ariyachandra, T., & Watson, H. (2010). Key organizational factors in data warehouse architecture selection. *Decision support systems, 49*(2), 200-212.

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks, 54*(15), 2787-2805.

Balakrishnan, S., & Rahul, R. (2018). Big data in business intelligence. *CSI Communications magazine, Volume, 42*(8), 21-23.

Baskerville, R. L. (1999). Investigating information systems with action research. *Communications of the Association for Information Systems, 2*(1), 1-32.

Baskerville, R. L., & Wood-Harper, A. T. (1996). A critical perspective on action research as a method for information systems research. *Journal of information Technology, 11*(3), 235-246.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS quarterly, 11*(3), 369-386.

Berndtsson, M., Forsberg, D., Stein, D., & Svahn, T. (2018). *Becoming a data-driven organisation*. Paper presented at the Twenty-Sixth European Conference on Information Systems, Portsmouth, United Kingdom.

Bhanot, V. (2017). How to transform your bank into a data-driven organisation? *Journal of Digital Banking, 1*(4), 296-308.

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

Brichni, M., Dupuy-Chessa, S., Gzara, L., Mandran, N., & Jeannet, C. (2015). *Business intelligence for Business Intelligence: A case study at STMicroelectronics.* Paper presented at the 2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS).

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review, 106*(5), 133-139.

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data science journal, 14*(2), 1-10. doi:http://dx.doi.org/10.5334/dsj-2015-002

Cavaye, A. L. (1996). Case study research: a multi-faceted research approach for IS. *Information systems journal, 6*(3), 227-242.

Chan, J. O. (2013). An architecture for big data analytics. *Communications of the IIMA, 13*(2), 1-14.

Chatterji, A. K., Findley, M., Jensen, N. M., Meier, S., & Nielson, D. (2016). Field experiments in strategy research. *Strategic Management Journal, 37*(1), 116-132.

Chen, W., & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information systems journal, 14*(3), 197-235.

Coyne, I. T. (1997). Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? *Journal of advanced nursing, 26*(3), 623-630.

Dawson, L., & Van Belle, J.-P. (2013). Critical success factors for business intelligence in the South African financial services sector. *South African Journal of Information Management, 15*(1), 1-12.

De Villiers, M. R. (2005). Interpretive research models for Informatics: action research, grounded theory, and the family of design-and development research. *Alternation, 12*(2), 10-52.

Devlin, B. (2010). Beyond business intelligence. *Business Intelligence Journal, 15*(2), 7-16.

Dhar, S., & Mazumdar, S. (2014). Challenges and best practices for enterprise adoption of big data technologies. *Journal of Information Technology Management, 25*(4), 38-44.

Elgendy, N., & Elragal, A. (2014, July 16-20, 2014). *Big Data Analytics: A Literature Review Paper.* Paper presented at the Advances in data mining. applications and theoretical aspects: 14th Industrial Conference, St. Petersburg, Russia.

Elgendy, N., & Elragal, A. (2016). Big data analytics in support of the decision making process. *Procedia Computer Science, 100*, 1071-1084.

Engelbrecht, A., Gerlach, J., & Widjaja, T. (2016). *Understanding the anatomy of data-driven business models–towards an empirical taxonomy.* Paper presented at the Twenty-Fourth European Conference on Information Systems (ECIS), Istanbul, Turkey.

Erez, A., & Grant, A. M. (2014). Separating data from intuition: Bringing evidence into the management classroom. *Academy of Management Learning & Education, 13*(1), 104-119.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods, 5*(1), 80-92.

Fidel, R. (1984). The case study method: a case study. *Library and Information Science Research, 6*(3), 273-288.

Frické, M. (2009). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of information science, 35*(2), 131-142.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The qualitative report, 20*(9), 1408.

Gable, G. G. (1994). Integrating case study and survey research methods: an example in information systems. *European journal of information systems, 3*(2), 112-126.

Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces, 4*(3), 28-36.

Goldkuhl, G. (2008, December 14, 2008). *What kind of pragmatism in information systems research.* Paper presented at the AIS SIG Prag inaugural meeting, Paris, France.

Goles, T., & Hirschheim, R. (2000). The paradigm is dead, the paradigm is dead… long live the paradigm: the legacy of Burrell and Morgan. *Omega, 28*(3), 249-268.

Gupta, B., Goul, M., & Dinter, B. (2015). Business intelligence and big data in higher education: Status of a multi-year model curriculum development effort for business school undergraduates, MS graduates, and MBAs. *Communications of the Association for Information Systems, 36*(1), 23.

Gustafsson, J. (2017). Single case studies vs. multiple case studies: A comparative study. Halmstad, Sweden: Academy of Business, Engineering and Science, Halmstad University.

Haddadi, A., Hosseini, A., Johansen, A., & Olsson, N. (2017). Pursuing Value Creation in Construction by Research-A Study of Applied Research Methodologies. *Procedia computer science, 121*, 1080-1087.

Han, J., & Chang, K.-C. (2002). Data mining for web intelligence. *Computer, 35*(11), 64-70.

Hand, D. J. (2007). Principles of data mining. *Drug safety, 30*(7), 621-622.

Hawking, P., & Sellitto, C. (2010). *Business Intelligence (BI) critical success factors.* Paper presented at the 21st Australian conference on informafion systems.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464-472.

Hedgebeth, D. (2007). Data-driven decision making for the enterprise: an overview of business intelligence applications. *VINE, 37*(4), 414-420. doi:10.1108/03055720710838498

Horton, J., Macve, R., & Struyven, G. (2004). Qualitative research: experiences in using semi-structured interviews. In H. Christopher & L. Bill (Eds.), The real life guide to accounting research (pp. 339-357): Elsevier.

Hove, S. E., & Anda, B. (2005). *Experiences from conducting semi-structured interviews in empirical software engineering research.* Paper presented at the 11th IEEE International Software Metrics Symposium (METRICS'05).

Jamali, M., & Abolhassani, H. (2006). *Different aspects of social network analysis.* Paper presented at the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06).

Kaplan, B., & Maxwell, J. A. (2005). Qualitative research methods for evaluating computer information systems. In A. James G. & A. Carolyn E. (Eds.), *Evaluating the organizational impact of healthcare information systems* (pp. 30-55): Springer.

Katal, A., Wazid, M., & Goudar, R. H. (2013). *Big data: issues, challenges, tools and good practices.* Paper presented at the 2013 Sixth international conference on contemporary computing (IC3).

Keim, D., Qu, H., & Ma, K.-L. (2013). Big-data visualization. *IEEE Computer Graphics and Applications, 33*(4), 20-21.

Knabke, T., & Olbrich, S. (2011). Towards agile BI: applying in-memory technology to data warehouse architectures. *IMDM 2011–Proceedings zur Tagung Innovative Unternehmensanwendungen mit In-Memory Data Management*.

Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of healthcare information management, 19*(2), 65.

