Risk assessment processes for big data based on cloud computing technologies: A comparative study

FMM Fataftah

orcid.org/0000-0002-3299-0972

Thesis accepted in fulfilment of the requirements for the degree *Doctor of Philosophy in Information Technology* at the North-West University

Promoter: Prof B Isong

Graduation: November 2022

Student number: 28175107

Declaration

I, Fadi Manna Fataftah, declare that this thesis titled **"Risk assessment processes for big data based on cloud computing technologies: a comparative study"** is my work carried out at the North-West University, Vanderbijlpark Campus, and has not been submitted in any form for the award of a degree to any other university or institution of tertiary education or published earlier. All the materials used as the sources of information have been dully acknowledged in the text and the references.

Signature:

Date: February 22, 2022

Supervisor:

Date: March 17, 2022

Acknowledgements

First, I would like to praise Allah the Almighty, the Most Gracious, and the Most Merciful for His blessing given to me during my study. May Allah's blessing go to His Prophet Muhammad (peace be up on him), his family, and his companions.

I am extremely grateful to my supervisor, Prof. Bassey Isong, for giving me this chance and for his continuous support, guidance, feedback, and patience during my study.

To all NWU staff members at the Vanderbijlpark Campus who supported and helped me during my residency at the campus, thank you.

I am grateful to my beloved mother, Etaf, for her continued support, phone calls and prayers during this long journey. For whom was always pushing me and had my back, my late father, whom I lost in November 2020.

To my wife, Eman and the kids' Manna, Lydia and Taym, thank you for being patient and supportive. In addition, for my brothers and sisters, especially Anas.

Language Declaration

DECLARATION OF LANGUAGE EDITING

13.03.2022

The Writer's Fix
Independently contracted editing

I, Natasha Ravyse, hereby declare that I have completed a language, technical and reference edit (NWU Harvard) for the thesis with the title:

"Risk assessment processes for big data based on cloud computing technologies: A comparative study"

for **F.M.M. Fataftah (28175107)** in order to meet the requirements for the degree *Doctor of Philosophy in Information Technology* at the North-West University.

Changes were suggested in the form of electronic track changes and comments. The implementation thereof was left to the discretion of the author.

N.E. Ravyse

Dr Natasha Ravyse
Independently contracted editor (Private and C-Trans SATI Accredited)
Full member of Professional Editors' Guild

Abstract

In recent years, the use of big data has dramatically increased in a variety of fields, for example, healthcare and management, among others. Because of the growing number of digital and linked devices and platforms, traditional data forms have shifted from structured to unstructured data, including various image and video formats, text, and other forms of data. As a result of the increasing volumes and frequency of data collection, improved data analysis, including risk assessment, is required to keep up with the times. Specifically, this study aims to give a comparative analysis of big data risk assessment approaches based on cloud computing technologies in four countries: Canada, Jordan, South Africa, and the UK. Additionally, cloud computing is examined in this study, which involves transporting information and computing from home and business computers to external data hubs.

This study followed an embedded mixed research methodology to collect quantitative and qualitative data, where the quantitative data play a supportive role in the qualitative data. A structured questionnaire was administered in four countries, targeting experts to understand how they use and store big data and the risks associated with the different types of big data. In addition, the study used quantitative data as a secondary form of data to supplement information not provided by the qualitative data. The utilisation of Atlas.ti, descriptive statistics analysis using Statistical Package for Social Sciences (SPSS), and Excel iterations was a component of quantitative analysis. The Atlas.ti program was used to analyze the data in the following ways: preparation and importation of data, familiarisation of oneself with the data by creating word clouds, coding of the data, and retrieving and examining codes and quotations. According to Woods et al. (2016), the grounded theory procedure was used to convert all transcripts to the qualitative analysis program Atlas.ti (version 9.1) and narrative analysis. For all transcripts, this entailed producing word clouds, quotations, and codes, as well as looking for conceptual connections, differences, and the most important and relevant information. These descriptive statistics were utilised to illustrate the current state of knowledge on big data, cloud computing, and risk assessment management.

This study found that the significant advantage of big data technologies is reducing the cost of storing, processing, and analysing massive volumes of data for organisations. Regarding risks in Canada, there are several requirements to comply with regulations. In Jordan, risks related to data loss due to malfunction or negligence with potential client lawsuits. Data loss, corruption, or disclosure in South Africa leads to legal issues. In addition, South Africa was the only country

challenged with risk assessment. South Africa and Jordan tend to run their big data applications on traditional storage systems. In the UK, individuals have the right to stop or restrict data processing. The study also noted several aspects to mitigate the risk by organisations, such as a mix of education and robust cybersecurity controls, the use of tools to automate the detection of unintended data access, implementation of secure keys, and following certificate management updates. The use of encryption at all points of the data journey and multifactor authentication was also highlighted.

There are several implications from the findings of this study. For instance, organisations need to invest in big data and cloud storage systems to help discover cost-effective and efficient business policies by providing more accurate data. In addition, organisations need to empower a new generation of employees, improve reliability, improve data control, improve availability, ensure automated updates, and ensure easy data backup, recovery, and scalability. Organisations need to find means to reduce the costs of big data storage, particularly in the cloud system; educate employees about best practices and robust cybersecurity controls; invest in infrastructure that enhances encryption at all points of the data journey, and ensure the usage of secure channels for transmission.

Most respondents from the four nations were found to have either intermediate or advanced levels of knowledge. Based on the countries, SA has 80%, Jordan has 60% of respondents having intermediate knowledge, while the UK has 70%, and Canada has 60% of respondents with an advanced understanding of cloud computing. Also, to find the existing risk assessment methods and policies in an organisation, the quantitative analysis of respondents' answers revealed the level of knowledge of risk management in each country. The responses gathered reveal that 50% of respondents in SA had advanced knowledge of risk, 50% of respondents in the UK also had advanced knowledge, 60% of the respondents in Canada had expert-level knowledge, and in Jordan, 40% belonged to advanced and intermediate group respectively. Only SA has 20% of respondents in the limited experience domain and faces challenges in risk assessment.

Keywords

big data; cloud computing; risk assessment; Canada, Jordan, South Africa, United Kingdom

TABLE OF CONTENTS

List of Tables

List of Figures

List of Acronyms

CC: Cloud Computing

CSCs: Cloud Service Consumers

CSPs: Cloud Service Providers

DaaS: Data as a Service

ENISA: The European Union Agency for Cybersecurity

GDPR: General Data Protection Regulation

GT: Grounded Theory

IaaS: Infrastructure as a Service

ICT: Information and Communication Technology

IS: Information Systems

IT: Information Technology

NIST: National Institute of Standards and Technology

NLP: Natural Language Processing

PaaS: Platform as a Service

PIPEDA: Personal Information Protection and Electronic Documents Act

POPIA: Protection of Personal Information Act

QoS: Quality of Service

RA: Risk Assessment

SA: South Africa

SaaS: Software as a Service

SLA: Service Level Agreement

SMEs: Small and Medium-sized Enterprises

UK: United Kingdom

VM: Virtual Machine

Glossary

Big data: Is a term used to characterise the massive amounts of structured and unstructured data that businesses face daily. However, it is not just about the quantity or quality of data but also about how businesses use it (SAS, 2016).

BYOD: Bring Your Own Device is a set of business and technology policies allowing employees to access company applications and services through their personal devices (VMware, 2022).

Cloud computing: Is the provision of various services over the Internet. Data storage, servers, databases, networking, and software are examples of these resources (Frankenfield, 2020).

CSC: is a cloud computing consumer, which encompasses cloud service subscribers and cloud service users. Users in the private cloud infrastructure domain are part of the same company as the provider (Open Group, 20212).

CSP: is a business model for cloud computing services (Garnter, 2013).

DaaS: A method of storing, integrating, processing, and/or performing analytics on large amounts of data in the cloud (data management) (Dataprise, 2017).

Data centre: A data centre (data centre/datacentre/datacenter) is a structure that houses and maintains back-end IT systems and data repositories, such as mainframes, servers, and databases (Gartner, 2013).

IaaS: Virtual machines and other computing resources are typically provided by cloud computing services in the form of IaaS (Dataprise, 2017).

NaaS: Is a category of cloud services that enable customers to access network/transport connectivity services and/or inter-cloud network connectivity services (Dataprise, 2017).

Network: A network of connected computers with the ability to exchange information. A network can range in size from the local area network (LAN) connecting a few personal computers to the Internet, a vast global network of computers. (Gartner, 2013).

On-site (On-premises): Is a technique for distributing software. Computer programmes used with on-prem are loaded directly on users' computers via CDs or USB devices. Off-premise, however, allows you to find the installation anywhere on the Internet. (Open Group, 20212).

PaaS: Platform as a Service; in this model, Cloud providers give a computing platform that includes an operating system, database, and web server (Dataprise, 2017).

Private Cloud (also called internal cloud or corporate cloud): A phrase that refers to a proprietary computing architecture that allows for the safe and reliable delivery of hosted services to a select group of clients. (Dataprise, 2017).

Risk assessment: Is a procedure for identifying prospective dangers and analysing the consequences of their occurrence (Cole, 2019).

SaaS: Software as a Service; a software delivery strategy where software and data are stored in the cloud. Users often access SaaS via a web browser on a thin client (Dataprise, 2017).

Server: A computer in charge of responding to requests from a client programme (e.g., a web browser or an e-mail application) or another computer. Also known as a "file server" (Gartner, 2013).

Virtualisation Is the process of creating a virtual (rather than physical) replica of something, like a hardware platform or an operating system. In hardware virtualisation, the host machine is the actual computer that is virtualised, whereas the guest machine is the virtual machine. Host and guest are often used to distinguish between physical and virtual machines. A hypervisor, Virtual Machine Monitor software, or firmware generates a virtual machine on the host hardware (VMware, 2022).

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 Introduction

This study intends to compare big data risk assessment techniques based on cloud computing technologies in four countries: Canada, Jordan, South Africa (SA), and the United Kingdom (UK). These four countries represent different continents, indicating how different regions handle big data and cloud computing risks. The interaction between big data and cloud computing is getting closer as the latter provides an optimal online working environment. Together, they create a new approach to obtaining, exploiting, and managing information (Deb and Fuad, 2021). Big data is the rapid and analytical use of information that results from processing large amounts of data. Companies generally use big data to understand customers, patterns, and appropriate behaviours that lead to strategic decision-making (Saif and Wazir, 2018), which can help increase competitive advantage (Xie *et al*., 2021). For companies, cloud computing means hiring an external service to process their data into information, which must be treated appropriately. The misuse of cloud information can cause important trust and security issues (Yang *et al*., 2021). In addition, it can be dangerous for customers and to the reputation of companies demanding and offering cloud services (Hashem *et al*., 2015). The amount of data generated by the world's existing devices is astounding. Almost every digital device with an internet connection leaves a digital trail, which necessitates storing data on servers that can be accessed routinely from any location, at any time.

As a result, cloud service providers (CSPs) must constantly assess the threats and risks that could jeopardise the integrity of processing and storing data in any data management stage (Mthunzi *et al*., 2020). Risk assessment (RA) can be essential in identifying and analysing these evolving risks.

Numerous domains, including statistics, economics, systems analysis, biology, and operations research, rely heavily on RA (Djemame *et al*., 2016). There are two sorts of RA in general — quantitative RA, which is objective and quantifiable, and qualitative RA, which is subjective and is based on expertise and know-how.

This study examines a particular component of the risk assessment process related to big data and cloud computing technologies and approaches. More precisely, a service provider (such as a CSP) might employ risk assessment techniques throughout the service's lifecycle: construction, deployment, and operation. Although there is research on RA in the cloud and large data situations, they do not address the subject thoroughly. This thesis seeks to address this gap by proposing a qualitative RA technique for information security in the context of big data and cloud computing.

This chapter begins with an introduction to the central topics of this research. The study's background is discussed in section 1.2, along with the concepts and components of big data, cloud computing technologies, and risk assessment models. Section 1.3 presents the problem statement and purpose of the study by emphasising the critical nature of risk assessment in the context of big data and cloud computing. Sections 1.4 and 1.5 discuss the theoretical and empirical research objectives and questions. Section 1.6 discusses the research methodology and design, section 1.7 outlines the research contribution, and section 1.8 outlines the study chapters.

1.2 Background

This section presents an overview of big data techniques, cloud computing, and risk assessment processes. It also explains the integration between these techniques and some of their main features.

1.2.1 Big data

Big data use has exploded in recent years in areas like healthcare, urbanism (Grybauskas *et al*., 2021), and management (Bifulco *et al*., 2021), among other areas, catching the attention of scholars. Big data represents large datasets that exceed the capability of traditional database software to obtain, clean and process data in a reasonable period. Moreover, large quantities characterise big data speed and diversity of information sets that need new ways of operationalisation to facilitate insight development, management, synthesis, and efficiency at the process level (Raguseo, 2018). As a result, new opportunities and benefits for people and companies arise (Zhang *et al*., 2012).

The main dimensions of big data are volume, variety, and velocity (Raguseo, 2018). The volume dimension shows the quantity of data produced from several sources. Big data involves handling extensive amounts, sometimes in the order of zettabytes. The variety dimension refers to the ways of processing and analysing data. With the increasing number of connected digital devices and platforms, usual data forms have moved from structured to unstructured data, including several images and video formats to text. Velocity dimension refers to the velocity of data frequency from diverse origins, for example, the speed at which platforms such as Facebook and Twitter generate data. A higher speed of data analysis is needed due to the increasing volumes and frequency of data generation.

Other authors have added the following features considering the challenges in saving and analysing data using big data processes (Zanoon *et al*., 2016). For example, veracity represents the data's quality; it represents the precision and reliability of the data. In this instance, data quality is heterogeneous, affecting the precision and processing of the results. Although there is a broad agreement on the potential value of big data, it can reduce its value when it is not precise and does not generate correct insights. The value features show the importance of large data. In addition, the relevance of data following analysis is demonstrated. Data is almost worthless if it does not generate value for companies and their customers. The worth centres on the accurate processing of accurate data and the information and insights it can provide. As the last step, the value comes after processing volume, velocity, variety, and veracity.

Big data usually uses cloud computing services such as those provided by Amazon or Microsoft as a base technology to operate, which eases the creation of decision-support systems (Stergiou *et al*., 2018). This relationship between big data and cloud computing centres on their combination: the cloud can be seen as the storage compartment, and big data can be represented as the item stored in the compartment (Zanoon *et al*., 2016).

1.2.2 Cloud computing

Cloud computing transfers information and computing from home and office computers into external data hubs. Through cloud computing, people can use computing services when and where they need them via the Internet, which facilitates the creation of applications in real-time and faster implementation of services (Mondragón-Ruiz *et al*., 2021). Cloud computing is based on processes, software, information, and storage services. The user or customer does not need to know the configuration and location of the entity that provides the service (Sahu and Pateriya, 2013). In general, cloud computing has several benefits: better resource utilisation and efficiency (Shu *et al*., 2021); optimisation of processes (Goswami *et al*., 2022); organisational agility and easy outsourcing of non-core applications (Karimi-Alaghehband and Rivard, 2019); and scalability and integration with several processes and applications (Belgaum *et al*., 2021).

According to NIST (National Institute of Standards and Technology), cloud computing gives ease of access, with minimal management labour and communication with the service provider, to a range of computing assets such as server farms, storage, and apps through the Internet. In addition, cloud services' overall cost decreases by sharing resources, making it a viable option for companies of many sizes (Sahu and Pateriya, 2013). The NIST defines cloud computing by

explaining five core features, three cloud models, and four cloud deployment models (see Figure 2: Cloud environment architecture in Chapter 4) (Potey *et al*., 2013).