Kothari, C. R. (2013). *Research methodology: methods and techniques*. New Delhi, India: New Age International.

Krauss, S. E. (2005). Research paradigms and meaning making: A primer. *The qualitative report, 10*(4), 758-770.

Kroeze, J. H. (2012). *Interpretivism in IS–a postmodernist (or postpositivist?) knowledge theory.* Paper presented at the Proceedings of the Eighteenth Americas Conference on Information Systems, Seattle, Washington.

Kwanya, T. (2014). *Big data in land records management in Kenya: a fit and viability analysis.* Paper presented at the International Conference on Knowledge Management in Organizations.

Lai, P.-L., Su, D.-T., Tai, H.-H., & Yang, C.-C. (2020). The impact of collaborative decision-making on logistics service performance for container shipping services. *Maritime Business Review, 5*(2), 175-191. doi:10.1108/MABR-12-2019-0061

Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management, 36*(5), 700-710.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60*(3), 293-303.

Lee, J.-G., & Kang, M. (2015). Geospatial big data: challenges and opportunities. *Big Data Research, 2*(2), 74-81. doi:https://doi.org/10.1016/j.bdr.2015.01.003.

Lethbridge, T. C., Sim, S. E., & Singer, J. (2005). Studying software engineers: Data collection techniques for software field studies. *Empirical software engineering, 10*(3), 311-341.

Liang, T. P., Huang, C. W., Yeh, Y. H., & Lin, B. (2007). Adoption of mobile technology in business: a fit-viability model. *Industrial management & data systems, 107*(8), 1154-1169.

Lönnqvist, A., & Pirttimäki, V. (2006). The measurement of business intelligence. *Information systems management, 23*(1), 32.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in educational research, 16*(2), 193-205.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 16*(3), 4-6.

Martins, D., Ramos, C. M., Rodrigues, J. M., Cardoso, P., Lam, R., & Serra, F. (2015). *Challenges in building a big data warehouse applied to the hotel business intelligence.* Paper presented at the 6th Int. Conf. on Applied Informatics and Computing Theory (AICT'15), in Recent Research in Applied Informatics.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: the management revolution. *Harvard business review, 90*(10), 60-68.

McGlothlin, J. P., Madugula, A., & Stojic, I. (2017). *The Virtual Enterprise Data Warehouse for Healthcare.* Paper presented at the Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies, Porto, Portugal.

McVilly, K. R., Stancliffe, R. J., Parmenter, T. R., & Burton-Smith, R. M. (2008). Remaining Open to Quantitative, Qualitative, and Mixed-Method Designs: An Unscientific Compromise, or Good Research Practice? *International review of research in mental retardation, 35*, 151-203.

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research, 93*, 67-78.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management, 16*(3), 547-578.

Mills, R. J., Chudoba, K. M., & Olsen, D. H. (2016). IS programs responding to industry demands for data scientists: A comparison between 2011-2016. *Journal of Information Systems Education, 27*(2), 131.

Mingers, J., & Walsham, G. (2010). Toward ethical information systems: the contribution of discourse ethics. *MIS quarterly, 34*(4), 833-854.

Mukherjee, R., & Kar, P. (2017). *A comparative review of data warehousing ETL tools with new trends and industry insight.* Paper presented at the 2017 IEEE 7th International Advance Computing Conference (IACC).

Myers, M. D. (1997). Critical ethnography in information systems. In A. S. Lee, J. Liebenau, & J. I. DeGross (Eds.), *Information systems and qualitative research* (pp. 276-300): Springer.

Myers, M. D. (1999). Investigating information systems with ethnographic research. *Communications of the Association for Information Systems, 2*(1), 1-20.

Myers, M. D. (2002). Qualitative research in information systems: a reader. *MIS quarterly, 21*(2), 241-242.

Nasser, T., & Tariq, R. (2015). Big data challenges. *Journal of Computer Engineering & Information Technology, 4*(3), 1-10.

Negash, S. (2004). Business intelligence. *Communications of the Association for Information Systems, 13*(1), 177-195. doi:10.17705/1CAIS.01315

Niraj, R., Gupta, M., & Narasimhan, C. (2001). Customer profitability in a supply chain. *Journal of marketing, 65*(3), 1-16.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*: Oxford university press.

Nonaka, I., & Toyama, R. (2005). The theory of the knowledge-creating firm: subjectivity, objectivity and synthesis. *Industrial and corporate change, 14*(3), 419-436.

Olszak, C. M., & Ziemba, E. (2012). Critical success factors for implementing business intelligence systems in small and medium enterprises on the example of upper Silesia, Poland. *Interdisciplinary Journal of Information, Knowledge, and Management, 7*(2), 129-150.

Ong, I. L., Siew, P. H., & Wong, S. F. (2011). A five-layered business intelligence architecture. *Communications of the IBIMA*.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research, 2*(1), 1-28.

Ossai, C., & Wickramasinghe, N. (2021). *Crowdsourcing Users' Comments for Clinical and Operational Features Analysis of Diabetes Mobile Apps.* Paper presented at the Proceedings of the 54th Hawaii International Conference on System Sciences.

Palvia, P., Midha, V., & Pinjani, P. (2006). Research models in information systems. *Communications of the Association for Information Systems, 17*(1), 47.

Peral, J., Maté, A., & Marco, M. (2017). Application of data mining techniques to identify relevant key performance indicators. *Computer Standards & Interfaces, 54*, 76-85.

Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: an assessment. *Journal of management information systems, 10*(2), 75-105.

Pope, C., Ziebland, S., & Mays, N. (2000). Analysing qualitative data. *British Medical Journal, 320*, 114-116.

Premkumar, G., Ramamurthy, K., & Saunders, C. S. (2005). Information processing view of organizations: an exploratory examination of fit in the context of interorganizational relationships. *Journal of management information systems, 22*(1), 257-294.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data, 1*(1), 51-59.

Puglisi, S., Rebollo-Monedero, D., & Forné, J. (2017). On web user tracking of browsing patterns for personalised advertising. *International Journal of Parallel, Emergent and Distributed Systems, 32*(5), 502-521.

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies, 3*(9), 369-387.

Ranjan, J. (2008). Business justification with business intelligence. *The journal of information and knowledge management systems, 38*(4), 461-475.

Richards, G., Yeoh, W., Chong, A. Y. L., & Popovič, A. (2019). Business intelligence effectiveness and corporate performance management: an empirical analysis. *Journal of Computer Information Systems, 59*(2), 188-196.

Ruzgas, T., & Dabulyte-Bagdonaviviene, J. (2017). Business Intelligence for Big Data Analytics. *International Journal of Computer Applications Technology and Research, 6*(1), 01-08.

Salinas, S. O., & Lemus, A. C. (2017). Data warehouse and big data integration. *International Journal of Computer Science & Information Technology (IJCSIT), 9*(2), 1-17.

Sant, M. (2019). WASP (Write a Scientific Paper): Qualitative research and evidence based practice: implications and contributions. *Early human development, 133*, 37-42.

Santoso, L. W. (2017). Data warehouse with big data technology for higher education. *Procedia computer science, 124*, 93-99.

Shoro, A. G., & Soomro, T. R. (2015). Big data analysis: Apache spark perspective. *Global Journal of Computer Science and Technology, 15*(1).

Smuts, H., & Juleka, B. (2018). Enabling Business transformation through enterprise architecture and the knowing cycle. *International Journal of Engineering & Technology, 7*(3), 95-102.

Smuts, H., Van Der Merwe, A., Loock, M., & Kotzé, P. (2009). *A framework and methodology for knowledge management system implementation.* Paper presented at the Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists.

Sun, Z., Zou, H., & Strang, K. (2015). *Big data analytics as a service for business intelligence.* Paper presented at the Conference on e-Business, e-Services and e-Society.

Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2019). Factors Influencing Effective Use of Big Data: A Research Framework. *Information & Management*.

Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative research journal, 11*(2), 63-75.

Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons, 62*(3), 347-358.

Tambe, P., Hitt, L. M., & Brynjolfsson, E. (2012). The extroverted firm: How external information practices affect innovation and productivity. *Management Science, 58*(5), 843-859.

Taylor-Sakyi, K. (2016). Big data: Understanding big data. https://arxiv.org/ftp/arxiv/papers/1601/1601.04602.pdf

Todorova, N., & Hoeben, K. (2016). *From Data to Decisions: A Case Study of Capability Assessment in a Health Organisation.* Paper presented at the The European Conference on Information Systems Management.

Tsang, E. W. (2014). Case studies and generalization in information systems research: A critical realist perspective. *The Journal of Strategic Information Systems, 23*(2), 174-186.

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems. *Information systems journal, 20*(4), 357-381.

Van Der Aalst, W. (2016). *Process mining: Data science in action*: Springer.

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research, 83*, 75-83.

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. doi:10.25300/misq/2013/37.1.02

Wang, B., Wu, C., Huang, L., & Kang, L. (2019). Using data-driven safety decision-making to realize smart safety management in the era of big data: A theoretical perspective on basic questions and their answers. *Journal of Cleaner Production, 210*, 1595-1604.

Watson, H. J. (2016). Creating a fact-based decision-making culture. *Business Intelligence Journal, 21*(2), 5-9.

Wiesche, M., Jurisch, M. C., Yetton, P. W., & Krcmar, H. (2017). Grounded theory methodology in information systems research. *MIS quarterly, 41*(3), 685-701.

Winter, R. (2001). *The current and future role of data warehousing in corporate application architecture.* Paper presented at the Proceedings of the 34th Annual Hawaii International Conference on System Sciences.

Wixom, B., & Watson, H. (2010). The BI-based organization. *International Journal of Business Intelligence Research (IJBIR), 1*(1), 13-28.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth, 10*(1), 13-53.

Yin, R. K. (1981). The case study as a serious research strategy. *Knowledge, 3*(1), 97-114.

Appendices

8.5 Appendix A: Interview Questions

Eighteen research participants (RP) were asked to answer the interview questions shown in Figure 1. The count of RP feedback column indicates the number of participants who gave feedback for a specific interview question.

Table 30 – Interview questions

Question number	Question description	Count of RP feedback
1	In your opinion, what is a data driven organisation?	18
2	What is the role of data in your organisation’s decision making?	18
3	In your opinion, what is business intelligence?	18
4	What are popular components of a traditional business intelligence system?	18
5	How would you define big data?	18
6	What is meant by big data analytics?	18
7	What are the known challenges of traditional business intelligence systems?	18
8	What are the challenges of adopting and using big data analytic systems?	18
9	Given an opportunity to integrate traditional business intelligence with big data analytics, which components would you consider combining from both systems, and why?	17
10	Are there any general comments that you would like to add?	13

8.6 Appendix B: Coded Interview Feedback

8.6.1 Appendix B1: Question 1 Interview Feedback

Table 31 – Question 1 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	An organisation where data drives decision making.	Primary
P2	Strategic decisions are guided by data.	Primary
P3	Operational decisions are guided by data.	Primary
P4	Predictive data analytics is essential.	Primary
P5	Historical data analytics is essential.	Primary
E1	Data governance is essential.	Emerging
E2	Data assist in strategic objectives tracking.	Emerging

Table 32 – Question 1 coded interview feedback

RP ID	Feedback	Themes						
		P1	P2	P3	P4	P5	E1	E2
RP1	A data driven organization is one that uses data by deriving useful insights to aid strategic and operational decision making	X	X	X				
RP2	An organization which makes decisions based on data they have. An organization that uses data when making organizational decisions.	X						
RP3	An organization that leverages the knowledge gained by analysing its available data store to correctly determine past behaviour and predict future behaviour.				X	X		
RP4	A data driven organisation is the one that has embedded data collection, wrangling, analysis and decision making into every aspect of its business, and uses the insights provided by the analysis of the data to perform evidence based decision making at all stages.	X	X	X				
RP5	It is an organization that understands and practices Information Lifecycle Management (ILM), uses data to make informed decisions, promotes or encourages continuous learning to its employees and customers, encourages a culture of collaboration, and has a strong data governance.	X					X	
RP6	An organisation which has been able to access its vast amount of data and use it to create reports/ MIS/ analytics which is accessible to all decision makers who can then use that data & insights to make their business decisions instead of relying on their gut.	X	X	X				
RP7	An organisation that is able to tracking their business performance using data, make rational decisions based on the information provided, and action those decisions. "Important to note that this is a cyclical process."	X						
RP8	Organization make decisions and forecasts with data.	X			X			
RP9	A data driven organisation would be an organisation that gives priority to having the correct data available when required to assist in making decisions whenever possible. This would not only be at a strategic level but also for the staff on the ground. Everyone needs the correct data at the right time to assist with decisions.	X	X	X			X	
RP10	A data-driven organisation relies heavily on the insights and analysis derived from their data in order to assist with their decision making process. Therefore, enabling these organisations to make informed decisions, and validating a course of action before committing to it. Which will lead to commercial growth, stability and obtaining their objectives and goals.	X						X
RP11	A data driven organisation makes decisions based on proof gained from data. Trends and deep dives inform how decisions are taken.	X						
RP12	Any organization that uses historical or predictive data to make decisions on moving the organization forward.	X			X	X		X
RP13	Data-driven organisation is an organisation that makes informed decision based on facts and data provided.							
RP14	A data driven organisation is an organisation that ensures processes are in place for accurate data capturing and storing in order for decisions to be made on actionable insights gathered from that data. It removes biases by having such insights that allows for unbiased decisions to be made.	X					X	
RP15	An organization which collects, stores data and structure it for usage by the end user, which can be used for various objectives.							
RP16	A data-driven organisation is the type of organisation that makes informed decisions based on analysed and measurable data. With the help of business intelligence tools, enriched data can help the organisation learn current state of the business and predict future trends.	X			X			
RP17	An organisation that uses insights gained from their data to determine what their next steps and future steps will be in order to achieve their objectives and goals.				X			X
RP18	An organization that gathers data in all aspects of decision making and enabling employees to use the data.	X	X	X				
Count	**18**	**14**	**5**	**5**	**5**	**2**	**3**	**3**