The basic objective of cloud computing is to make greater use of distributed web resources (Kumar *et al*., 2018) and combine them for improved performance to solve big-scale computing issues. However, cloud computing has five essential characteristics, as Potey *et al*. (2013) have stated. Firstly, on-demand self-service has almost no human intervention, and the customers independently access their required services. Customers use information, apps, and storage via the browser independent of other software and hardware. Secondly, broad network access; for example, cloud customers can access several services available via the Internet. Clients do not depend on any platform to gain accessibility. Cloud services are always active no matter the accessing place. Thirdly, resource pooling or computational assets are shared among several users depending on their requirements. Fourthly, in measured service, for example, customers do not have to control and optimise cloud services because the cloud system manages them automatically. Resource usage can be assessed and managed with high transparency for the user and the service provider. Fifthly, selecting providers, which refers to the choice of the cloud service provider in this instance, is central to getting excellent service.

Cloud computing has been used in healthcare (Kishor *et al*., 2021), cryptocurrency applications (Tang and Zeng, 2020), planning (Abdalkafor *et al*., 2021), and education (Mukred *et al*., 2021), among other areas. Cloud computing generates the possibility of managing a large amount of information over the Internet via process and hardware virtualisation, facilitating the modularity, universality, and capability to scale big data (Saif and Wazir, 2018). The life cycle of data contains six steps: create, store, use, share, archive, and destroy. When the data is established, it can move freely between stages (Kumar *et al*., 2018). The main objective of cloud service companies is to maintain information security in each state.

1.2.3 Risk assessment

Cloud computing, a relatively new advanced technology, is susceptible to a wide range of threats if poorly handled (Alshammari and Aldribi, 2021). It generates new security threats that must be examined and reduced, such as data breaches, service disruption, and data confidentiality (Kumari *et al*., 2018). Cloud providers should reassess their delivery and management methods to offer a better value proposition and a more secure process that matches customers' needs (Drissi *et al*., 2013). Regarding big data and risk assessment, the most relevant works focus on big data risk in

supply chain management. For example, Singh and Singh (2019) explore how companies develop risk resilience via big data analytics during events that impact their supply chain. Despite the relevance of this research, it is based on surveys, which may prevent in-depth explorations of the issues analysed. Another work carried out by Araz et al. (2020) evaluates recent research on operational risk management in a big data context. However, this study is based on a literature review. Other research focuses on big data and risk assessment in the context associated with educational administration (Sorensen, 2018), politics (Zhang et al., 2019), and health care (Richter and Khoshgoftaar, 2019) but do not integrate the cloud computing aspect and neither use interviews across countries, which is a major differentiation and contribution this research intends to make. Consequently, the evaluation of security risks is fundamental.

A "risk" is the probability of a threat entity taking advantage of a weakness, which usually impacts businesses and their processes and services (Drissi *et al*., 2013). Risk management involves the steps and procedures to ensure that an organisation mitigates or eliminates risks and reaches its goals. Risk assessment (RA) is an important step in risk management. An organisation can spot threats through RA and assess the likelihood of materialisation, effects, and measures to control or eliminate the effects. Risk assessment's primary objective is to define appropriate controls for reducing or eliminating those risks (Drissi *et al*., 2013). In information systems, there are several critical layers for risk evaluation: information security, software security, hardware security, management security, and environment security (Peisheng *et al*., 2020).

Risk assessment is essential in risk management, defining reasonable measures for eliminating or minimising risks. The following four steps categorize RA (Dioubate *et al*., 2015): 1) Threat identification step, which spots the potential risks that can impact a particular system and its origin; 2) vulnerability identification which seeks to create a set of weaknesses (vulnerabilities) that a threat can take advantage of; 3) risk determination, which evaluates the system risk level; and 4) control recommendation, which seeks to propose measures to minimise or eliminate the threats to the system.

Risk analysis methods can be qualitative and quantitative (Drissi *et al*., 2013). Although many well-developed and large companies use quantitative risk assessment methodologies, they are not frequently implemented in IT. Qualitative risk assessment describes the likelihood of consequences in detail; this method can be implemented when a numeric-based indication of risk is hard to obtain. The European Union Agency for Cybersecurity (ENISA) identified thirty-five incident contexts that fall into one of these classifications: policy and organisational, technical,

legal, and the last one is related to other contexts different from cloud computing (see Table 1: ENISA's list of risk scenarios and their categories in Chapter 3) (Cayirci *et al*., 2016; Zanoon *et al*., 2016). There are also semi-quantitative risk assessment methods that use qualitative and quantitative parameters to assess risk exposure more accurately and comprehensively (Kusi-Sarpong *et al*., 2021); these methods mitigate some of these shortcomings of the qualitative and quantitative methods.

1.3 Problem statement and study motivation

Big data and cloud computing are being critically developed, expanded, and consolidated. Both have enormous advantages, hazards, and problems from various sources and agents (Zhang *et al*., 2012). Information security and anonymity are some of the most crucial problems for cloud services due to their open environment with low user control and high levels of automation. Big data storage and processing affect security and privacy due to the many uses of third-party services and the infrastructure to host and analyse critical data. As big data organisations grow, the risks and security concerns affecting big information operations must be addressed to avoid negative repercussions (Zanoon *et al*., 2016).

Big data security is critical in cloud computing owing to the following elements. Firstly, big data and cloud computing are increasingly used in companies' core activities and critical processes (Kusi-Sarpong *et al*., 2021). Data heterogeneity, volume, and fast data generation and use can increase the risk of data quality degradation (Taleb *et al*., 2021). Secondly, malicious activity and cyber threats keep growing as more operations migrate to the cloud. Thirdly, the algorithms implemented to analyse large volumes of data are opaque. As these algorithms learn by themselves, it is increasingly difficult to know how they operate, reducing transparency. Fourthly, more operations are migrating to the cloud. Therefore, it is critical to understand how the data is stored, managed, and eliminated.

The importance of risk assessment processes in big data and cloud technologies is a ramification of the pressing importance of supporting several incumbents in making insightful decisions, forecasts, and data visualisation. A core element in the characterisation and reporting of threats is implementing the following steps appropriately: 1) analyse the origin of risk events and understand them to visualise their structure; 2 correctly) assess the potential losses related to every event in case they materialise ; 3) predict the probability or potential for the event using either statistical

approaches with probabilistic evaluations or subjective judgments with approximate reasoning (Djemame *et al*., 2016).

The following seven security issues must be considered despite improving cloud technologies when using cloud services: theft, destruction of equipment, data, facilities, updates, patches, and passwords (Williams, 2018). These are privileged user access where data and information generated and processed outside the company develop new risks because it has bypassed its internal controls. Regulatory compliance, whereby, in the end, cloud users are accountable for the accuracy and safety of the data they send to the cloud, even when a service provider holds it. Most of the time, data location references cloud users who are not aware of the location of their information (where the data is stored). Data segregation implies that information in the cloud is generally stored in shared servers. Although encryption is used, it is not infallible. Recovery keeps the information stored if any problem causes failure in the cloud server. Investigative support, in this case, some cloud service programs tend to be complex to investigate because of the multiple settings and places where the information is stored.

In the big data context, there are several dimensions of risks: technological, organisational, institutional, and human (Kusi-Sarpong *et al*., 2021), with technological risks being one of the most important to consider. An adequate risk assessment process in big data services in the cloud allows for keeping the data safe, which is one of the most important issues in this service model so that the clients can trust them. This study analyses different risk assessment processes to determine more robust work paths.

Big data is a critical component of today's businesses of all kinds. Big data presents countless possibilities and a wealth of intellectual capital (Raja and Hanifa, 2017). Using cloud computing, huge amounts of data may be easily shared, archived, and destroyed in various ways (Rao and Rao, 2021). However, some inherent hazards must be considered regarding big data. In the event of a data breach, confidentiality issue, or threat arising from data availability, firms and their systems could suffer significant damage (Rajen and Prasad, 2015). There are dangers to the companies from the deliberate misuse of data-driven technologies by malicious parties and the dangers of falling into corporate vulnerability. There are several sources of big data, including websites, electronic check-ins, databases, sensors, call records and many more types of information. All data and information can be stored in service providers' hardware via network-based storage in cloud computing. These can be protected using risk assessment procedures (Sivasubramanian et al., 2017).

The empirical evidence demonstrating the risks associated with the cloud-based management of large data has been derived. Van Der Schyff and Krauss (2014) conducted a theme analysis of semi-structured interviews with twelve cloud computing security specialists in South Africa. The researcher analysed virtualization-related security vulnerabilities by combining the issues in a classified manner, and the results revealed that multitenancy, harmful insider presence, and shared application usage raise data privacy and protection risks.

Using a quantitative analysis of one hundred and thirty-four participants, Hammouri and Abu-Shanab (2020) investigated the significance of cloud computing and the usefulness of cloud services in Jordan. The findings revealed that a user's decision to adopt cloud services is influenced by cost reduction, quality, and control augmentation. Machuga (2020) investigated cloud computing utilisation in European nations by secondary research based on Eurostat figures, and the results indicated that the adoption of cloud technology was extremely effective in European nations. People's familiarity with the technology and the existence of a robust infrastructure contributed to the safety and security of cloud computing usage. Using thematic analysis and the expert elicitation method, Cains et al., 2021 interviewed forty-four cyber security professionals about cyber security offences and likely attack causes. The outcomes of the interviews highlighted a need for developing an integrated sub-system that allows for human intervention to preserve secrecy, ensuring a fast and reliable flow of data, regular compliance, and control over intrusion via confidentiality.

1.4 Study objectives

This study offers a comparative analysis of big data RA procedures based on cloud computing. An outline of this study's primary and secondary goals is provided below.

1.4.1 Core objective

The primary purpose of this study is to examine and compare large data RA procedures relying on cloud computing technologies in four countries: Canada, Jordan, South Africa, and the United Kingdom.

1.4.2 Secondary objective

This section presents both theoretical and empirical secondary goals.

1.4.2.1 Theoretical purpose

This study uses a literature review to try to understand RA in big data and cloud computing. This study further seeks to analyse the literature on the theoretical dynamics between RA, cloud computing, and big data. A deeper theoretical knowledge of these processes will allow the use of RA to mitigate the risks for big data in cloud services. Specifically, the theoretical objectives involve:

- outlining the evolution of the research on RA involving big data in cloud services;
- compiling the main results of the studies on RA in big data and presenting a theoretical framework aggregating these studies, and
- compiling the best practices of RA in big data and cloud computing emerging from the scientific literature.

1.4.2.2 Empirical objective

This study is conducted practically using a qualitative methodology from the interpretive research paradigm. Specifically, the empirical objectives are to:

- Explore the critical features of RA in Canada, Jordan, South Africa, and the UK environments involving big data on cloud services; and
- Develop managerial and policy recommendations on RA on big data in cloud computing environments using data from Canada, Jordan, South Africa, and the UK.

1.5 Research questions

The main research question explored in this study is "What are the risks associated with big data based on cloud computing technologies in Canada, Jordan, South Africa, and the United Kingdom?"

1.5.1 Secondary research questions

Specifically, the following theoretical research questions are asked —

- How has the research on RA involving big data in cloud services evolved as it impacts risk levels in Canada, Jordan, South Africa, and the United Kingdom?

- What theoretical framework can be proposed by aggregating different studies to inform the best practices of RA in big data and cloud computing in Canada, Jordan, South Africa, and the United Kingdom?

1.5.2 Empirical research questions

Since this study was conducted practically from the interpretive research paradigm using a qualitative methodology, the following empirical research questions are posed —

- What are the critical features of RA in Canada, Jordan, South Africa, and the UK environments involving big data on cloud services?

- What managerial and policy recommendations can be developed to mitigate big data risk in cloud computing environments using data from Canada, Jordan, South Africa, and the UK?

1.6 Methodology

1.6.1 Methodology of research

This research aimed to conduct a comparative study of risk assessment processes for big data based on cloud computing technologies in Canada, Jordan, South Africa, and the UK. This increased a profound understanding of the phenomena being investigated within these countries' contexts and interpreted the meanings individuals attribute to the observed events (Orlikowski and Baroudi, 1991). Due to its ability to gain an in-depth understanding, this study is ideal for an interpretative research paradigm. Moreover, this paradigm works well when the issue is analysed and understudied, such as risk assessment in big data in cloud services. Since the issues are analysed in their context in interpretative research, obtaining more meaningful and grounded results is possible. Furthermore, the results will be attached to cloud computing and big data; hence, they will be more pertinent, leading to better policy and managerial recommendations.

This research implemented some grounded theory techniques. Consequently, the results were tied to the gathered data and information, increasing the conclusions' veracity and trustworthiness. In grounded theory, the researcher is forced to compare and analyse different constructs, which allows a critical understanding of them.

This study was complemented by elements of a systematic literature review, which involves selecting preliminary keywords and validating them; selecting inclusion and exclusion criteria; selecting indexing databases to analyse; screening and scanning documents; reviewing the full text of selected documents, and analysing and generating the theoretical framework and the conclusions. In addition, the study collected qualitative data to extract experts' knowledge regarding a clear risk assessment plan in their respective countries, what is covered in their current risk plan and their knowledge about the risks related to data protection. Atlas.ti Software was used to analyze the qualitative interview data, allowing for the creation of word clouds. These word clouds provided useful information on the codes used to generate the themes. The themes were further investigated in Atlas.ti to find patterns that would respond to the research goals. A thematic analysis was used to classify patterns or themes from the qualitative data. Furthermore, the thematic analysis allowed for identifying noteworthy or intriguing patterns in the data and using the themes to answer research questions or express an opinion on a topic. This was much more than summarizing the material and making sense of it as a sound thematic analysis (Braun and Clarke, 2006).

The paradigm chosen to frame this investigation is described and discussed in detail in Chapter 4. It shows the numerous research approaches and reasoning processes used in information systems (IS) and digs into the application thereof to this study. The justification for the methodology's adoption is also explained. In addition, Chapter 4 outlines many forms of research designs before coming to a final one for this study. The techniques of data gathering, target population, sample procedure, size, and data processing approach are all represented. Finally, Chapter 4 illustrates the quality and rigour criteria in data dependability and validity, as well as the ethical concerns for conducting this study.

1.6.2 Rigour and evaluation of the method

Several techniques were implemented to increase the rigour of the method and ensure the quality of the results. These techniques include data and researcher triangulation, where several sources of information were used to contrast claims and increase trustworthiness. Several researchers can assess if the data analysis is correct and if this analysis leads to the proper conclusions. In addition, data gathering and analysis were thoroughly described so an external researcher or reader could understand the results. Lastly, the use of examples and description of research events were conducted during the interviews. Examples of real issues were asked to ensure that the information provided is grounded, relevant, and pertinent.

Klein and Myers (1999) propose principles to evaluate interpretive research in information security. Firstly, the hermeneutic circle's fundamental principle is that human knowledge is reached by iterating independent or isolated and comprehensive or systematic meanings. Secondly, this principle serves as a base for the other following principles —

1 *Contextualisation* is important to explain the context and history of careful research to facilitate others to understand the question being examined.

2 The *principle of interaction between the researcher(s) and their subjects* refers to how the elements being studied and the researchers interact to build socially constructed theories, hypotheses, or arguments.

3 The *principle of abstraction and generalization* demands the contrast of the idiographic detail provided by the data references the general constructs that define the nature of human knowledge and social action through principles one and two.

4 Dialogical reasoning involves understanding conflicts in the previously formed theories, which leads to studying design and outcomes with consecutive iterations and revisions (what the data shows).

5 The *multiple interpretation principle* requires an understanding that there can be various interpretations of the same study object.

6 The *suspicion principle* includes knowing that the information gathered and conclusions can be partial or erroneous.

The above principles assure strict implementation of the study technique in this research.

1.7 Research contribution

These study findings are expected to help cloud service providers improve risk assessment during the construction, deployment, and operation of cloud-integrated big data services. In addition, a comparative analysis allowed the development of strategies for improving risk management of the cloud services offered by companies. Cloud computing is pervasive. Small, medium, and large companies, in one way or another, use these services to monitor their resources, carry out planning activities, monitor their processes, make payments, buy supplies, and sell their product online. This study identified critical elements to be considered when safely deploying these services.

This research intends to generate new knowledge that contributes to cloud computing, big data, and risk assessment literature by offering a deep understanding of the interactions between these concepts. A better understanding of these concepts will facilitate the scholarly debate and discussion on risk assessment in the cloud and big data environments, taking more and more relevance in academic discussions.

1.8 Classification of chapters

This study is divided into the following chapters:

Chapter 1: Introduction

This chapter provides an overview of the introduction to the study, background and problem context, problem statement and study motivation, the study's objectives, the research questions, methodology, and the research contribution.

Chapter 2: Literature review: Big data and cloud computing

This chapter presents the main research areas and literature on big data and cloud computing.

Chapter 3: Literature review: Risk assessment in big data and cloud computing

This chapter presents the main research areas and literature involving risk assessment in big data and cloud computing contexts.

Chapter 4: Research methodology and design

This chapter describes the methods implemented to reach the objectives and how to ensure the results' quality.

Chapter 5: Results

The results of this study and its importance for the risk assessment and big data environment are presented in this chapter.

Chapter 6: Discussion

This chapter critically assesses the results and integrates them with the literature review. Finally, the chapter outlines further questions and research paths to explore.

Chapter 7: Conclusion and recommendations

This chapter presents the main conclusions and highlights the study's main results.

1.9 Summary

This chapter provided the background and problem context, the problem statement and study motivation, the study objectives, the research questions, the methodology and the research contribution of this study. The notions of "vulnerability," "threat," "risk," and "exposure" are different; however, they are mistakenly used as if they are the same. A good result of RA can generate the right risk management mechanisms to minimise risks and attain adequate levels of safety. When integrating cloud computing with big data, various drawbacks must be considered. The safety of the big data cloud environment is the most critical consideration. Several vulnerabilities concerning big data integration into a new and unfamiliar cloud platform have emerged. One of the main big data cloud security vulnerabilities is based on platform differences. Many big data applications require setting a new platform in the cloud while the cloud's existing security tools and methods may not properly work for these implementations — these new platforms necessitate developing new safety measures. Encrypting data, authenticating processes, regulating access, detecting interferences, and constantly monitoring and tracking occurrences are some security measures that could be used. Besides security policies, plans for consolidating big data should be considered during integration with the cloud environment. This research explores these issues via literature review in chapters 2 and 3, and interviews and discussion are presented in chapters 5 and 6. The next chapter outlines the literature involving big data and cloud computing contexts.

CHAPTER 2: BIG DATA AND CLOUD COMPUTING

2.1 Introduction

This chapter provides a synthesis and a brief introduction to big data and cloud computing literature. Section 2.2 focuses on big data and discusses its historical elements, features (volume, variety, velocity, integrity, and value); data type (unstructured, structured, and semi-structured); and technologies. This section also presents some concepts associated with big data (big data engineering, non-relational model, big data analytics, big data models, schema-on-read, NoSQL, and the big data paradigm); processing and resource management; benefits and uses of big data; and finally, challenges and problems.

Section 2.3 emphasises cloud computing. This section briefly presents the origins of cloud computing; its key technologies (virtualisation, mass distributed storage, parallel programming, and data management); and its uses (infrastructure and platform cloud storage, disaster recovery, test and development, analytics, and data backup). This section also discusses the cloud computing environment's framework; cloud computing characteristics (on-demand self-service computing, resource pooling, and measured service); broad network access and provider selection; cloud delivery models ((Software as a Service (SaaS), Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Data as a Service (DaaS)); as well as deployment models (private, public, hybrid, and community). This section discusses big data and cloud computing (data storage, variety of data, data transfer, and privacy and ownership).

2.2 Big Data

This section discusses the evolution of big data since 2000. The size and comprehension of data evolve as we humans evolve in culture and technology. The evolution of big data discussed in this section reflects that when the amount of data becomes too large and uncontrollable, humans build systems to interpret and analyze the data. The characteristics of big data are discussed as they affect the risk environment. Big data processing and resource management impact risk assessment processes for big data based on cloud computing technologies in Canada, Jordan, South Africa, and the United Kingdom.

2.2.1 Early days

Since 2001, the problem of big data has been progressively debated in many areas when Gartner Co. issued their "3Vs" definition of big data (volume, speed, and variety) (Laney, 2001). However,

a single definition of big data is unavailable; its meaning differs between academia and industry. "Large, diversified, complex, longitudinal and/or distributed data sets from instruments, sensors, Internet transactions, email, video, clickstreams, and any other existing digital sources, both today and in future, is described in the US National Science Foundation" (National Science Foundation (NSF), 2014), while Gartner (2013) argues that big data is characterised by huge volumes of information, high speed and/or wide diversity, which necessitate new kinds of processing to allow for better decision-making, insight, and optimisation of processes. While attending the Beijing Xiangshan Science Congress, at which data science and big data were discussed, specialists from outside China came up with two definitions of "big data" (XSSC, 2013). The simplest way to describe the academic and business worlds is a complicated, diverse, heterogeneous set of data that has a high potential value, making it challenging to manage and evaluate in a reasonable length of time. Strategic resources in the digital era are a unique characteristic of strategic resources crucial to the innovation that alters how people innovate and interact (Xu and Shi, 2015). Since the amount of information created has increased substantially (Figure 2.1), reaching enormous amounts of big data is more crucial.

Real-World Data Storage Capacity	
ITEM	**CAPACITY REQUIERD**
Amount of test on an average print page converted to HTML	1 kilobyte
Typewritten page	2 kilobytes
Short novel	1 megabyte
Downloadable MP3 music file	2 megabytes to 5 megabytes
Pickup truck filled with books	1 gigabyte
50,000 trees made into paper and printed	1 terabyte
Information contained in all U.S. academic research libraries	2 petabytes
All words ever spoken by human beings	5 exabytes

Figure 2.1: Real-world data storage capacity

Source: Adapted from Wenhong and Yong (2014)

"Big data" became a new trend in the Internet and IT industries in 2009. Initially, most uses for big data were in the Internet industry: Internet data are up 50% annually and double every two years. Most Internet firms developed the "big data" age and the significant relevance of data (Wenhong and Yong, 2014).

To correctly run big data applications, it needs vast amounts of data, which can be processed via several models depending on the goal of the analysis (Rayala and Kalli, 2021). Depending on the

computational power, some big data applications can consume a lot of energy and time (Rayala and Kalli, 2021).

Big data have been used in many applications. Some include medical diagnosis and imaging (Xin and Fan, 2021). One of the significant challenges in this application is to compress images without distorting them (Xin and Fan, 2021); some researchers use powerful compression algorithms to avoid distortion. Big data applications are also used in innovative city projects to process the tremendous amount of information that cities generate to facilitate policy-making, real-time decision-making, assessment of resources, and planning activities at the city level (Xie and Zhang, 2021). Big data is also being used to analyse urban mobility issues and traffic patterns (Sadowski *et al*., 2021); detect patterns and behaviours related to autism (Uljarević *et al*., 2021); and predict consumer behaviour (Al-Marsy *et al*., 2021).

2.2.2 Big data characteristics

Big data have several features. Firstly, heterogeneity; big data applications can handle different data types from diverse origins (Taleb *et al*., 2021). Secondly, big data can be suited to process large amounts of data while consuming less time and resources (Banchhor and Srinivasu, 2021). Thirdly, insight generation; some big data processes can uncover surprising and unforeseeable correlations and links (for example, mobility and pandemic development) that are impossible to detect through qualitative methods (Sadowski *et al*., 2021). Fourthly, predictive capabilities; some big data applications retrieve historical data and turn it into valuable information to predict the likelihood of several events (e.g., purchases, customer behaviours) to occur (Chaudhary *et al*., 2021). Fifthly, speed; some processes are carried out faster through particular and optimised applications that would take much time if carried out via traditional methods (Taleb *et al*., 2021).

Moreover, according to (Zanoon *et al*., 2016), big data has several additional features, also known as the *five Vs*:

Volume: indicates the quantity of data generated from many sources; the amount of data is frequently in the order of zeta. The volume is the most obvious dimension in big data.

Variety: denotes different types of data; as the number of Internet users, cell phone users, and social network users increase, the familiar form of data has evolved from structured database to unstructured and now encompasses a diverse range of formats, including images, audio and video clips, SMEs, and GPS data.

Velocity: indicates the speed of data frequency created, transported, and stored from multiple sources, that is to say, the speed. The huge rise in the amount and frequency of data necessitates a super-speed data analysis system.

Veracity: indicates data quality and correctness. The quality of the data collected might vary considerably, affecting the accuracy of the analysis and the findings. While there is widespread agreement on the potential use of big data, it is almost worthless if inaccurate.

Value: It demonstrates the relevance of large data (the importance of data after analysis; data on its own is almost worthless if it is not valorised). The value lies in thoroughly analysing the data and its information and ideas. In the final processing stage, volume, speed, diversity, and integrity determine the value.

Big data is derived from several sources, including sensors and unstructured data from social media, metadata, other online blogs, geographical data from GPSs, and medical devices. For Zanoon *et al.* (2016) and Shameli-Sendi and Cheriet (2014), big data is collected from many sources and produced in diverse forms:

Structured data: The organised data is handled in the form of tables or databases.

Unstructured data: Individuals create most of the data daily, including texts, images, videos, communications, logs, and clickstreams.

Semi-structured data: It is regarded as organised data, but not as tables or databases, in the same way XML documents or JavaScript Object Notation (JSON) are regarded.

2.2.3 Big data technologies

Big data technologies are software tools for analysing, processing, and extracting data from extremely complex and huge data sets that typical management systems cannot handle. Therefore, investment in big data technologies can be seen as mitigating risks in Canada, Jordan, South Africa, and the United Kingdom.

Without a doubt, data is the foundation of contemporary information technology and scientific research services, and big data processing technologies have been a key focus of current information technology development. Five distinct fields of big data technology exist, each with its tools (Figure 2.2). On the one hand, Big Data Processing's blossoming technology marks the commencement of a new wave of the information technology revolution. On the other hand, as

economies restructure and industries modernise, information processing technology will grow in importance. Big data processing technologies are the most effective means of developing core technology, pursuing development, innovating applications, and reducing the nation's pillar industries' computerization lock-in (Wenhong and Yong, 2014).

Classification of big data technologies	Big data technologies and tools
Infrastructure support	Cloud Computing Platform Cloud Storage Virtualization Technology Network Technology Resource Monitoring Technology
Data acquisition	Data Bus ETL Tools
Data storage	Distributed File System Relational Database NoSQL Technology Integration of Relational Databases and Non-Relational Databases In-Memory Database
Data computing	Data Queries, Statistics, and Analysis Data Mining and Prediction Graph Analysis BI (Business Intelligence)
Display and interaction	Graphics and Reports Visualization Tools Augmented Reality Technology

Figure 2.2: Classification of big data technologies

Source: Adapted from Wenhong and Yong (2014)

2.2.4 Big data: General concepts

The aim of storing and collecting vast data volumes is to conduct analyses that provide further information and insights. Selected data at random has typically been researched in the past. According to El-Seoud *et al*. (2017), there are several concepts associated with big data —

Big data engineering uses a network of horizontally linked resources to achieve near-linear performance scalability. The data layer's innovative technology was driven by the growing relevance of data types that cannot be managed well within a relational paradigm. Due to the

requirement for scalable access to organised and unstructured data, columnar (big table) and document- and graphics-oriented software has been developed.

Non-relational model refers to logical data models such as the document, the graph, the key value, and others used to increase the efficiency of non-tabular datasets.

NoSQL: NoSQL (alternately called “no SQL” or “not only SQL”) refers to data storage and interfaces not linked to close relations.

Big data models refer to logical data (non-relational and relational) and processing/computing (batching, streaming, and transactions) architectures in which data is stored and manipulated over horizontally scaled resources.

Schema-on-read is frequently kept in a raw form depending on production, with the system needing to organise (and often clean) the data detected and processed when the data is being requested. Keeping Scheme-on-read in a raw form is essential because the data must be organised to execute many analytics’ processing frameworks or algorithms.

Big data analytics capabilities and programming approaches are rapidly evolving. These analytical functions enable the simultaneous integration of findings from disparate components of one or more data sources.

The big data paradigm comprises the horizontally coupled independent resource distribution to provide the scalability required to process large data volumes efficiently.

Big Data and Cloud: The connection between big data and cloud computing is based on their integration; the cloud is a warehouse, and big data is a digital element stored in the warehouse since storage spaces without digital resources are difficult to build (Zanoon *et al*., 2016).

2.2.5 Big data processing and resource management

Proper data processing and management of big data can reduce risks for big data based on cloud computing technologies in the four countries under study. The management of big data processes is increasingly taking energy-efficient and green methods into account. They allow carrying out highly performant activities while using fewer resources (Stergiou *et al*. 2020). Another critical element in big data management is harmonising several applications and data sources and normalising the information to facilitate operations and decision-making, all in an automated way (Sellami *et al*., 2020). Some authors (Oktian *et al*. 2020) propose a framework for big data

management that involves three steps: transit, which consists of the transmission of information from several resources; rest, which involves storing the data while conserving its integrity; and process; which implies carrying out operations on the data to generate useful information.

When attempting to comprehend the concept of big data, it is impossible to ignore the terms "MapReduce" and "Hadoop."

Hadoop is a free Java-based software platform for distributed data processing. It is an Apache project. Hadoop clusters are designed in a Primary/ Secondary approach. Hadoop can process enormous data volumes utilising a cluster of servers and applications on systems with thousands of nodes and terabytes of storage. The distributed Hadoop file system speeds up data transmission and ensures that the system continues to work even if some nodes fail. This method dramatically reduces the likelihood of a system-wide failure, even in the event of numerous node failures. Hadoop is a cost-effective, scalable, adaptive, fault-tolerant distributed computing platform. Well-known firms like Google, Yahoo, Amazon, and the International Business Machines Corporation (IBM) use the Hadoop Framework to enable applications that require massive amounts of data. Hadoop is divided into two major subprojects: MapReduce and the Hadoop Distributed File System (HDFS) (Wenhong and Yong, 2014).

MapReduce is a framework for writing programs that process huge volumes of data in a reliable, fault-tolerant manner concurrently with commodity hardware clusters. A MapReduce task divides the data into multiple parts that map jobs execute in parallel. The map outputs are then added to the reduced jobs in the order in which they were generated. A task's input and output are frequently kept in a file system. The framework is responsible for scheduling, monitoring, and re-executing of tasks.

Hadoop Distributed File System (HDFS) is the Hadoop data storage network's file system. It connects local file systems to a large file system. The HDFS improves dependability by multi-source data replication to overcome node outages.

2.2.6 Big data benefits and uses

The benefits of big data can be viewed from the ability of organisations to establish adequate risk assessment processes for big data based on cloud computing technologies in the four countries under the study. It is possible to decide the complete scope of organised, semi-structured, and/or unstructured data using big data. Because vast values cannot be analysed with small data sets, large

data sets offer enormous possibilities for developing expectations in all areas through rapidly updated information technology. For instance, scientific investigations in biological sciences, high-energy physics, astronomy, geosciences, and remote sensing may improve their potential to discover previously unknown knowledge using big data. Policymakers can use massive data to examine government policies and plans methodically, while businesses mine big data for profit/earnings, useful consumer insights, and marketing shares (Xu and Shi, 2015).