8.6.2 Appendix B 2: Question 2 Interview Feedback

Table 33 – Question 2 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	Decision making.	Primary
P2	Product offering and customer service improvement.	Primary
P3	Predictive modelling.	Primary
P4	Business process and business operations optimisation.	Primary
E1	Business opportunity identification.	Emerging

Table 34 – Question 2 coded interview feedback

RP ID	Feedback	Themes				
		P1	P2	P3	P4	E1
RP1	The role of data in the organization is at the forefront in making strategic and operational decisions, whether it is to improve customer service or provide better products offerings based on data and analytics	X	X			
RP2	To gauge customer satisfaction with the products and services which are offered to them as well as using data to form and make solid competitive strategies and the overall business culture and satisfactory of the stakeholders. As well as to create a solid competitive analysis strategy.	X	X		X	X
RP3	Used to manage operational business processes and perform basic predictive modelling			X	X	
RP4	A large part of data collection is mainly to do reporting on monthly metrics for this different business units which (I hope) they use to inform business strategy. There are also one of pieces of analysis to provide pockets of understanding of certain issues or segments to make tactical decisions. We also use data to perform predictive modelling and A/B testing to decide which models to scale to a wider base.			X	X	
RP5	Data allows the company to create new opportunities, have actionable insights, predict future trends, optimise current processes			X	X	X
RP6	It is very important, and is definitely being used to aid decision making, however not enough of our data is being mined and not enough of our data is available in near real time to allow for quick decision making.	X				
RP7	A lot of data in BI reporting is used to measure success of channels to KPI's and is used motivate and prioritize business actions.		X		X	
RP8	Database administrator.					
RP9	Data is used at all levels to report and track sales performance and operational information. It is used to assist the staff in making good sales and keeping customers on books. The sales data is used to improve and optimize the sales process and provide insight into areas that need attention.		X		X	
RP10	The role of data in OM's decision- making should be the following: to enable OM to make more confident decisions, to enable OM to become more proactive, and to enable OM to save costs.	X		X	X	
RP11	Data is used in a number of strategic decisions but "it is not fully mature."	X				
RP12	Data plays a substantial role in my organizations decision making but "we could do more in the predictive analysis space."	X		X		
RP13	The role of data in our organisation is to use that data to drive business decisions.	X				
RP14	Data in Old Mutual is there to provide information on what has happened. Decisions are then made based on that information. "Old Mutual is not really data-driven organisation since actionable insights aren't really gathered from the data, instead data is just used to provide information."	X				
RP15	This relates where insights are obtained from data and decisions are then based on these insights. it plays a key role in decision making, hence the data element becomes critical.	X				
RP16	At my organisation data is a big part of decision-making, it helps the organisation understand their customers better, it helps with the understanding of the latest trends and with making strategic business decisions that will grow the business and give it a competitive edge. For an example at my organisation, before they launch a new product they will do campaigns for a group of selected customers and depending on how the campaign performs, they will decide on what products to launch.	X	X			X
RP17	It is playing a growing role. My organisation wants to improve its ability to use data to drive the decisions that are made. "Progress has been made but there is still a lot of work to do."	X				
RP18	Analysing financial reports, KPI, and performance management for our employees, customer feedback on surveys. Keeping up with our competitors.		X		X	X
Count	**18**	**11**	**6**	**5**	**8**	**4**

8.6.3 Appendix B3: Question 3 Interview Feedback

Table 35 – Question 3 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	Means of transforming data into information.	Primary
P2	Software tools and processes used to transform data to information.	Primary
P3	Entails data analytics and reporting.	Primary
P4	Use of data to support decision making.	Primary
E1	Involves data and information quality.	Emerging

Table 36 – Question 3 coded interview feedback

RP ID	Feedback	Themes				
		P1	P2	P3	P4	E1
RP1	Business Intelligence is the use of analytical software to change data into useful insights which can assist an organization with business decisions.	X	X		X	
RP2	The use of software and processes to turn raw data into meaningful information which can be used to make informed decisions.	X	X		X	
RP3	Having the correct data available at the correct time at the correct grain for interpretation to guide business decisions.				X	X
RP4	BI what most companies in SA have set up. SQL servers that manage tabular customer data at volumes that are not too great or frequent. Reporting is often done, nothing particularly proactive.		X	X		
RP5	It is the strategies or processes and technologies or tools that used to collect and analyse data that help the organization to make informed and insightful decisions.	X	X	X	X	
RP6	Traditionally BI has been used to refer to more MIS or operational reporting so for example daily or weekly reporting which assists with the business operations so a good example of this is weekly sales reporting, daily call centre reporting etc.			X		
RP7	Retrospective measuring the success of the business against KPI's.					
RP8	It's the process to build, transform, and present data for business consumption.	X	X	X		
RP9	BI is about turning data into information. This now information is used to assist all levels of the business in improving performance. The information needs to provide insights in an easy to understand and simple manner. This should also include both external and historic data giving the ability to expand beyond operational reporting and into predictive tools/analysis. This assists to create a business that is proactive rather than re active.	X	X	X		X
RP10	Business Intelligence (BI) refers to the transforming of data into business information and insights about an organisations current state. Although BI does not tell the business users, what to do, or what will happen if they take a certain course. "Neither is BI solely about generating reports. BI rather offers a way for people to examine data to understand trends and derive insights necessary to make sound decisions."	X		X	X	
RP11	Business intelligence is the 'art' of using software to create information out of data.	X	X			
RP12	Providing decision makers with the correct data to derive the right information at just the perfect time in the right format.	X			X	X
RP13	Business intelligence is a set of processes that converts raw data into meaningful information so that the business can make informed decisions based on facts.	X			X	
RP14	Business Intelligence is the process of collecting data and storing it in a uniform manner in order to gain insights from the data by analysing and reporting on it. This enables the business to make better decisions.	X		X	X	
RP15	This will relate to usage of the data in its various forms or structures and then to provide valuable views / insights of the data to the business for reporting/decision making.	X		X	X	
RP16	Business Intelligence is the use of tools that organise, analyse, enrich data, present it in a way that makes sense and can be used to make business decisions. This can be in a form of reports, dashboards, visualisation etc.	X		X	X	
RP17	The collection of relevant data that is required to be viewed by various stakeholders in various formats and via various mediums.			X		X
RP18	The software and services that transform data into actionable insights for business decisions.	X			X	
Count	**18**	**13**	**7**	**10**	**11**	**4**