Big data has many uses; some of them include (Bello-Orgaz *et al.*, 2016):

Marketing: Marketing academics think that large cloud computing and social media analytics provide companies with a unique chance to collect feedback from numerous clients and improve existing methods. Leading e-commerce companies like Amazon and eBay have undergone a major market revolution through their unique e-commerce systems that are extremely scalable and recommended (Dauriz *et al*., 2014). Social network research collects user information and allows companies to develop more specific marketing and advertising campaigns.

Crime analysis: Criminals tend to exhibit repeated pattern behaviour, which depends on situational variables. Crime is concentrated in surroundings with traits that encourage crime (Wortley and Mazerolle, 2013). Crime data analysis aims to find such crime patterns, enabling the identification and discovery of crimes and their linkages to offenders. Data mining tools can aid law enforcement officials in gaining knowledge. Telephone, face-to-face conferences, email, and other digital forms of communication are the primary ways of communication between public and government institutions. Most of these discussions are recorded or translated into text and then kept in digital format, allowing for automated text analysis employing natural language processing (NLP) technology to increase law enforcement's effectiveness (Knutsson *et al.*, 2012). Ku and Leroy (2014) present a decision-making assistance system that integrates NLP methods, classification methodologies, and similarity measures to automate and ease criminal analysis. Filtering and detecting reports of the same or similar incidents can assist in analysing the crime pattern, enable suspects to be apprehended, and enhance crime prevention.

Epidemic intelligence: The early detection, evaluation, and checking of possible hazards in public health can be characterised as epidemic intelligence (Paquet *et al*., 2005) and the supply in good time of the corresponding notifications. Social networks, blogs, digital media, and government-sponsored sources of unstructured free text and media output are all included in this issue, as are monitoring systems for automated and ongoing analysis. The biomedical text was mined using

approaches for recognising and categorising texts, extraction of terminological terms, and extraction of relationships (Cohen *et al*., 2018). Human language processing algorithms translate unstructured text data from large collections into a predefined format and filter it. In published articles, terms connected to illnesses or symptoms are discoverable (Lampos and Cristianini, 2012). However, terms can be tricky because the same term might refer to several matters according to the context.

User experiences-based visualisation: For improved user experiences and services, big data from social media must be shown. For example, the large numerical volume of data (typically in tabular form) can be converted into other forms. User understanding ability can, therefore, be improved. For different areas, such as corporate success, treatment, cyber, national security, and disaster management, it is necessary to enable quick choices based on the visualisation of such vast data (Keim *et al*., 2013). Therefore, user experience-based visualisation is vital to helping decision-makers make better judgments. Visualisation is also an important social media data analytics tool (Kotval and Burns, 2013). It is vital to understand the demands of consumers on social networking platforms. Many visualisation techniques have been developed to gather (and improve) user experiences. Interactive data analysis is one of the best known. Users can interact with the visualisation-based analysis system by selecting a subset of the given big data's characteristics.

2.2.7 Big data challenges and problems

Understanding big data challenges is an important step toward finding solutions for risk assessment processes for big data based on cloud computing technologies in Canada, Jordan, South Africa, and the United Kingdom.

2.2.7.1 Big data challenges

Big data, as with many other technical advances, faces many challenges

Quality: Since big data applications work with vast and diverse amounts of data, it is sometimes challenging to normalise and clean them to have accurate insights (Taleb *et al*., 2021). Hence, it is essential to put in place quality management measures in big data processes.

Non-regulation: Some big data applications have not been regulated in some countries. Some companies, for example, use big data to analyse detailed information about people to offer targeted ads and lure them into buying, in a very subtle way, things they do not need (Picciotto, 2019).

Faulty and biased algorithms: Data can be processed with algorithms filled with errors and biases (Kuhlman *et al*., 2020). This constant challenge is being addressed by increasing, when possible, the transparency of the calculations and processes carried out by those algorithms.

Moreover, according to Xu and Shi (2015), there are numerous problems connected to big data:

The first essential problem is how large data gathering and management can be efficiently acquired, stored, and documented. The majority of big data is unstructured and unorganised. While MapReduce (Hadoop) technology may be used to collect huge amounts of data, computer science's traditional data acquisition and management techniques must be augmented with management science skills. For example, the organisational strategy of using big data must be addressed before gathering the big data. A large database's basic design and administration should be based on data capacity, value, ethics, ownership, policy, the guarantee of quality, etc. (Sivarajah *et al*., 2017). Big data may play a vital part in making a practical choice using management science.

The second major problem is the availability and processing of big data. The various forms and characteristics of large data contribute to difficulties in evaluation, particularly in data processing and interpretation. Numerous existing techniques for information science are capable of addressing this challenge. Because most data mining and machine learning techniques are designed to handle structured data, it is impossible to directly evaluate vast amounts of semi-structured and unstructured data. Notably, contemporary IT cannot reasonably process enormous amounts of semi-structured and unstructured data, such as clustering millions of text files, photos, or both. A method for transforming semi-structured and unstructured data into structured or pseudo-structured data is necessary. Many existing data mining or machine learning algorithms can be analysed (Shi, 2014). This transformation process can be accomplished through current document retrieval methods, such as document metadata and web pages. Certain information recovery techniques may turn each text file into a single record in a structured or pseudo-structured format with various features for a particular transformation purpose.

Similarly, a picture may be converted to record the changed format using a pattern recognition technique. The structured or pseudo-structured layout will also vary as the objective transformation changes. As a result, information science skills may be successfully applied to access and process large amounts of data.

The third major problem is applying mathematical and statistical ideas and methods to the mining and interpreting of large amounts of data. The analysable large data formats enable all current

mathematical and statistical tools for big data analysis. Modelling procedures may include, but are not limited to, parenting and sampling of spaces; grouping, regression, classification, and relevance analysis; forecasting and variable selection for data mining methods; and latent analytical variable and statistical inference for analysis methods. The problem reflects when and how the big data mining situation is applicable. Because big data transformation is subject to a predetermined aim, choosing a data mining or knowledge discovery approach might be helpful (Baesens, 2014). Like conventional data mining methods, experimental technique selection design should, for the most part, be carried out in big data mining. However, big data mining discoveries must be assessed against the user's judgment, as knowledge varies by person and circumstance (Yu *et al*., 2014). To aid in the user's comprehension of big data mining, simplified representations of the complexity of big data may be demonstrated using various representation and visualisation techniques, such as standardised systems.

The fourth critical issue is determining how to apply big data analysis skills in real-world applications. This might become an engineering challenge. Superior data understanding contributes to advancing current scientific, economic, or social conditions in most situations. Today, big data pervades every facet and event of human society. Finally, the data-driven approach is always the most dependable option. A strong technical design for big data applications is the greatest approach to gaining scientific, societal, and/or commercial advantages (Baesens, 2014).

2.2.7.2 Big data problems

According to Alnemr *et al*. (2016), big data for the potential value it holds is now becoming an unseen "gold mine." As production, transactions, management, control, sales, customer service, and other data are growing, and the number of users is increasing, evaluating correlation patterns and trends from a huge quantity of data enables efficient management, accuracy, and targeted marketing. However, standard IT infrastructure, data management, and analytical techniques are difficult to adapt to the rapid growth of massive data, resulting in a slew of complications (Alnemr *et al*., 2016):

Classification of big data problems	Description
Speed	Import and export problems
	Statistical analysis problems
	Query and retrieval problems
	Real-time response problems
Types and structures	Multisource problems
	Heterogeneity problems
	The original system's infrastructure problems
Volume and flexibility	Linear scaling problems
	Dynamic scheduling problems
Cost	Cost difference between mainframe and PC servers
	Cost control of the original system's adaptation
Value mining	Data analysis and mining
	Actual benefit from data mining
Security and privacy	Structured and nonstructured
	Data security
	Privacy
Connectivity and data sharing	Data standards and interfaces
	Protocols for sharing
	Access control

Figure 2.3: Classification of big data problems

Source: Adopted from Alnemr *et al*. (2016)

Data leakage issues: There are two consumer data modifications in the cloud. First, the data will be stored away from the local system of the consumer. Secondly, the data moves from a single tenant to a multi-locator environment. These modifications may pose a major risk known as data leaking. Data leaking has become one of the most serious security corporate threats (Santos *et al*., 2014).

Cloud security issues: Malicious apps are prevalent in big data operations and constitute a continual danger (Han *et al*., 2020). The Internet is a communication infrastructure for cloud service providers that provides Internet authentication via well-known TCP/IP protocols. Like a computer connected to the Internet, a virtual machine has an IP address. These IP addresses can also be found by a malicious user, internally or externally. In this scenario, a malicious user can detect the physical servers that the victim uses by inserting a malicious virtual machine to conduct an attack. If a hacker steals or takes control of a virtual computer, they access all user data within it. The hacker can thus copy the stolen information to its local computer before the cloud provider recognises that the virtual machine is controlled by a hacker (Santos *et al*., 2014).

Reliability: Big data applications work with the data they gather. That data may underemphasise or overemphasise the available data. For example, when certain big data algorithms make decisions on overpopulations or situations in which it has not collected information, reliability problems may arise because it makes wrong generalisations (Picciotto, 2020).

Erroneous data gathering: When the gathered data is faulty or retrieved from the wrong resources and in the wrong way, the operations and calculations implemented on them may lead to inaccurate results (Saez and Corchado, 2019) with huge economic, social, and even health consequences depending on the sector in which the big data applications are being used.

Sample sizes: Some big data applications cannot work with small amounts of data. This is problematic since, in some cases, there is not enough data and information to explore, assess and study several phenomena and objects (Konietschke *et al*., 2021).

2.3 Cloud computing

In this section, the origins of cloud computing are discussed. Understanding the origins of cloud computing is important to provide a clear picture of how the various aspects of cloud technology worked in the past and how it came to work the way it is now. In addition, the origins of cloud computing help to appreciate the evolving ideas and recognize them as meaningful products of specific times and places. With the study focusing on risk assessment processes for big data based on cloud computing technologies in Canada, Jordan, South Africa, and the United Kingdom, understanding the origins of cloud computing will mitigate risks and improve the management of such risks.

2.3.1 Origins

Regalado (2011) asserts that the term cloud computing emerged in 2006 when huge corporations such as Google and Amazon began referring to a new paradigm in which users increasingly access software, computing power, and data via the Internet rather than their PCs. However, the Technology Review tracked the term's creation at an office park outside Houston around the middle of the 1990s. Netscape's web browser then sparked interest in the technology. A tiny group of technology executives at Compaq Computer's headquarters monitored the Internet industry's future and coined "cloud computing." The group has a precise and proactive vision. Not only would all business software migrate to the Web, but it became common to refer to consumer file storage as "cloud-enabled computing apps." For George Favaloro (Compaq), Cloud computing might establish a $2 billion-a-year business offering servers to Internet service providers.

In contrast, cloud computing is a formula for dissatisfaction and insolvency for young technologist Sean O'Sullivan. However, its adoption is accelerating. Cloud computing reflects a historical transition in the information technology sector, with more computer memory, processing capacity, and applications hosted in remote data centres or the cloud. The term *cloud computing* has become a contested prize for billions of dollars in IT investment. In 2008, Dell dragged out the programmers' indignation over gaining a cloud computing trademark. Other technology suppliers, including IBM and Oracle, were accused of "cloud washing" or misusing the word to describe outdated products. Like "Web 2.0", cloud computing has become a universal jargon that many technical managers find irritating and impossible to ignore. In information technology, computation and data are moved from desktop and portable computers to large data centres. Cloud computing enables anybody to access on-demand software and computer services from any location, at any time, over the Internet. The term "cloud computing" encompasses a wide range of technologies and services, including computers, software, data access, and data storage, all of which do not require the end user to be aware of or configure the system's physical location. (Sahu and Pateriya, 2013).

2.3.2 Key technologies of cloud computing

Several technologies make possible the implementation of cloud computing. Technology in cloud computing is seen as having a mitigating role in the risk assessment processes for big data based on cloud computing technologies in the four countries under study. According to Zhang *et al*. (2012), the main technologies involve:

Virtualisation: This is a way to deploy computer resources. To achieve dynamic architecture and the goals of central management and dynamic use of real and virtual resources while increasing the flexibility of application systems, including hardware, software, data, networking, and storage, it connects the data centre and all of these elements. Cloud virtualisation integrates servers, storage systems, network devices, software, and services. These solutions combine virtualisation technologies such as hardware, network, application, and desktop virtualisation. The mechanisms of virtualisation, different environments, and experiments can be quickly imitated without the necessary hardware and physical resources, as well as the purpose of building operative systems and applications to increase the security and managing environment and to put them in the production environment in a simplified and effective manner at a later stage. This gives more flexibility and identifies possible conflicts faster.

Meanwhile, server virtualisation techniques can be used to integrate a large number of dispersed and underperforming physical servers with more autonomous and aggregated physical servers or even to form a massive virtualised network system, thereby replacing thousands of servers and ensuring that they continue to operate at high levels of utilisation for an extended period. Additionally, storage virtualisation techniques may be employed to enable many disk storage systems inside a network environment by pooling storage capacity. We can also help IT systems simplify storage foundation structure, manage information systems lifecycle, and maintain business continuity (Zhang *et al*., 2012).

Mass Distributed Storage: Cloud computing saves data by distributed storage, utilising redundant storage to maintain the integrity of stored data, and utilising highly credible software to compensate for the hardware's incredibility, resulting in affordable and sensible computing systems and mass-distributed storage. Cloud computing data storage technologies are Google File System (GFS), created by the Hadoop team, and Hadoop Distributed File System (HDFS).

1) Google File System: The GFS is a distensible distributed file system. It is utilised in big dispersed applications requiring safe access to bulk data to be processed and employed (Elomari *et al*., 2021). The GFS design concept is distinct from a traditional file system built to manage massive amounts of data and Google application property. It is based on affordable and widely available hardware. However, fault tolerance can be implemented. It can provide a high-performance service to a large number of users. Google File System clusters consist of a master server and numerous chunk servers that various clients may visit. The file is split into pieces of a specified size. When a block is created, the server distributes 64 handles to identify it as unaltered and globally unique. The block server processes block the same way Linux files are processed and saved to the local hard drive, reading and writing block data according to specified handles and byte ranges. Each block is replicated to many block servers to verify trustworthiness. The master server handles all the file system metadata, including namespace, file information, access control, and block position. Google File System has the benefit that it can manage multiple users efficiently and safely concurrently (Mailavaram and Rani, 2019). Each program includes GFS client codes that implement the GFS API, allowing the application to connect with and then read or write data from a master server and block the server. The interaction between client and server is limited to metadata activities; all data transmission is directly connected to the block server, which consumes and improves the system's efficiency by preventing the master server from overflowing.

2) Hadoop Distributed File System: The HDFS is a distributed file system for commodity hardware operating. It has been created in addition to saving enormous volumes of data and streaming it at high speed to satisfy the demands of consumers (Kaseb *et al*., 2019). The distributed file system is quite similar to GFS, but, at the same time, there is also a big distinction between the two data storage technologies. For example, HDFS is very tolerant of faults; it can copy data and produce several block copies (Kaseb *et al*., 2019). In addition, it can run on low-cost hardware; HDFS can give high-performance data access to large-scale data collections. The Master/Slave Architecture of HDFS is adopted (Mailavaram and Rani, 2019), and the HDFS cluster consists of nomadic and many data nodes. NameNode is the central server that controls the file system's namespace and client access to files. Generally, a node has a data node that manages the node's storage. A document is segmented into blocks from the inner portion stored on a set of data modes. NameNode executes namespace actions of the file system, including opening, closure, rename, or directory; it is also responsible for deciding mapping from data block to data node and for the deletion and replication of data in name nodes. The mode unitary streamlined the structure of the system consumedly. NameNode must preserve and manage all HDFS Metadata and read or write user data, not NameNode, on DataNode.

Parallel programming model: In this model, several computers or servers are linked together to carry out operations faster and more efficiently (Alrawais, 2021). Parallel programming is used in many applications, including Artificial Intelligence and the Internet of Things (Yin and Shi, 2021) and a simulation of complex phenomena (Eichstädt *et al*., 2020). Cloud computing programming paradigms must make task scheduling and concurrent delivery transparent to users and programmers to enable users to utilise cloud computing resources and enjoy cloud computing services more easily. Cloud computing utilises the MapReduce programming model, dividing work into several subtasks and reducing the time required to plan and distribute resources in the large node via two phases (Map and Reduce) (Joe *et al*., 2021). MapReduce is a fast and efficient (Pandey and Saini, 2020) mechanism for simultaneously executing many programs. Parallelism and fault tolerance are implemented in the database, and data distribution and load balancing. All data operations are summarised as follows: The Map and Reduce approach is used. If the programmer has used MapReduce to build its parallel processing approach, two functions must be defined: Map and Reduce. MapReduce may automatically initialise these to numerous instances of the same Map and Reduce tasks by performing the Map and Reduce functions on various data blocks based on the input data size and configuration information. A client, a master, and a worker make up the MapReduce architecture. For parallel processing, the client submits user assignments

to the master node and then automates them into Map and Reduce missions before distributing them to the worker nodes. Finally, the worker nodes contact the master node and request work, handled by the distributed file system of many worker nodes simultaneously (Zhang *et al*., 2012).

Data management: Enormous data sets must be efficiently handled by data management systems for cloud computing to store and analyse large volumes of distributed data. Google BigTable and Hadoop's HBase are the two types of data management technologies commonly found in cloud computing systems. Scheduler, Lock, and GFS are all used by BigTable, a MapReduce application. Two-dimensional sparse maps are used to create each table's rows and columns. The BigTable comprises four parts: a row, a column, a tablet, and a timestamp. Rows are found on a tablet. Tables in BigTable are dynamically assigned to tablets based on the dictionary keyword sequence. Each node manages around 100 tablets. A timestamp is appointed as an integer in 64 bits for each file version.

The family column is a collection of many columns, the granularity defining the authority level. BigTable requires a database connected to each client, a master server, and many tablet servers. The master server oversees the distribution of tablets to tablet servers, load balancing, and trash collection, among other functions. Tablet servers are used to manage a tablet set, handle read and write requests, and perform other operations (Zhang *et al*., 2012). HBase is a column-oriented database (Addakiri *et al*., 2020) and distributed data storage system (Xiao *et al*., 2017). It is open-source, developed on HDFS, and carries out queries quickly (Xu, 2021).

2.3.3 Uses of cloud computing

Cloud computing has boosted competitiveness by lowering costs, improving flexibility elasticity, and maximising resource usage (IBM Cloud Team, 2020). Moreover, it has scalability, efficient data storage, and easy management (Carvalho *et al*., 2021). Cloud computing can be used in several ways. Firstly, as a tool to share resources and data through the Internet while ensuring the integrity of the information (Altaee and Alanezi, 2021). Secondly, as a platform to process data from different sources and facilitate complex operations in a distributed way (Al-Marsy *et al*., 2021). Thirdly, as a tool to access information remotely efficiently and at a low cost (Leung *et al*., 2021). Cloud environments use encryption algorithms to increase security and confidentiality (Altaee and Alanezi, 2021).

Cloud computing applications have been used to store and manage data in the health care sector — it has been used, for example, to process data on patients with chronic diseases (Kishor *et al*.,

2021); to develop emotion recognition systems (Tian, 2021); to deploy blockchain applications (Tang and Zeng, 2020); to serve as a platform to schedule and plan processes (Abdalkafor *et al*., 2021); and to store and retrieve information from educational institutions (Mukred *et al*., 2021), among other applications.

Additionally, according to IBM Cloud Team (2020), cloud computing has several uses —

Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS): Infrastructures-as-a-Service (IaaS) provide users with on-demand, online, and pay-as-you-go computing and storage capabilities, and network. Utilising an existing Pay-per-Use Infrastructure seems to be a logical alternative for firms to save on their investment costs in IT infrastructure acquisition, management, and maintenance. Platform-as-a-Service offers a full platform for clients (hardware, software, and infrastructure) to develop, operate and manage applications without expense, complexity, and flexibility in establishing and maintaining this platform. For the same reasons, organisations can employ PaaS while looking for IaaS to enhance the development pace of a ready-to-use app platform (IBM Cloud Team, 2020). For example, digital twin apps operate cloud computing widely.

Hybrid cloud and multi-cloud: Hybrid cloud computing combines private and third-party public cloud services into one adaptable infrastructure to operate and operate its applications. As a result of this unique mix of public and private cloud resources, a corporation may select the most suitable cloud for any service or activity and can transfer workloads freely between the two clouds as needed. It is more cost-effective and efficient than utilising solely public or private clouds to achieve technical and business objectives (IBM Cloud Team, 2020) while ensuring appropriate levels of sustainability (Pańkowska *et al*., 2020).

Test and development: A test and development environment is an excellent candidate for cloud computing. Test and development involve planning and establishing an environment through physical assets, a big workforce, and time. Following that, the platform is installed and configured. All of this typically extends the time required to complete a project and causes milestones to be missed. Cloud computing enables access to settings that are tailored to specific needs. This generally includes but is not limited to automated provisioning of physical and virtual resources (IBM Cloud Team, 2020). For example, hundreds of educational apps were tested and built on cloud platforms to accommodate needs during Covid-19 lock-down situations (Qasem *et al*., 2019).

Big data analytics: One advantage of cloud computing is the ability of big data analytics to extract commercial value from massive amounts of structured and unstructured data (Li *et al*., 2021). Retailers and suppliers increasingly collect information on customers' purchase habits to focus advertising and marketing initiatives on a certain population sector. Social networking sites increasingly include behavioural analysis companies utilise to obtain useful information (IBM Cloud Team, 2020).

Cloud storage: Cloud storage allows users to save, access, and retrieve data from any web-enabled interface (Trinath Basu and Sastry, 2020). The interfaces of web services are generally basic and are always available, fast, scalable, and safe. Companies in this scenario pay for the cloud storage they use, with no regard for the ongoing costs of maintaining the storage infrastructure. Data can also be stored on-site or off-site based on regulatory compliance needs. Third-party hosts virtualise storage pools for the consumer's benefit (IBM Cloud Team, 2020).

Disaster recovery: Disaster recovery (DR) technologies that enable speedier recovery are also made more cost-effective with cloud computing (Cao *et al*., 2020). These DR technologies have no permanent assets, strict procedures, or a substantially higher price tag than a standard disaster recovery site (IBM Cloud Team, 2020).

Data backup: Data backup has always been a time-consuming and tedious procedure and necessitates the maintenance of a set of tapes or disks, their manual collection, and transport to a backup facility, which brings with it all the inherent issues connected with the source and backup locations. This method ensures that a backup is vulnerable to problems (such as running out of backup media). Before completing a restoration operation, backup devices must be loaded. This process might be prone to mistakes and human error. Since the beginning of cloud-based backup, many advancements have been made (Cao *et al*., 2020); data can automatically be sent across the cable, ensuring no safety, availability, or capacity problems (IBM Cloud Team, 2020).

While the list of cloud computing applications above is not exhaustive, it demonstrates the value of cloud computing compared to conventional choices for increasing IT infrastructure flexibility and implementing big data analytics and mobile computing (IBM Cloud Team, 2020).

2.3.4 Cloud computing: Environment architecture

Cloud computing architecture impacts risk assessment processes for big data based on cloud computing technologies. For example, cloud computing affects how software is built, servers are

used, and computational tasks are carried out and implemented (Poniszewska-Maranda *et al*., 2019). It has also played an important role in how industries (Albelaihi and Khan, 2020); educational institutions (Backialakshmi and Sumalatha, 2020); health institutions (Cao *et al*., 2020); and governments (Shafiu *et al*., 2016) migrate to the digital age. In the cloud computing model, information is stored, processed, and retrieved in a distributed and secure way (Bermani *et al*., 2021). As defined by the US National Institute of Standards and Technology (NIST), cloud computing is a network access approach providing simple, on-demand network access to a shared pool of configurable computer resources that can be rapidly provided and released with low administrative effort and/or service provider (NIST). There is a reduction in the total cost of cloud resource sharing (Sahu and Pateriya, 2013). The European Community for Software and Software Services (ECSS) characterises cloud computing as providing computational resources from a place other than the existing one (Drissi *et al*., 2013). The NIST states that the cloud computing system includes five key features, three cloud service models, and four cloud deployment models.

Layer	Cloud computing Components
Five Characteristics	On-demand self-service Broad network Access Resource Pooling Rapid elasticity Measured Services
Three Delivery Model	IAAS PAAS SAAS
Four Deployment Model	Public Private Community Hybrid

Figure 2.4: Cloud environment architecture

Source: Adopted from Potey *et al*. (2013)

2.3.5 Characteristics of cloud computing

Understanding the characteristics of cloud computing will help identify areas for improvements in managing risk assessment processes for big data based on cloud computing technologies. Cloud

computing is on-demand, virtual, elastic, cost-effective, locational, device-independent, and all-time. The cost of third-party and worldwide resource usage is considerably lower, preventing resource waste and computing power. Cloud computing's primary goal is to maximise distributed resources to improve performance and handle large-scale computing issues. Cloud computing has several functions —

Optimisation: Cloud computing can serve as a platform to optimise operations assignments (Sundararaj, 2019) and workflows (Saeedi *et al*., 2020).

Support: It can be used to implement various applications in several domains and through several technologies such as IoT (Yassine *et al*., 2019).

Service improvement: Due to its high levels of flexibility and low cost, it can increase the quality of its deployed services. Cloud computing is improving, for example, the efficiency of information technology healthcare services (Darwish *et al*., 2019).

Integration: It is possible to integrate different devices and applications (Zamora-Izquierdo *et al*., 2019), facilitating connectivity and operationalizing tasks in many fields like precision agriculture and smart farming.

Computational capability: Cloud computing can increase the computational capabilities of firms and institutions in a fast, scalable, and efficient way (Zheng *et al*., 2019).

Moreover, according to Liu *et al*. (2013), several of the fundamental aspects of cloud computing include the following —

On-demand Self-Service Computing: Resources are delivered online at a specified moment without human involvement according to customer requirements. Users just utilise a browser to access the data, apps, or other services in the cloud regardless of other software and hardware (Qasem *et al*., 2019).

Broad Network Access: Cloud users have a wide range of cloud services available on the Internet (Ahmed *et al*., 2019). There is no reliance on a customer platform for cloud services. Services are always on, wherever, and whenever.

Resource Pooling: Cloud computing resources are pooled to meet customer demand for numerous consumers (Shen *et al*., 2020).

Measured Service: Cloud users do not need to control and optimise computer resources because the cloud system automatically controls most. The use of resources can be monitored, regulated, and reported to make the service's supplier and customer more transparent.

Selection of Provider: The cloud service provider selection is the key to achieving good service. Users can pick the proper service provider according to their preference and expertise of cloud providers. It is important that the supplier is dependable and well-known for its customer service and has a demonstrated track record of performance in IT businesses.

2.3.6 Cloud delivery models

For Sahu and Pateriya (2013) and Sohaib *et al*. (2019), there are three cloud delivery models: Platform as a Service (PaaS), software as a Service (SaaS), and Infrastructure as a Service (IaaS). Moreover, there is an alternative model, data as a service. These models can be deployed in public, community, private and hybrid clouds (Sahu and Pateriya, 2013).

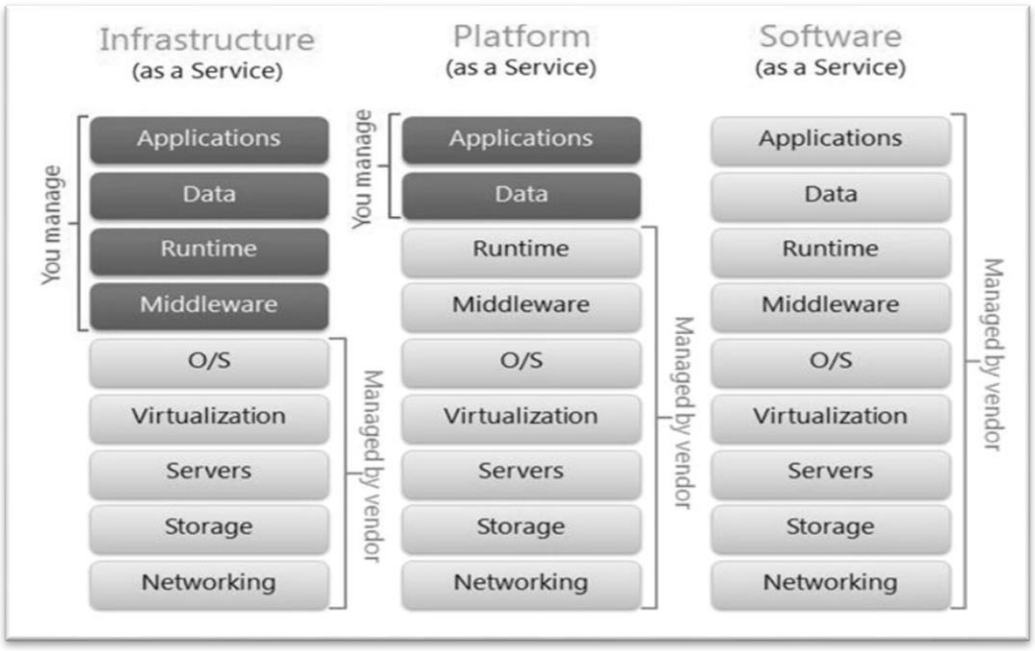

Figure 2.5: Cloud computing models

Source: Adopted from Sahu and Pateriya (2013)

Software as a Service (SaaS): Cloud computing enables users to access a SaaS service without handling any gear or software installation and configuration. In this mode, the provider offers applications that run in the cloud and can provide software already developed by a third party (Sohaib *et al*., 2019). A vendor or cloud provider manages all the installation and configuration of

services. Examples include Google, MS Office, and Google Docs. The major advantages include lower upfront costs, lower lifetime costs, eliminating licensing risks, eliminating version compatibility, and reducing the hardware footprint. Software as a Service's primary drawback is billing administration and customer and vendor migration synchronisation.

Platform As A Service (PaaS) integrates hardware and software to develop, improve, and implement algorithmic applications (Sohaib *et al*., 2019). It configures the Web delivery of a computer platform, where the user may build and install its application. A vendor or cloud provider handles the computer platform and server setup. Web applications may be developed fast without the complication of purchasing and operating the storage server, database, and other software/hardware. The Google App Engine is an example of PaaS. Platform as a Service allows developers to focus on code development, global platform access, hardware dependency elimination, and scalability. Some drawbacks include tight control to avoid the construction of apps without IT engagement by business lines.

Infrastructure as a Service (IaaS): The cloud provider provides the infrastructure, including servers, software, and network equipment, as a pay-as-you-go service (Sohaib *et al*., 2019). Infrastructure as a Service is generally used to work on large-scale and complex computational tasks and problems (Abdullahi *et al*., 2019). It may be utilised to eliminate the costs and management associated with purchasing, housing, and administering the necessary hardware and software infrastructure. It enables rapid scaling up and down to meet demand.