8.6.4 Appendix B4: Question 4 Interview Feedback

Table 37 – Question 4 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	Extract Transform and Load (ETL).	Primary
P2	Data Warehouse.	Primary
P3	Analytical tools.	Primary
P4	Reporting.	Primary
E1	Data sources	Emerging

Table 38 – Question 4 coded interview feedback

RP ID	Feedback	Themes				
		P1	P2	P3	P4	E1
RP1	Databases or Files; Extract, Transform, Load (ETL); Data Warehouse; Application or analytical tools; Reporting.	X	X	X	X	X
RP2	Data warehouse; Data source systems; Power BI; Real time and online analytical; processing of data.		X	X		X
RP3	Data warehouse with data pipelines leading to and from for data collection, aggregation and export in varying forms, data marts or canned reporting or pivot table data structures			X	X	X
RP4	SQL servers, SQL server management studio, SAS, tableau/Power BI, Alteryx, excel. Mainly to access, wrangle and report on data.			X	X	
RP5	Data warehouse; OLAP; ETL; Dashboards; Reporting.	X	X		X	
RP6	An underlying data mart or data warehouse and then a transformation or business rules layer and then the visualization or front end layer which the business users interface with like your Power BI or Tableau as examples.	X	X	X	X	
RP7	An integrated mix of admin systems, warehouses and data cubes used to pull data into manual and automated reports.		X	X	X	
RP8	data warehousing, analytics, performance, and user interface.		X	X	X	
RP9	Data warehousing, OLAP (Aggregated reporting), Multiple sources of data, ETL and static reports/metrics (dashboards/extracts).	X	X	X	X	X
RP10	The following components are what a good BI Architecture consist of: Collection of Data: Gathering data from various sources, in order to be able to manipulate with it. Data Integration: Extracting data and loading it into the data warehouse. This is the ETL process. (Extract Transform Load). Data Storage: Data warehousing and Business Intelligent concepts have certain differences; however, they cannot function without each other. Below are the main differences between the 2: BI goals are more focussed on business insights, whereas the DWH stores the company's data. BI outputs information to provide insights, the DWH outputs data in dimension and fact tables and cubes. BI tools translate the heavy IT data into insights that an average business user can understand, while the DWH is used by, data engineers and backend developers. Analysis of Data: Modern BI tools empower business users to create queries with a few clicks and without profound technological knowledge. The DWH work behind this process and makes the overall architecture possible. • Data Distribution: This can be performed in 3 ways: (1) Reporting via automated emails. (2) Dashboarding. (3) Embedded BI. Reactions based on generated insights: Without the structure of DWH and business intelligence, business won't be able to progress.	X	X	X	X	X
RP11	Data sources; OLAP; Data warehousing; Advanced Analytics		X	X		X
RP12	a standard loading process, plus business and data rules, plus a data warehouse , plus a reporting layer.	X	X		X	
RP13	OLAP; Data warehousing; data sources.		X	X		X
RP14	Popular components of a BI system are: ETL tools, A data warehouse, Data marts, cubes, and Reports.	X	X	X	X	
RP15	The structure of the data - easily to understand and interpreted. The Accuracy of the data Real-time updates Data warehouses.		X		X	
RP16	The major component of Business Intelligence system is a data warehouse, this is critical because that is where data is stored and organised before it is presented to the end user. Data marts are part of the data warehouse and they are useful for partitioning data into different segments that will be easily accessible for reporting. ETL is also very important because this is the layer of the system, which extracts data from source systems, transforms and conforms data before it is stored. Reporting tool is another important component because this is the presentation layer of BI. Cubes are great for analysis but BI does not need them to be effective.	X	X	X	X	
RP17	A staging area, an ETL process, storing of data according to a particular methodology, example Kimball, a semantic layer, metadata, cubes and reports.	X	X	X	X	
RP18	Data warehouse, data sources, ETL and excel files, predefine data view, and hardware limitations.	X	X		X	X
Count	**18**	**10**	**16**	**14**	**15**	**8**

8.6.5 Appendix B5: Question 5 Interview Feedback

Table 39 – Question 5 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	Volume	Primary
P2	Velocity	Primary
P3	Variety	Primary
P4	Veracity	Primary
P5	Value	Primary
P6	Beyond traditional BI tools processing capacity.	Primary

Table 40 – Question 5 coded interview feedback

RP ID	Feedback	Themes					
		P1	P2	P3	P4	P5	P6
RP1	Big data are large volumes of information which are continuously increasing.	X	X				
RP2	Data that is too big and seem to be impossible to store. it is can be unstructured, structure or semi structured data and can be obtained from different source systems including from websites.	X		X			
RP3	Very large data stores that grow exponentially in size rapidly.	X	X				
RP4	Huge datasets(potentially unstructured data) generated at a high velocity that is difficult to access and wrangle with existing BI tools such as SQL server.	X	X	X			X
RP5	This is the large amount of data both structured and unstructured that is normally extracted from social media, pictures, websites, GPS, etc. This data is difficult to understand the relationship between the database and sometimes difficult to understand.	X		X			
RP6	Data which is too big to open or look at using traditional means like Excel, I would think anything where the data is to wide or too long to actually be looked at that I refer to as big data. However in research you will often see it being referred to as data sets which are millions of records and so you need special software or systems to be able to handle it.	X					
RP7	Using large volumes and several forms of structured and unstructured data to any purpose i.e. reporting, analysis or modelling.	X		X			
RP8	It's the increase of data in large diverse sets.	X		X			
RP9	Large amounts of data from any/multiple sources most likely in an unstructured format that is constantly available/giving new data and increasing in volume.	X	X	X			
RP10	The definition of big data refers to the large sets of information that grow at an increasing rate. It includes the volume of the information, and the velocity/speed at which it is created and collected. The variety of the data points being covered and it arrives in multiple formats.	X	X	X			
RP11	Big data consists of data sets that are generally too large to be analysed through traditional means.	X					X
RP12	Can simply be defined as all data. This is not just all the data that can be generated by an organization in an organizational setting. This is all the data available on all electronic platforms available to you. This data is also not structured in any conventional way but can be structured or grouped by text, graphic (diagrams/ pictures), sound etc. An Unlimited source of unorganized, unstructured and uncategorized data.			X			
RP13	Big data refers to a large, complex combination of structured and unstructured collection of data from various sources that is difficult to use traditional BI tools to convert data into meaningful information	X		X			X
RP14	Big Data are very large volumes of structured, semi-structured and unstructured data. It is produced at a high speeds and can be difficult to control its quality, but that is very valuable due to insights that can be gathered from it.	X	X	X		X	
RP15	In an organization where there are different areas using data in different ways or receive data from different sources, is to then consolidate this data and use it to obtain insights across the value chain incorporating all areas of the organization.					X	
RP16	Big data is large volumes of data that continue to grow exponentially with time and traditional data management tools cannot handle its complexity. The data can be structured, unstructured or semi-structured. Big data also refers to how fast large volumes of data is generated and processed. In addition, being its ability to handle the inconsistency of the data.	X	X	X	X		X
RP17	A large volume of data that can be used to gain insights about various aspects of the processes that the data relates too.	X					
RP18	The concept of 5V's Volume huge amount of data Velocity high speed of accumulation Variety different formats of data from different sources inconsistencies uncertainty in data Value Extracts useful data.		X	X	X		
Count	**18**	**15**	**8**	**12**	**2**	**2**	**4**