Data as a Service (DaaS) is the alternate cloud computing model. It varies from existing models like (SaaS, IaaS, and PaaS) by providing data to network users. DaaS is tightly linked to large data; both employ comparable technologies (Rjoub *et al*., 2020). Data as a Service offers very efficient data distribution and processing technologies. Data as a Service, together with SaaS, is strongly linked to storage as a service and software as a service. Furthermore, DaaS typically relies on cloud storage to keep the company that owns information and the supplier it works with ongoing data access. It is thought to be hosted on the cloud (Elhoseny *et al*., 2018).

2.3.7 Cloud deployment models

According to Sabahi (2011), there are four cloud deployment models.

Public cloud: Users can access the public cloud freely using Web browser interfaces. The user is only charged for the length of the service (pay-per-use). Consider the electricity we receive in our

houses as a contrast. We pay for the resources we consume. The same strategy applies here and may aid in lowering operating costs associated with IT expenditures.

On the other hand, public clouds are less secure than other cloud architectures. All apps and data stored in public clouds are more vulnerable to malicious cyberattacks; security measures must avoid such attacks (Sabahi, 2011). The public cloud deployment model tends to be cheaper than the private model (Andreadis *et al*., 2015) and offers elastic computing resource availability (Azumah *et al*., 2021).

Private cloud: A private cloud operation occurs within the internal data centre of a business. The major advantage here is that safety, maintenance, and upgrades are easy to monitor (Reena and Shajin Nargunam, 2019). This model fully controls the information and implements applications (Azumah *et al*., 2021). A private cloud may be likened to the intranet and gives greater control over the implementation and use of cloud applications. All private cloud resources and apps are pooled and made available to users on an organisational basis. The organisation controls the resources and applications itself. This deployment strategy improves security because only companies can access the private cloud (Sabahi, 2011).

Hybrid cloud: This model offers total control of information and programs and greater computing capabilities (Agarwal *et al*., 2020). This concept connects a private cloud to one or more external cloud services. It is safer to manage and access data and apps through the Internet. It enables the company to satisfy its demands in the private cloud, and in infrequent needs, it requests intensive computer resources from the public cloud (Sabahi, 2011). The user can leverage the advantages or disadvantages of the private and public model via a hybrid cloud depending on the particular needs of the process implemented (Azumah *et al*., 2018).

Community cloud: A community cloud is configured when many firms collaborate to build and share a cloud infrastructure and associated needs and regulations (Elzamly *et al*., 2019). Usually, the users of this model have similar interests, goals, and concerns (Baig *et al*., 2015). The cloud infrastructure might be hosted by a third-party supplier or an entity inside the community (Sahu and Pateriya, 2013).

2.3.8 Challenges in big data and cloud computing

Understanding challenges in big data and cloud computing is an important milestone toward finding the solution to managing risk assessment processes for big data based on cloud computing

technologies. Big data and cloud computing applications pose many problems despite their growing benefits.

Increased energy use: Carrying out large-scale operations in big data and cloud environments has received criticism for consuming large amounts of energy (Guo *et al*., 2019). The server farms where these applications operate use large amounts of energy resources.

Security: Cloud computing implies sharing data, which involves many risks (Shen *et al*., 2019). These risks are leakage, data distortion, cyber-attacks, and malicious threats.

Quality: One of the most pressing challenges of the providers of cloud and big data platforms and tools is ensuring the quality of the services offered (Ma *et al*., 2019) because customers' requirements and operations are broad, and the set of data generated is diverse and sometimes complex to manage.

Additionally, Zanoon *et al*. (2016) describe several challenges that emerge when big data and cloud computing intersect.

Data Storage: Large data storage through conventional storage is a challenge since the hard drives often fail, data protection measures are not sufficient, and the pace of big data demands efficient storage systems to increase quickly, which with conventional storage systems is difficult to do (Lin *et al*., 2019). Cloud storage services offer nearly infinite capacity and a lot of fault tolerance to solve huge data storage issues.

Variety of data: big data grows, rises, and fluctuates organically due to the growing number of data sources. This increase contributes to the heterogeneity of big data.

Data transfer: The data is collected, entered, processed, and produced through many steps. Big data transfers are a difficulty. Data compression techniques must minimise volume when data volumes impede transfer speeds. It also impacts prices, whereas cloud computing enables shared storage and data transport over high-speed lines, lowering expenses via virtual resources and on-demand use.

Privacy and data ownership: The cloud environment is open, and the user's function is restricted in monitoring. Privacy and security are key problems with large data (Xiong *et al*., 2019). Big data and cloud computing are frequently used to tackle these privacy problems in practice.

Big data and cloud computing are emerging technologies that provide new security risks that must be assessed and addressed. As a result, safety risk assessment is critical; the traditional technical risk assessment method centred on assets should create new business opportunities that focus on the unique characteristics of cloud computing and technological advancements that have resulted in a new way for cloud providers to provide their services to cloud consumers. Cloud computing users should be aware of the threats and vulnerabilities of the cloud computing environment. Risk assessment for information security is a critical component of this process (Drissi *et al*., 2013).

2.4 Summary

Cloud computing and big data are closely interrelated. The literature reports that cloud computing environments play a key role in storage and management due to the huge amounts of data generated. To literature reports, cloud computing can be useful for testing and development, analytics, storage, disaster recovery, and backup. Literature showed that understanding the characteristics of could computing and big data is an essential step toward finding solutions for better management of risk assessment processes for big data based on cloud computing technologies.

CHAPTER 3: RISK ASSESSMENT IN BIG DATA AND CLOUD COMPUTING

3.1 Introduction

This chapter presents risk assessment, analytics, and evaluation in large data environments and cloud systems. The risk assessment process (threat identification, risk identification, and control recommendation), qualitative and quantitative risk analyses, and risk assessment methodologies are described in this Chapter (needs statement and identification of security objectives; evaluation of operationally critical threats, assets, and vulnerabilities; and harmonised risk analysis). This Chapter also highlights some major threats to cloud computing that should be controlled properly (technology, data security and regulatory, vendor, operational and financial).

3.2 Risk assessment and risk management in the big data era

Virtual work structures are taking over conventional patterns, and big data's role is becoming increasingly imperative daily. With increased dependency and widespread utility of big data based on cloud services, risk assessments become important for organizations and service providers to continue the faith people bestow in the cloud service providers. Therefore, organisations and cloud service providers are adopting increased RA methods to manage and access the risks and potential threats. Assessment and management of risks are critical Risk assessment processes for big data based on cloud computing technologies in Jordan, Canada, South Africa, and the United Kingdom.

According to Drissi et al. (2013), the risk is not inherently negative; taking risks is necessary for making progress, and experiencing failure is frequently an integral factor in learning. Nevertheless, we need to acquire the skill of learning to strike a balance between the potential drawbacks of risk and the various potential advantages of the chance it presents. Drissi et al. (2013) posit that taking a risk is not inherently negative; rather, it is necessary for growth, and failure is frequently a crucial component of learning. But we must learn to weigh the potential drawbacks of risk against the advantages of the possibility it presents. A "risk" is the possibility that a dangerous agent is vulnerable (Potey et al., 2013). The standard ISO 27005 states that information security risk may benefit from hardware and software flaws, negatively affecting the institution and its processes (ISO, 2018). The risk associated with incidences can be understood as follows:

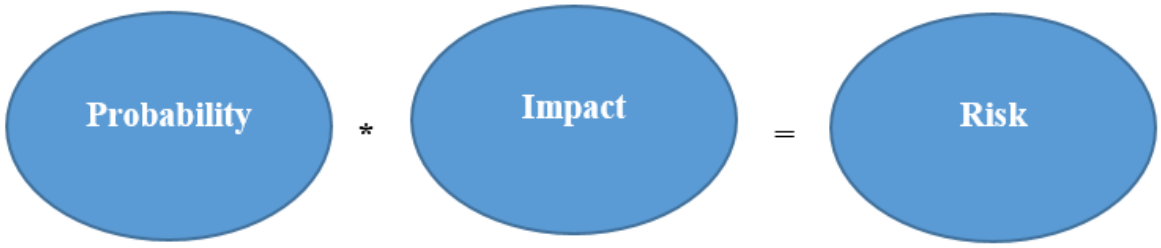

Figure 3.1: Formula for risk associated with an incidence

Source: Adapted from Malathi et al. (2019)

Risk management is one of the key processes in information systems in general and in cloud computing and big data applications in particular; ignoring risks can have disastrous consequences at the infrastructure, design, and process level (Turskis *et al*., 2019) in safety-critical industries (Paltrinieri *et al*., 2019). Risk management is a coordinated collection of activities and strategies used to guide and control the various risks that can jeopardise an organisation's ability to accomplish its objectives. Additionally, risk management refers to the architecture utilised to manage risk (ISO 31000, 2009). In cloud-based operations, the data is highly vulnerable; therefore, protecting the data and information is crucial against possible threats from online frauds, data leakages, misinterpretations, and other unintentional misuses. Risk management activities and techniques closely monitor and allow only the right information to be flown in the system at right time. These services closely monitor the suspicious activities in real-time and curb the present and future losses. Such evaluations are the need of present time to correspond the present day rising demand of virtual operations

Risk evaluation is one phase in risk management. The approach is to discover a system's safety concerns and identify its probability, impact, and protections that lessen that impact. For Drissi *et al*. (2013), the major purpose of the risk assessment is to identify effective controls to reduce or eliminate these hazards. Risk includes three elements: what can go wrong (scenario); what chance it will have (possibility of occurrence); and implementing strategies that can handle this risk (Paltrinieri *et al*., 2019). Liu *et al*. (2013) interpret risk assessment as identifying threats and vulnerabilities in a corporation's information resources and deciding what countermeasures the business should take to mitigate these risks to an acceptable degree in light of the organisation's information resources' worth. Despite all the negative consequences of materialisation risks, they can be minimised to an adequate level by implementing proper assessment and mitigation techniques (Qin *et al*., 2020).

Cloud clients can use the risk assessment approach to evaluate the risks of selecting a cloud service provider. It analyses historical data from cloud clients and service providers to assess risk. This simplifies the process of finding a cloud service provider with a low-risk profile based on security, privacy, and service delivery threats (Cayirci *et al*., 2016).

3.3 Risk assessment processes

The security of information is a dynamic and complicated task. There are various comprehensive ways, but none can be considered fully accurate and applicable to all big data sets. Hence, numerous risk assessment processes operate on multi-variate cloud patterns and element model correlation. As previously discussed, risk assessment is a key phase in risk management that identifies appropriate control techniques to reduce or eliminate hazards. According to Turskis *et al*. (2019), the most common risk assessment tools and techniques used at the industry level are the Failure Modes and Effects Analysis (FMEA) (Baynal *et al*., 2018); Bow-Tie Analysis (Muniz *et al*., 2017); Fault Tree Analysis (FTA) (Giraud and Galy, 2018); Layer of Protection Analysis (LOPA) (Yan and Xu, 2018); and Hazards and Operability Studies (HAZOP) (Taylor, 2017). For to Dioubate *et al*. (2015), in the risk assessment process, there are four steps:

1) *Identifying potential threats:* This initial phase identifies all potential system threats. It recognises potential threats to the system.
2) *Identifying vulnerabilities:* The second stage aims to provide a list of system vulnerabilities (faults or flaws) that potential threat sources can exploit.
3) *Risk assessment:* The third phase determines the system's risk level.
4) *Suggestion for control:* The fourth stage's objective is to recommend rules consistent with the organisation's operations and to mitigate or eliminate identified risks. The proposed regulations are intended to reduce system risk.

For Dioubate *et al*. (2015), the risk assessment method for information security is divided into six steps:

1) *Determine the aim of the assessment:* This step defines the information system's data, hardware, and software.
2) *Evaluation performance:* To improve the evaluation plan, develop evaluation processes and pick appropriate evaluation methods and instruments.
3) *Risk identification:* The evaluation scope identifies the assets, danger, vulnerability, and existing security measures.

4) *Risk analysis:* The potential and effects of threats and vulnerabilities should be analysed.
5) *Assessment:* Assess the results to generate an expert's risk assessment report.
6) *Risk control:* consists of applying appropriate risk transfer procedures and avoiding or reducing the system risk.

3.4 Risk analysis methods:

Risk analysis is an intricate and multi-step process that involves an in-depth analysis of the exchange of information within the system. It aims to create a proactive control whose main purpose is to mitigate the risk associated with cloud-based big data. Drissi *et al*. (2013) believe that methodologies for risk analysis are often separated into quantitative and qualitative studies. Both methods test the risk probability and aim to highlight the chances of data and information-based risks and threats (Jingrui, 2017). Both methods are fundamentally aimed at identifying, analysing, and gauging the external risk factors affecting the confidentiality and integrity of the data.

Quantitative risk assessment techniques: Quantitative risk assessment methods are the objective methods that use verifiable data and analyses the risk matrix based on a mechanistic perspective (Ahmad et al., 2022). Although many well-developed firms employ quantitative risk management, it is frequently not used in information technology. Quantitative risk assessment is a statistical technique for coping with unforeseeable situations (Alali *et al*., 2018). Various risk assessors considered this model to be based on simple approximations that may not be highly dependable in the case of complex data structures.

Qualitative risk assessments: This strategy is utilised when a numerical assessment of risk is difficult to communicate. It is performed by risk managers and assessors who address the qualitative information and make inferences based on personal judgment based on past experiences and competencies (Raja and Hanifa, 2017). It discusses in full the likelihood of outcomes. It relies on experts' judgment rather than statistical information (Alali *et al*., 2018).

Semi-quantitative risk assessment techniques: certain techniques work on the amalgamation of qualitative and quantitative techniques. These systematic procedures have minimum risk factors as they are based on the elements of both processes (Jain and Mahajan, 2017). For Alali *et al*. (2018), a semi-quantitative method uses numerical data supported by qualitative information to estimate the risks better. Potey *et al*. (2013), Liu *et al*. (2013), and Drissi *et al*. (2013) Compiled some of the most relevant approaches for risk assessment accessible today.

- Expression of needs and identification of security objectives (EBIOS) is a risk assessment and management method to determine safety measures for implementing and safety expressions.
- An operationally critical threat, asset, and vulnerability assessment (OCTAVE) is a technique for assessing vulnerabilities and threats against a business's functioning assets.
- Method of harmonised risk analysis (MEHARI) is an information system security risk assessment method designed to match the demands of each firm.

Thus, there are various methods for risk analysis, and their implementation depends upon the nature of operations in the organisation. If companies have relevant, structured and reliable data, the likelihood of risk occurrence can be examined using any of these methods. In different countries, these methods are implemented with different intensities (Mastorakis, 2018). Western nations focus on risk assessment methods as they consider audit and assessment opinions crucial. However, in some developing nations like Asian countries, South Africa and Jordan, the risk assessment methods are still inconsistent with the business operations (Bai, 2014). Thus, there is an intrinsic need to have centralized or integrated risk assessment standards that could exemplify the division of categories based on which risk assessments can be done in developing nations worldwide.

3.5 Risk assessment in big data and cloud computing environments

The Risk Assessment approach is designed to assess risks by picking a specific cloud service provider for cloud and big data customers. It is an expert system that assesses and analyses several risk scenarios and evaluates background information from cloud consumers, cloud service providers, and other external public sources. For Cayirci *et al*. (2016), this would enable the implementation of large data operations by cloud customers in making educated decisions to choose a preferential risk profile cloud service provider. Four major management techniques exist risk acceptability, risk mitigation, risk reduction, and risk transfer (Turskis *et al*., 2019).

The risk inventory involves two main elements:

- *Assets:* Virtual Machine (VM), physical host, and SLA, describe their features. Risk events are evaluated in this regard.

- *Incidents/risk scenarios:* Any occurrence, circumstance, or combination that may impair the capacity or availability of an asset should be described.

Incidents consist of the following:

- *Vulnerabilities:* Report the vulnerabilities inherent in the asset (e.g., malfunctioning hardware); their impact indicates the risk for a hazardous incident (e.g., the asset-inherent service level agreement (SLA) and breaches of quality of service (QoS) indicators.

- *Threats:* Represent the risk on the other side based on asset-specific considerations (e.g., loss of connectivity of a physical host).

- *Capacity for adaptation:* Description of mitigation strategies for the particular asset (e.g., server replication).

- *The impact/consequences of a risk incident:* Refers to the effects of a risk occurrence on the assets (i.e., the breakdown of a physical host), with impaired performance, data loss, or unavailability. The evaluation is based on the indicators chosen for the asset and related costs (e.g., cost of not meeting predefined service levels).

The EU Cybersecurity Agency (ENISA) recommendations are particular to cloud computing, and there is no generic framework for mapping the peculiarities of cloud service consumers (CSCs) and service providers (CSPs) into the 35 risk scenarios. Since European nations are proactively diligent in their risk assessment concerning big data on cloud services, the European integrated list of cyber security issues and malpractices has been included to provide an outlook for the rising catastrophic incidences against the virtual world developed robustly around the communities globally.

Table 3.1: List of risk scenarios and categories of ENISA (The European Union Agency for Cybersecurity, 2009)

Risk Category	**Risk's full name**	
Policy and Organisational	P1	Lock-in
	P2	Loss of governance
	P3	Compliance challenge
	P4	Loss of business reputation due to co-tender activities

Risk Category	Risk's full name	
	P5	Cloud service termination or failure
	P6	Cloud providers acquisition
	P7	Supply chain failure
Technical	T1	Resource exhaustion (under or over-provisioning)
	T2	Isolation failure
	T3	Cloud provider malicious insider abuse of high privilege role
	T4	Management interface compromise (manipulation, infrastructure availability)
	T5	Interpreting data transit
	T6	Data leakage on up/download, intra-cloud
	T7	Insecure or ineffective deletion of data
	T8	Distributed denial of service (DDoS)
	T9	Economic denial of service (EDoS)
	T10	Loss of encryption keys
	T11	Undertaking malicious probes or scans
	T12	Compromise service engine
	T13	Conflicts between customer hardening procedure and cloud environment
Legal	L1	Subpoena and e-discovery
	L2	The risk from changes in jurisdiction

Risk Category	Risk's full name	
	L3	Data protection risk
	L4	Licensing risk
Not specific to the cloud	N1	Network breaks
	N2	Network management (i.e., network congestion/misconnection/non-optional use)
	N3	Modifying network traffic
	N4	Privilege escalation
	N5	Social engineering attacks (i.e., impersonation)
	N6	Loss or compromise of operational logs
	N7	Loss or compromise of security logs (manipulation of forensic investigation)
	N8	Backup lost, stolen
	N9	Unauthorised access to premises (including physical access to machines and other facilities)
	N10	Theft of computer equipment
	N11	Natural disasters

The table above highlights that risk assessment processes for big data based on cloud computing technologies are defined into policy and organisational, technical and legal risks frameworks. The various risk scenarios mentioned in the table depict the natural and deliberately created malfunctioning by malevolent players that affect the information security and data quality. These

probable risks are the most commonly found threats to big data in cloud computing that needs special focus on risk assessment using appropriate methods in the technological market.

Elahi *et al*. (2021) propose a semi-quantitative risk assessment method with eight steps for cloud computing scenarios: 1) characterise the system to be evaluated; 2) identify vectors of attack or threat circumstances and their potential gravity at the level of privacy, security, and trust; 3) classify the attack vectors and thread circumstances by several categories (i.e., reputation, information disclosure, denial of service, the elevation of privileges); 4) the impact of threats at three levels is assessed at low (limited), medium (severe), and high levels (powerful implications); 5) establish the probability of dangers; 6) calculate the serious threat risk by analysing the likelihood of occurrence together with its impact; 7) identify and effectively identify security controls; and 8) systematically analyse the net risk considering the initial five processes.

3.6 The risk associated with cloud computing:

Cloud computing greatly benefits organisations and facilitates easy and integrated operations from different locations. However, along with various advantages and scalability, numerous risks come with it, as data is available and vulnerable in the cloud domain. Various issues and risks include compromised service standards and embedded risks associated with internet networks.

For Swathy Akshaya and Padmavathi (2018), nine cloud security threats are based on the characteristics and vulnerabilities of attacks; as cloud applications become more prevalent, these assaults become more frequent (Swathy Akshaya and Padmavathi, 2018):

- Infringement of data
- Data loss of critical nature
- Obtaining account or service-related traffic
- Interfaces and APIs that are insecure
- Denial of service
- Malicious insiders
- Cloud service abuse
- Inadequate diligence
- Vulnerabilities in shared technology

According to Maniah *et al*. (2019), cloud computing has three security risks: resource use deviations, data misuse, and crimes against data. By using another classification, Gadia (2018)

posits five different types of risks present in cloud environments: data security and regulatory risk, operational risk, technology risk, financial risk, and vendor risk.

Figure 3.2: Main types of risks present in cloud environments

Source: Adopted from Gadia (2018)

3.7 Main Types of Risk:

The risk assessment techniques applied to big data using the cloud are an integral part of organisational information system as it helps the firm mitigate the risks and threats arising from virtual working. It is necessary to understand and fragment the common risks and threats that can hamper the data available in the cloud. Data security and regulatory risk could be linked to loss, leakage, or unavailability of data. This might result in company interruption, revenue, loss of reputation, or compliance with regulations. Regulatory risk is linked to failure to comply with different legal and regulatory standards, such as the Gramm-Leach-Bliley Act (GLBA), Sarbanes-Oxley Act (SOX), Health insurance portability, and accounting law (HIPAA), or the Data Protection Directive of the European Union (EU). Data risk does not alter much for a private cloud as traditional computing. Organisations have a greater grasp and control over applying numerous government rules, legislation, and regulations. Furthermore, data is not mixed between multiple cloud users. However, the private external cloud has added dangers:

– Visibility of commencement, recording, authorization, processing, or reporting checks.

Unauthorised data access by a service provider and/or a lack of control over which contractors or third parties view the service provider may use data.

The data dangers connected with the private external cloud are applicable in a public cloud. The following hazards also apply:

– Leakage of data or access risk due to multi-tenancy/shared infrastructure.

– Flexibility of data protection measures, such as encryption and applying controls on specific data types. Different enterprises may have distinct encryption and control requirements, and a public cloud provider may be incapable of designing or managing their infrastructure. This is especially true for solutions delivered via SaaS and PaaS models.

Technology risk (Gadia, 2018) may be linked to continuously evolving technology and a lack of standardisation in integration or interoperability. Technological hazards could lead to costly re-architectural efforts to integrate or implement new technologies.

A technological risk could arise for a private cloud:

– A continually changing technological landscape could compel the firm to modernise its computer resources or to educate its support personnel.

– Human mistake possibilities are due to the number of customisable locations and the deployment frequency.

For a public cloud, the developing technological risk may be reduced, as service providers will be responsible for upgrades/rearchitects and retraining of users. However, more technological problems could be included in a public cloud, including the following:

– The organisation may need to redesign its cloud apps more frequently than established technologies because technology characteristics are continually evolving. Multiple instances of a company's cloud architecture being created before deploying advanced security and control services have been observed. Management did not update the cloud architecture to enable management to use the supplier's extra functionalities for free or at a minimal cost. For example, in 2015, AWS implemented hundreds of improvements to its cloud services.

How much of an organisation's infrastructure, platform, or apps can be customised depending on the type of service used?

– CSP dashboards may provide limited visibility due to several cloud providers and hundreds or thousands of cloud-based server instances. Typically, each public cloud solution provider gives its management console. Software as a Service solution in companies is rapidly developing. This becomes increasingly challenging because of the lack of a single/unified management dashboard.

Operating risk (Gadia, 2018) can be linked to the performance of IT services and business tasks. Migration to the cloud has also brought a new method dubbed DevOps, where the responsibility for development and operations merges. This reduces deployment timelines to days instead of weeks or months. Moreover, IT operations can influence, and development teams will require cloud deployment and system administration training, although IT hardware and network administration will be less critical.

Some of the operational hazards for a private cloud are:

– Suboptimal service reliability and uptime because a company could not use leading cloud computing technologies may deliver higher service reliability and uptime.

Some of the operational hazards for a public cloud are:

– Customised service levels for various information technology services that the company may require, including those connected to availability and disaster recovery, to choose an acceptable level of service.

– Less control over service quality.

– Lower control on the availability of important applications and disaster recovery.

The vendor risk (Gadia, 2018) is derived through leverage or seller relationship. Unexpected selling conditions, such as bankruptcy, lawsuits, an SEC investigation, or other defamation against the seller, might significantly harm the company's reputation and goodwill. This danger applies to private external and public cloud computing, as it is related to and dependent on the service provider. As a result of simple access to IaaS, innovative SaaS start-ups have sprouted—some with unique solutions that fill gaps left by traditional suppliers. Some of these suppliers may not be viable for large companies that want to transmit expanding data while meeting strict control standards. By utilising third parties for cloud computing, service providers rely on them to ensure adequate controls and compliance with applicable laws, rules, and regulations.

– According to Skyhigh Networks' Cloud Adoption Risk Report for Q4 2015, the company now has access to over 16,000 cloud services and an average of 1,154 cloud services.

– According to the CIO blog of the Wall Street Journal, "Cisco Systems Inc.'s Vice President and Chief Information Security Officer Steve Martino discovered 607 cloud services when the company began to use a security broker. About half of those were cloud services that Cisco already had a relationship with, so it meant investigating a few hundred other services, said Mr Martino, speaking at the RSA conference."

–Many firms are battling with Shadow IT, which increases their organisation's risk by using non-risk cloud public solutions. According to Skyhigh Networks, less than one in ten providers maintain encrypted data and even less enable a customer's capacity to encrypt data with their encryption keys.

Overspending and earnings loss might be related to financial risk. According to the December 2015 North Bridge and Wikibon's Future Cloud Computing Survey, cloud service costs are three times more today than five years ago.

Financial concerns for a private cloud are:

– The initial costs of establishing a private cloud are grossly overestimated.

– Maintain capital expenditures on hardware and software.

Financial hazards associated with public clouds are mostly due to the changeable nature of costs, which raise the cost of using the public cloud due to insufficient business planning and demand. Consider electricity as a metaphor. During the winter, a consumer can use an electric heater all day without realising the expense until the end of the month, when the electricity bill arrives. Cloud expense management necessitates a level of concentration, expertise, and technology that was not previously required.

Alali et al. (2018) state that additional risks can impact cloud applications. One of them is criminal activities carried out through new technologies, mainly focusing on information and digital systems. The other is fraud and stealing illegal images or information; this risk is materialised via digital platforms. These risks, if set, can significantly impact firms' processes in the service and manufacturing sectors; hence, they are top security priorities globally (Kshetri, 2015). Cryptographic attacks are one of the worrying elements of threats popular nowadays wherein a

malevolent intermediary player places himself between two users and intercepts the conversation (Trimintzios and Gavrila, 2013). These types of privacy issues risk the entire system of organisations, the service providers cannot provide extensive security and risk assessments, and any outage results in great losses for organisations.

Several elements must be properly managed to control the risks described above: confidentiality, integrity, availability, accountability, and privacy (Tchernykh *et al*., 2016). Confidentiality involves the controls that restrict access to information and systems. Integrity considers the processes and structures implemented to prevent data and systems from being manipulated without permission from the owners and administrators. Availability involves access to the required information when needed. Accountability implies the strategies to protect the data observed and facilitate traceability. Privacy consists of the freedom to grant (or not) access to private information.

3.8 Associated publications on risk assessment in big data and the cloud

The most recent work on risk assessment and cloud computing is very diverse. Some works focus on risk assessment methods to mitigate threats (Mackita *et al*., 2019). Other authors focus on risk management methods in e-commerce contexts (Li *et al*., 2019), risk management methods in resource provisioning (Halabi and Bellaiche, 2019), and risk management in logistic processes based on the cloud (Maniah and Milwandhari, 2020). The second area integrates research to identify threats and risks in cloud environments (Maniah *et al*., 2019). This line of research is based on surveys attempting to assess the most important risks firms face. An important work carried out by Jones *et al*. (2019) explores the risk factors associated with implementing cloud applications in the cloud; the authors found that loss of control and lack of data ownership can be an important risks that firms can face when implementing cloud applications. Although these works provide important insights on risk assessment in cloud environments, they mostly focus on quantitative methods and do not focus on several countries, which is a restriction that this research tries to overcome.

Regarding big data and risk assessment, the most relevant works focus on big data risk in supply chain management. For example, Singh and Singh (2019) explore how companies develop risk resilience via big data analytics during events that impact their supply chain. Despite the relevance of this research, it is based on surveys, which may prevent in-depth explorations of the issues analysed. Another work carried out by Araz *et al*. (2020) evaluates recent research on operational

risk management in a big data context. However, this study is based on a literature review. Other research focuses on big data and risk assessment in the context associated with educational administration (Sorensen, 2018), politics (Zhang *et al*., 2019), and health care (Richter and Khoshgoftaar, 2019) but do not integrate the cloud computing aspect and neither use interviews across countries, which is a major differentiation and contribution this research intends to make.

Despite the risks mentioned, the cloud cannot be placed out of businesses nowadays. It provides firms with reduced capital costs, minimum material cost, modernization of legacy IT, fast and remote application, and development of new capabilities in companies (Yang et al., 2016). Forsaking the cloud due to privacy and confidentiality issues is not thinkable after its many exponential benefits. It is an integral part of the organisational function nowadays and is inseparable (Manogaran et al., 2016). Therefore, organisations across the globe are looking for opportunities through which risks can be minimized. Before implementing cloud in businesses, organisations must implement strategic and critical functions to understand the operations' sensitivity and curb future risk factors.

Further, before cloud adoption, organisations must have a realistic assessment of risk factors through information system audits and other techniques that could bring sustainable benefits (Raja and Hanifa, 2017). In the UK, the USA, and western nations, standards and best practices for cloud services are high, protecting the firms from potential destruction (Ahmad et al., 2022). Nevertheless, in developing African and Asian nations, a comprehensive methodology to support risk mitigation needs to be developed and penetrated the community to avoid risks associated with this environment

3.9 Summary

The literature shows that risk assessment in cloud computing and big data environments is increasingly relevant; privacy and data leakage risks must be carefully identified, assessed, and managed to avoid catastrophic consequences for companies and their customers. To properly and safely implement the different cloud and big data applications, it is necessary to identify potential threats, vulnerabilities, and associated risks and put certain practices to mitigate the risks, considering its goals and operations.

CHAPTER 4: RESEARCH METHODOLOGY AND DESIGN

4.1 Introduction

This chapter outlines the research methodology used in this study to develop a comparative analysis of risk assessment processes for big data based on cloud computing technologies. This chapter describes the paradigm selected to frame this study. It represents the various research methodologies and reasoning processes implemented in big data using cloud services and Information Systems (IS) based platforms and delves into this study's use. The rationale for the adoption of the methodology is also presented. This study explains various research designs, selecting the most appropriate design for the study. The data collection methods, target population, sampling technique, size, and data analysis strategy are illustrated. Finally, this chapter shows quality and rigour conditions in data reliability and validity and the ethical considerations to carrying out this research.

4.2 Research paradigms

A paradigm can be defined as a "pattern of perspectives that can be used and binds the work of a group of theorists together" (Burrell and Morgan, 1981: 23). Kuhn (1962) defines a paradigm as a collection of widely recognised scientific achievements that have long served as models for a community of practitioners (horizontal perspective). Rather than discussing the evolution of fields, Burrell and Morgan (1981) use this phrase to refer to a fundamental assumption, that underpins all coexisting ideas (vertical perspective). Herein, this research adopted an interpretive paradigm; and, in the second instance, a pragmatist and functionalist paradigm.