8.6.6 Appendix B6: Question 6 Interview Feedback

Table 41 – Question 6 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	Modern big data analytics tools.	Primary
P2	Same as traditional data analytics.	Primary
P3	Predictive modelling and analytics.	Primary
P4	Data mining.	Primary
E1	Machine learning.	Emerging
E2	Hadoop.	Emerging

Table 42 – Question 6 coded interview feedback

RP ID	Feedback	Themes					
		P1	P2	P3	P4	E1	E2
RP1	Big data analytics is the use of advanced analytical techniques to analyse large volumes of data from a range of sources which could be structured, semi structured and unstructured	X					
RP2	The process of analysing data using some sort of analytical methods, such as predictive analysis to find patterns and form of how the data can be stored or used.		X				
RP3	manipulating said data to gain buss related insights		X				
RP4	Using the relevant tools to perform data wrangling or analysis or modelling on these large data sets that you would not be able to with current technology. E.g. Apache Spark and Hadoop	X					X
RP5	It is the strategies or processes and technologies or tools that are used to collect and analyse huge and complex data from pictures, social media, videos, etc. that help the organization to make informed and insightful decisions.	X					
RP6	Analytics or insights generation using all this large data to look for underlying insights or trends or patterns which are useful for business decision making. This might be exploratory data analytics or could go all the way to advanced analytics such predictive modelling and machine learning.	X	X	X	X		
RP7	Integrating these several sources of data together to either to understand and develop insights of processes and business actions or developing predictive models for decision making.		X	X			
RP8	To build enterprise trends, customer behaviour and correlation.		X				
RP9	The use of Machine learning, Data mining predictive analysis and other advanced tools/methods to gain insight and make decisions. The data used by these tools and methods will come from multiple sources in both structured and unstructured format including traditional data and big data. (see definition of big data).	X		X	X	X	
RP10	Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes. Analysis of big data allows analysts, researchers and business users to make better and faster decisions using data that was previously inaccessible or unusable. Businesses can use advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics and natural language processing to gain new insights from previously untapped data sources independently or together with existing enterprise data.	X		X	X	X	
RP11	Big data analytics is the process, sometimes complex, of finding trends etc. in large datasets.		X				
RP12	The process by which information is obtained through extrapolating from both a Traditional Business Intelligence system as well as an Unlimited source of unorganized, unstructured and uncategorized data. You Traditional Business Intelligence system will generally initiate your questions to the unstructured sources.	X	X				
RP13	Process used to breakdown complex big data so that it can be easy to analyse the data and other insights.	X					
RP14	Big Data analytics refers to the process of collecting, organizing and analysing Big Data using technologies like NoSQL databases, Hadoop, Data Mining and Predictive analytics.	X		X		X	X
RP15	Data analytics is the analysis of raw data in order to make conclusions about that information. Many of the data processes have been automated be it into programs or algorithms which will then provide end results in the form that the end user can understand and use.		X				
RP16	Big data analytics is the process of collecting, organising and analysing large volumes of data from different sources with the intention of discovering patterns or trends in the data. The information generated from big data is useful in making sound business decisions that will potentially grow the organisation.	X					
RP17	Doing analysis on large volumes of data to identify trends and to convert data to information.	X					
RP18	Collecting, organizing, and analysing big data. This includes both data from internal and external	X					
Count	**18**	**12**	**8**	**5**	**3**	**3**	**2**

8.6.7 Appendix B7: Question 7 Interview Feedback

Table 43 – Question 7 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	Expensive.	Primary
P2	Lack of self-service functionality.	Primary
P3	Poor performance due to processing of large data volumes.	Primary
P4	Inability to process unstructured data.	Primary
P5	Lack of flexibility.	Primary
E1	Hinder business innovation.	Emerging
E2	Inability to analyse desperate data sources.	Emerging
E3	Lack of mobile devices support capability.	Emerging

Table 44 – Question 7 coded interview feedback

RP ID	Feedback	Themes							
		P1	P2	P3	P4	P5	E1	E2	E3
RP1	Legacy issues - unable to handle large amounts of data so performance is affected, unable to process unstructured data.			X	X				
RP2	The use of software's and process of transforming raw data into useful information for reporting purposes. Source system files structures, Technological infrastructure, Storage space.								
RP3	complexity and the fact that not one size fits all. Business tends to run 5 years the technology that is current.						X		
RP4	Proprietary(and expensive) "ubiquitous(everyone uses them so you have to as well)" Sticky(companies do not like to change from what they have). Inflexible/potentially outdated. Unable to handle larger datasets effectively/easily.	X		X		X	X		
RP5	Disk Space because data continues to grow. Speed to do ETL, the more data is available the more time it takes to ETL.			X					
RP6	The user usually has to work quite hard to generate their own insights, most reports are pre built for a specific purpose and so if you need to answer a different question, then it's not easy to do this.		X						
RP7	Too make different legacy data systems that do not recon with each other.								
RP8	Cost, Business adoption, Analysing, mobile capability, and self-service.	X	X						X
RP9	Traditional BI is expensive and requires a large infrastructure and process to support its operation. Businesses also often struggle to adapt to a data driven decision making process and can't always see the ROI. Pulling data from multiple sources structured and some rare or unstructured sources is often difficult. Combining these sources into one warehouse poses a challenge. Traditional BI does has not yet adapted to the mobile world and is not readily available on mobile devices. Self-service also often poses a challenge with users either not using or understanding what is available to them. Poor source data is often difficult to process and include into the warehouse without issues and re occurring data concerns. Traditional BI often lacks a strategy that provides data in a simple and easy to understand manner that allows insights to assist with decision making. The gathering of data and providing information is often slow with poor database performance.	X	X	X	X		X		X
RP10	• Too expensive and hard to justify ROI of BI. • Lack of company- wide adoption. • Analysing data from different data sources. • Dealing with the impact of poor quality data.	X						X	
RP11	Analysing data from different data sources, expensive.	X						X	
RP12	1. Traditional Business Intelligence system tend to focus on historical data generated inside an organization. This makes it almost impossible to shift strategies at the same pace the world is changing. Decisions and Strategies is created based on historical trends which does not predict how the world has changed. 2. Convincing Business stakeholders the value BI brings.	X					X		
RP13	I think it is not flexible, not easy to make changes. Also difficult for traditional BI to create self-service analysis.		X			X			
RP14	Some of traditional BI challenges are: Processing doesn't happen in real time. It cannot process semi-structured and unstructured data.				X				
RP15	1. Too expensive and hard to justify the ROI of BI 2. Not all areas in a company adopts BI. 3. Analysing data from different data sources. 4. Providing true self-service analytics.	X	X					X	
RP16	Traditional business analysis systems struggle to process very large volumes of data; they often cause performance issues. They are also not dynamic enough to source data from different data platforms; this means unstructured or semi-structured data is a challenge for traditional BI systems. Another challenge is not having access to data in real time and that means the business will not have access to the latest data so they can be able to make informed decisions quicker.			X	X	X			
RP17	Up to date metadata, self-service analytics, expensive exploration tools	X	X						
RP18	Speed, static, few people have access/knowledge to it and separates systems.		X	X					
Count	**18**	**8**	**7**	**6**	**4**	**3**	**4**	**3**	**2**

8.6.8 Appendix B8: Question 8 Interview Feedback

Table 45 – Question 8 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	Lack of skillset.	Primary
P2	Computer power requirements.	Primary
P3	Expensive.	Primary
P4	Legacy system integration.	Primary
P5	Undefined processes.	Primary
E1	Validation of fast growing data.	Emerging
E2	Data security risk.	Emerging
E3	Difficult to prove its ROI.	Emerging

Table 46 – Question 8 coded interview feedback

RP ID	Feedback	Themes							
		P1	P2	P3	P4	P5	E1	E2	E3
RP1	Not many people are adequately skilled in big data analytics since it is a new concept. Since big data deals with large volumes of data which can be unstructured it will require more computing power as well as sufficient memory/storage to process the data	X	X						
RP2	Infrastructure compatibility. Not having skilled professionals i.e., it requires a lot of specific skillsets, Architectures, data engineers to name a few. Tools and process are not clearly identified or perhaps lacking. Data storage, Undefined data structures.	X	X			X			
RP3	"If the goals and expectations are managed correctly no diff to instituting a normal BI solution".								
RP4	It's a new skill set to learn and upskill your data and analytics department. A different way of warehousing data, a different way of accessing and wrangling it. It is hard to show the value in that immediately, even if you are performing things like A/B testing. You know it is keeping you relevant and up to date, but it's a big cost to just stay in touch instead of expecting it to pull you ahead.	X	X	X					X
RP5	Lack of required skill set. In terms of price, it is more expensive. Because the data is huge, reporting from the big data is not easy.	X		X					
RP6	I think they can be quite intimidating to some business people who do not understand big data or advanced analytics, because machine learning and predictive analytics can be quite difficult to understand. So getting business to buy in and really see and understand the value of using big data analytics can sometimes take a long time and there might be resistance at first.	X							X
RP7	Integrating the new models to old systems. Misunderstanding of how data and predictive models work to generate output lack of business ability to interpret and visualize data in order to make rational decisions.	X			X				
RP8	Data growth, fast insights, and data validation.						X		
RP9	"Big data is a new concept or tool for many organisations. It is not understood or is interpreted to imply only large amounts of data." The skill set in the existing organisation workforce does not always cater for all required skills. Organisations do not plan for the growth in volume and number of sources and increased availability of new data. Organisations are not willing to adapt existing processes to a more data driven decision making model. Organisations will resist and prefer to stick to what they know making a ROI investment slow. Many if not all big data solutions will centralize an organisations data making a large security risk that needs to be properly managed and secured.	X						X	X
RP10	The following are challenges of using big data analytics systems: Dealing with Data growth. Generating insights in a timely manner. Recruiting and retaining big data talent. Integrating disparate data sources. Validating data. Securing big data. Organizational resistance.	X					X	X	
RP11	No use cases defined, data not properly structured in big data landing zones (data lakes)					X			
RP12	knowledge, infrastructure and ethics. Knowledge - Education is not as readily available as data warehousing and other BI Tools, and specialist IT courses are costly. Infrastructure - In our country technology is a lot more expensive than other countries. Ethics - Protection of Personal Information (RSA POPI Act) Any and all organizations operating in South Africa is required to protect personal information of their customers as well as their employees, yet big data technology gives us the ability to gather the same information from Public sources (such as Facebook, Instagram, YouTube, etc.	X		X				X	
RP13	"With so much data available, it's difficult to dig down and access the insights that are needed most, this may cause not too fully analyse the data or focus on measures that matters".								
RP14	People with the big data analytics skills are still high in demand so organisations lack some of those skills and it takes longer for the concept to become more of a norm within the organisations and thus to adopt too. It is also difficult to control the quality of data since it comes from a variety of sources and in different formats.	X					X		
RP15	Data privacy, the points to be considered are the source of data and the way it is used i.e. can it be trusted, is it data that can be shared etc. Integrating legacy systems with big data technology. Lack of resource skills – analytics. The cost of big data tools.	X		X	X			X	
RP16	There are some challenges that come with big data analytics systems such as the integration of legacy systems with big data systems, it is not an easy transition and it can be costly. With the large amounts of data that will continue to grow, storage might also be a challenge so that will require the amount of storage that will cater for the rapid growth of data. The other challenge of big data systems is the high cost of setting up the infrastructure for big data. The lack of skill sets might also be a big challenge; there is not a lot of skilled workers for big data analytics this means training will be required.	X	X	X	X				
RP17	Cost, it takes time to get all the required sources.			X					
RP18	Skills and resources, Inaccessible data, poor quality, multiple data sources and visual representation of data.	X					X		
Count	**18**	**13**	**4**	**6**	**3**	**2**	**4**	**4**	**3**

8.6.9 Appendix B9: Question 9 Interview Feedback

Table 47 – Question 9 interview feedback theme legend

Theme ID	Theme description	Primary/Emerging
P1	The two system architectural components can be combined.	Primary
P2	Use ETL and data warehouse for structured data.	Primary
P3	Use big data tools for unstructured data.	Primary
P4	Use BI for operational reporting.	Primary
P5	Use big data analytics for advanced analytics.	Primary
E1	The two system architectural components should not be combined.	Emerging
E2	Explore cloud computing.	Emerging

Table 48 – Question 9 coded interview feedback

RP ID	Feedback	Themes						
		P1	P2	P3	P4	P5	E1	E2
RP1	Databases / Files/ Unstructured data. ETL Data warehouse for structured data, big data capabilities for unstructured data. Advanced Analytical Tools Reporting. The above components as they will complement each other where both unstructured and structured data can be analysed efficiently.	X	X	X				
RP2	Both frameworks can be used together or even interchangeably at any point depending on the specific use case. Both BI and Big data have some components borrowed from each other which then makes it easy to integrate both frameworks. i.e. BI can make use of big data analytics as a tool during the BI processes and big data framework can also rely on BI tools during its processes depending on the objectives.	X						
RP3	"I can't really see the diff to be honest - think the high level data flow remains the same - just a change in toolset"							
RP4	Knowing how to work in the traditional BI world will never go away. There will always be a need to operate there, or where a big data implementation would be overkill; so having this base would be great, especially the thinking you gain on how to wrangle and use data. I think the more coding heavy big data tools like Spark provide greater flexibility to your analytics teams and should definitely be included (as opposed to the GUI tools). This can filter down to usage in traditional BI as well. High volume/unstructured data should be moved on to big data platforms where appropriate. "In short, I guess traditional BI is the base, and appropriate use of big data and the related tools sit on top of that(and does not work as a replacement)."	X	X	X				
RP5	The ETL and Reporting components - Data can be taken used together by taking structured data that is used for day to day reporting and putting in the warehouse, then run the daily or recurring reports.		X		X			
RP6	Maybe combining the operational reports and then moving into advanced analytics so for example if you see something interesting when you as a sales manager are looking at your weekly sales reports, you could dig a little deeper into that data looking for insights and then if needed even result in some advanced analytics being undertaken like a forecasting model, or a propensity model or even some machine learning being kicked off. Also being able to use dashboards and MIS type reports to visualize the results of advanced analytics like a propensity model to allow business users to understand the finer details of the model and how it can be used to generate value for the business.	X			X	X		
RP7	Traditional BI is well understood and can be presented to users in a simple way. Big data should then be used to supplement the understanding from traditional BI.	X			X	X		
RP8	Structured data for high transact queries. Data visualization for reporting and Analysis Data flow post big data implementation.							
RP9	I would keep the multiple sources of data from big data. This would include both structured and unstructured data. I would create a "ETL" process that controls the extract and loading of data while reducing the importance of data transformation and placing more importance on ensuring data quality. A structured layer would sit on top of the data providing a more traditional BI star schema view and transformation of the data. This layer would need to be flexible to adapt quickly to change and add new data as needed. This data can be used to create OLAP models or dashboards as required. Slowly changing dimensions would be the exception and not the rule. This solution would work best on cloud which allows for scalability and has tools to assist with data management and predictive analysis.	X	X					X
RP10	Yes, this could definitely work, as unstructured data cannot be stored in the warehouse. "However, not sure if this will this allow for interaction of" the structured data stored in the warehouse and the unstructured data stored on image processing.	X	X	X				
RP11	Data warehouse and data lake. A data lake is a perfect environment to store data from various sources e.g. structured/unstructured and a data warehouse conforms data in a structured way for particular use cases. This is called a Lake House.		X	X				
RP12	Data warehousing and all its underlying infrastructure as well as Standards and Best practices. Cloud computing allows for less expensive and more flexible infrastructure but always be cognizant that this is only true if the environment is managed correctly. It makes processing both traditional data warehouse and big data possible because of its flexible scaling. Just constantly be aware of the Information overload concept. "Integrating Big data and its concepts in any organization can be of great use for predictive modelling, but care should be given that not unnecessary data is extrapolated just because it is available." If big data components can provide valuable insights, it should be considered. Big data reporting tools has become very popular but can be expensive if the tools aren't utilized to its fullest and adds the value that drives the organization's strategy.	X						X
RP13	More information allows you to make informed decisions so collecting data from different sources will give you more insights on the decision. Secondly having structured and unstructured (to cater for a number of different structures) will help the business to move forward. Lastly integrating BI with big data will help with creating flexible self-services and meaningful patterns in data.	X						
RP15	1. Tools use to extract data - why should have the same type of tools? 2. Data store warehouses - both should have platforms where data are stored						X	
RP16	Traditional Business Intelligence systems and big data can be integrated to save costs. For an example, Data Warehouse can be used in conjunction with a big data platform. The data warehouse can leverage of the big data platform, which will be collecting and organising data from different sources and in different forms. Big data would replace the ETL part of BI and having access to information quicker will mean the organisation will have access to real time data for analysis and reporting.	X						

RP ID	Feedback	Themes						
		P1	P2	P3	P4	P5	E1	E2
RP17	To be honest, I do not see why traditional BI cannot be used for big data. With the right infrastructure and the right toolset, I believe traditional BI can accommodate big data.						X	
RP18	Data Warehouse: storage and processing of data, Cloud: pay as you use or software as a service. Reporting services: updating reports and dashboards Power BI.		X					X
Count	**17**	**10**	**7**	**4**	**3**	**2**	**2**	**2**

8.6.10 Appendix B10: Question 10 Interview Feedback

Table 49 – Question 10 coded interview feedback

RP ID	Feedback
RP2	Yes. BI uses set of processes and tools in order to refine data which can be useful for organizations. BI data is clean and transformed and can be used meaningfully, while Big data on the other hand is data which can be as big as imagined and can be structured or unstructured or semi structured thus if not BI tools and processes are not adopted in this kind of framework, the data might be too big to be understood, therefore a bit of refinement is required, hence the BI tools and process can be useful in big data.
RP3	Think Big data means different things to different people, depending on the business requirement; a one size fits all approach is not always helpful.
RP4	Big Data is a big buzzword at the moment, but it needs to be carefully evaluated if it's really what you need to get the job done as it's not always an appropriate tool.
RP5	This was an eye opener. Thank you for the invitation to participate.
RP7	Decision making processes are themselves not well understood and cultural of the busy is to rely on past experience. Hence without shifting towards using data (as opposed to gut feel) to make decisions, the value from shifting from traditional BI to big data will not be achieved.
RP9	Big data is a dream, buzz word or goal used by organisations. I am not sure it is always understood or implemented effectively. A design or framework is needed to that would guide organisations to a solution not only to "implement big data" but one that quickly gives a ROI that draws organisations in leaving them wanting more.
RP10	In my years of experience working in an organisation who moved from making decisions on very limited information or gut feel, to a more data driven decision making organisation, Executives are more confident on their decisions and results in growth and stability for the organisation.
RP12	My personal opinion/hope is that big data will not truly replace what is now seen as traditional BI but will in future become traditional BI. We have the power to create a world where the two environments can be completely symbiotic and provide our organizations with insights a lot more complementary to the world we live or will live in.
RP15	I think both the BI and big data plays a role in an organization. both can be used to derive great business benefit. what is key is the structure of the data which needs to be easily interpretable for the end user. BI Areas should be the enablers for usage of data by the end users in a simple non complicated structure, real-time and extraction of data in the shortest possible time and real-time.
RP16	I am glad I took part in this survey, It was very informative.
RP17	There is currently a lot of hype around data and the growth of data that companies are collecting. Data can definitely be used to aid decision making. Large volumes of data though are not necessarily better than smaller volumes of data. Data quality is very important and should not be overlooked. Too much time is still being spent to clean up data, something which I feel should not be the case. At the same time, all systems that will result in data been captured, should be created with input from data analysts and data scientists etc. to ensure that the captured data will aid decision making.
Count	**11**