In Information Systems research, the role of the research paradigm is to provide a distinct direction based on which information collection and research analysis can be carried out. This research applies the interpretivism paradigm to create a philosophical foundation, and some elements are taken from the pragmatist and functionalist paradigms, aiming to generate constructive and valuable knowledge (Goldkuhl, 2012).

The main principle of interpretivism is to engage with personal symbols and meanings in society. That implies acknowledging their existence, reconfiguring, understanding, and using them as a base for theorising (Goldkuhl, 2012). For Orlikowski and Baroudi (1991: 14), "Interpretive systems of information research presuppose that society (such as social organisations, connections, and labour divisions) are not given ontologically. On the other hand, humans develop and reinforce

it by their actions and interactions." The pragmatic paradigm concerns actions and change, with individuals engaging in an ever-changing reality. According to the interpretivist paradigm's fundamental assumptions, what an idea or concept means is the result of its actual application (Goldkuhl, 2012).

Pragmatism emphasises the distinction between what is and what might be, the beginning of an unrealised future world. The functionalist worldview central to the functionalist worldview is understanding the status quo, social order, social integration, need fulfilment, consensus, and rational choice. Its goal is to comprehend how the interactions of individuals in society lead to them cooperating as a group (Burrell and Morgan, 1981).

The philosophical motivation of this study is defined as an unstructured collection of logically connected presumptions, ideas, or statements that serve as a basis for analysis and investigation (Cohen *et al.*, 2018). Cohen *et al*. (2018) were the first to use the term paradigm to explain the historical development of the scientific sciences, particularly physics and astronomy; it has been one of the most contentious concepts.

Instead of a paradigm, some authors prefer to use terms such as knowledge claims (Creswell, 1994), epistemology or ontology, or even research technique (Neuman, 2002; Knipe and Mackenzie, 2006). There are four basic paradigms (Table 4. 2): positivist, post-positivist, critical, and interpretive (Abdel-Fattah, 2015; Goldkuhl, 2012).

4.2.1 Positivism paradigm

This section details various paradigms to highlight the philosophical foundations used by various researchers and entails the most suitable paradigm selected for this study. The empirical-analytical interest (also known as the technical interest) governs positivist social sciences that seek to comprehend and anticipate phenomena in the realm of labour (which includes material production and systems, people, and objects) to achieve effective technical control (Cecez-Kecmanovic, 2011: 442). "Based on the rationalistic, empiricist philosophy that began with Aristotle, Francis Bacon, John Locke, Auguste Comte, and Emmanuel Kant" (Mertens, 2019), "positivism reflects a deterministic worldview in which causes almost certainly influence consequences or outcomes" (Creswell, 1994: 7). The goal of a positivist is to "test a theory or characterise an experience through observation and measurement to forecast and regulate forces that surround us" (O'Leary, 2004).

Positivism is an ontology that holds that reality exists objectively and irrespective of human experiences. However, this study emphasises realities produced and remade (Burrell and Morgan, 1981). Positivism is concerned with the hypothetic-deductive testability of theories from an epistemological standpoint. Scientific information should be able to be verified or refuted, and generalisable results should be sought. As a result, stating a causal relationship is common, and there should be a close relationship between explanation, prediction, and control (Orlikowski and Baroudi, 1991).

This research collected knowledge by understanding human and social interaction with cloud computing and massive data systems, which resulted in subjective interpretations of reality (risk factors as an example) (Walsham, 1995). Positivists claim that the hypothetic-deductive theory should be evaluated in terms of methodology. The research approach should be value-free, and an objective assessment should collect evidence. A traditional positivist instrument is a quantitative procedure, such as a survey.

In addition, this research is based on the idea that comprehending the meaning contained in human and social interactions is critical. Researchers must immerse themselves in the issue's social environment and examine the conversation from the participants' perspectives. Interviews, case studies, and grounded theory helped us understand risk variables in the cloud and significant data-based situations.

4.2.2 Post-positivist paradigm

Post-positivists believe that several well-developed hypotheses impact any piece of research in addition to and apart from the one being investigated (Cook, 1980). Furthermore, because Khun's (1962) views are preliminary, fresh insights may doubt the entire theoretical framework. According to O'Leary (2004), post-positivists see the world as ambiguous, diverse, and multifaceted in its realities: "What is "real" to one person or society may not be "true" to another." According to O'Leary (2004), Post-positivism is intuitive, and this definition contradicts Mertens' more well-known definition of post-positivism (2005).

This research is not carried out from a positivist or post-positivist paradigm because it assumes a 'value-free research' (Cecez-Kecmanovic, 2011). This research adopted a value-neutral position, predominant in the interpretive paradigm; researchers and practitioners judge when discussing risk because it can be perceived differently. However, the researcher can be neutral in how he analyses it in the context of large data and cloud computing.

4.2.3 Interpretivism paradigm

Interpretivism is not a single, unmistakable tradition. Interpretivism can take numerous shapes. Butler (1998) distinguishes between conservative, constructivist, critical, and deconstructionist perspectives. The interpretive paradigm requires understanding people's subjective meanings in the investigated domains (Goldkuhl, 2012). There is no such thing as universal interpretationivism. The term "interpretivism" can refer to a variety of concepts. Butler (1998) distinguishes four viewpoints: conservative, constructivist, critical, and deconstructionist. Understanding the subjective meanings of people who are examined domains is critical in the interpretative paradigm. According to Goldkuhl (2012), in any interpretive research, the goal is to show how members of a social group build their unique realities and imbue them with meaning by participating in social processes and demonstrating how these members' meanings, ideas, and intents interact to form their actions.

Qualitative research methodologies, including interpretative case studies, ethnography, discourse analysis, and action research, are associated with the interpretive paradigm. Such linkages represent various cognitive objectives in knowledge diverse ontological and epistemological mutually exclusive assumptions (Cecez-Kecmanovic, 2011). Some authors use interpretivism and constructivism interchangeably (Cohen *et al*., 2018). Interpretivist and constructivist studies aim to understand the world of human experience (Cohen *et al*., 2018), which means that reality is socially constructed (Cohen *et al*., 2018). The interpretive/constructivist researcher is interested in participants' viewpoints on the issue under consideration. (Creswell, 1994) and how their background and experiences affect the research. Constructivists, unlike post-positivists, do not begin with a theory; rather, they construct or inductively develop a theory or pattern of meanings as they conduct research. (Creswell, 2002). Qualitative data collecting and analysis methodologies and combining qualitative and quantitative procedures are more common among constructivist researchers (mixed methods). To add depth to a description, quantitative data can augment or expand on qualitative data (Mackenzie and Knipe, 2006). This study is framed under the interpretative paradigm in IS for various reasons.

Firstly, this paradigm simplified the interpretation of risk variables. In big data and cloud computing environments, the researcher deciphers the significance of risk factors. Secondly, this paradigm helped the researcher understand the context of security risks. Thirdly, the researcher can establish emerging relationships among different aspects of big data and cloud computing risks via this paradigm. Fourthly, rather than testing and validating hypotheses, interpretivism

encourages researchers to be more interpretive and inductive (Abdel-Fattah, 2015); this study inductively considered participants' varied points of view and created risk knowledge in cloud significant data scenarios. Fifthly, interpretive research is versatile; action research and grounded theory (GT) are just a few methods to which it is applicable. These approaches enabled iterative questioning and response interpretation (Abdel-Fattah, 2015). Interpretivist approaches aim to comprehend social phenomena, behaviours, processes, and institutions as subjectively meaningful and socially produced entities (Cecez-Kecmanovic, 2011). Since the concept of risk is socially constructed and risky for some people and institutions may not be dangerous for others, the interpretive paradigm is well suited for this study. The paradigms of positivism and interpretivism are usually seen as the most applicable (Niehaves and Stahl, 2006).

A comparison of the four paradigms is indicated in Table 4.1.

Table 4.1: A comparison of research paradigms (Abdel-Fattah, 2015)

Element	Positivist paradigm	Post-positivist paradigm	Critical paradigm	Interpretivist paradigm
Purpose	Primarily, the goal is to explain what causes phenomena to be predicted and controlled, and positivists strongly emphasise theory verification. The research goal is etic, seeking to discover the "truth."	It admits that an objective reality can only be grasped imperfectly. According to this point of view, human intellectual systems are flawed, and life phenomena are intractable; thus, it is impossible to capture a "true" reality fully. The post-positivist emphasises "theory falsification."	The ultimate goal is to eliminate oppressive human interactions (oppressive defined as forced assimilation). The goal is to motivate participants to reform to achieve group empowerment and freedom from oppression.	Understand the "as-is" society, which was built subjectively. Interpretivism seeks both idiographic and emic goals. Its goal is to define people's perceptions of the world, how they interact with one another, and the environments in which these interactions occur.
Ontology	There is only one reality, which can be observed by someone with little control over the object in question. Positivists believe that there is only one observable, identifiable, and quantifiable reality (a position known as naive realism). Positivism holds that natural laws govern a single simple world and that knowledge is context-free and governed by cause-and-effect laws.	Post-positivists acknowledge reality but believe it can only be grasped and assessed in imperfect ways (a critical realism position). Post positivism assumes an imperfect "reality" due to human cognitive limitations and the complexity of things.	Ethnic, cultural, gender, social, and political values all impact reality. This "reality" becomes accepted as "real" over time. There are many different versions of reality, each of which is produced. Internal and external indicators are used to decipher reality. As a result, it may be studied, bargained over, and reached an agreement.	Reality comprises people's mental constructs of the objects they interact with and how that interaction affects the observer and the scenario under consideration. Reality is internal and multifaceted; it is subjective and influenced by the individual's experience, circumstance, social environment perceptions, and the researcher's relationship.
Epistemology	For positivists, dualism and objectivism are essential. This implies that the researcher, the research participant, and the study topic are all assumed to be self-contained entities. "Knower (research participant) and known (researcher) are dualism," and the researcher can research the participant and issue without bias by using rigorous, standard methods (objectivism).	Post-positivists advocate a hybrid of dualism and objectivism. Although the researcher may influence the research, impartiality and researcher subject independence are essential standards for the research process.	A researcher's and a participant's connection is transactional and subjective, as well as a discussion.	Interpretivism advocates for a transactional and subjectivist approach to reality, arguing that it is socially produced. As a result, recording and characterising the "lived experience" hinges on the dynamic interaction between researcher and participant. The point of view is subjective. There is no separation between the knower and the one who is known.
Methodology	It entails carrying out empirical research to put hypotheses to the test. The experiment's conditions are meticulously controlled to avoid bias.	It places a strong emphasis on debunking theories. Data gathered about a scenario allows for knowledge acquisition, and opinions are sought without interactions to interpret people's behaviour.	Requires a dialectical process to change misconceptions into a well-informed comprehension of the research topic.	Its purpose is to understand the phenomenon's social surroundings and process, which is accomplished using a hermeneutic (interpretive) dialectic (represents) circle.

Some authors (Goldkuhl, 2012; Mackenzie and Knipe, 2006; Porra *et al*., 2014) have pointed out that pragmatism, design science, and action research paradigms have also influenced IS research. According to Burrell and Morgan (1981), functionalism and social relativism are also relevant IS paradigms. The researcher briefly describes them in the following section.

4.2.4 Pragmatism

The interpretive paradigm requires understanding people's subjective meanings in the investigated domains (Goldkuhl, 2012). There is no such thing as universal interpretationivism. The term "interpretivism" can refer to a variety of concepts. Butler (1998) distinguishes four viewpoints: conservative, constructivist, critical, and deconstructionist. According to one of the critical tenets of pragmatism, the meaning of an idea or a concept is the idea's/practical concept's ramifications (Goldkuhl, 2012). This means that pragmatism concerns what "could be," a vision of a future world that has yet to be realised, and what "is." Pragmatism is concerned with a practical understanding of knowledge and makes a significant difference in practice. This includes normative knowledge of aims and values and prescriptions for means.

Table 4.2: Pragmatism vs Interpretivism: Ideal-typical differentiation (Goldkuhl, 2012)

	Pragmatism	**Interpretivism**
Ontology Emphasis on data Types of expertise The significance of knowledge Investigation Type Generation of data The researcher's role	Symbolic realism is a type of realism based on changes and actions. The useful knowledge Action-oriented Inquiry Assessment and intervention provide data Change-oriented	Understanding Constructivism Beliefs (Socially Constructed Cognition) Interesting Observational research The interpretation of data Interested in learning more

As can be deduced from Table 4.2, any philosophical or practical framework does not constrain pragmatism. Pragmatist researchers are concerned with the research problems 'what' and 'how' (Creswell, 1994). While some mixed-methods researchers adhere to the transformative paradigm philosophically, pragmatism is widely regarded as the fundamental philosophical underpinning of mixed-methods research (Somekh and Lewin, 2005; Tashakkori and Teddlie, 2003).

As a consequence of the fact that philosophical or practical frameworks do not constrain pragmatism, hybrid techniques can be used with any paradigm. The pragmatic paradigm prioritises "the research challenge" and considers numerous options (Creswell, 2002). 'Central' data collecting and analysis procedures are the most likely to deliver answers with the study topic in

mind, regardless of philosophical affiliation with any other paradigm (Mackenzie and Knipe, 2006).

Pragmatism suited this research well because this paradigm approaches knowledge from an instrumental perspective. It is assumed that learning can be helpful in practice; these two elements are critical in research that addresses real problems associated with cloud computing risks mediated by big data. Pragmatism is an excellent fit for this investigation because it is not discriminatory towards any theory, technique, or procedure. It serves as a filter for seeing the other study levels (Goldkuhl, 2012). As a result, pragmatism provides a distinct perspective on theories and techniques. It is a form of knot that connects threads that go to several different regions of thought (Garcia and Quek, 1997). The goal of a pragmatist is to achieve perfect clarity in our thinking about a thing, which includes emphasising its good practical effects (for example, risk considerations in big data) (Garcia and Quek, 1997).

4.2.5 Design science research

Design science research (DSR) has evolved as a significant study paradigm in the field of IS over the previous decade (Myers and Venable, 2014). According to Venable and Baskerville (2012), design science research creates a new purposeful object to address a generic type of issue and assesses its value in solving that type of issue. Instead of the social science or behavioural science paradigms that have come to dominate IS research, DSR is founded on the engineering research tradition, where the primary purpose is to design new technologies or artefacts that may be used to change (and preferably enhance) the world.

The two basic components of DSR are build and assessment. In natural science, these are analogous to the discovery justification pair. Building an artefact for a given goal is called building, and the evaluation determines how well the item functions. The design science build process is not as well-known as the natural science discovery process (March and Smith, 1995). This research did not emphasise developing artefacts or evaluating their specific performance; hence, this paradigm does not fit this study well.

4.2.6 Action research

Power and Naysmith (2005) give a simple definition for Action Research (AR), stating that it consists of two constructs: “action” and “research,” as well as the linkages between them. It is possible to act without conducting research or to conduct research without affecting. It is important

to remember that action research is distinguished from other types of inquiry by the unique mix of the two notions. Action research was created to solve real-world issues and close the gap between theory and practice (Vaccarino *et al*., 2006); and is also distinct from typical problem-solving research in that it involves participants in the study process. Action research is classified into four types: traditional action research, contextual action research (action learning), radical action research, and educational action research (O'Brie, 1998). Traditional action research encompasses the concepts and practices of Field Theory, Group Dynamics, T-Groups, and the Clinical Model. Insofar, action research is contextual as it entails reconstituting the structural relations among actors in a social environment. Radical action research has a strong focus on emancipation and overcoming power imbalances. Educational action research has its foundations in the belief that professional educators should become involved in community problem-solving (O'Brie, 1998).

Rather than isolating the observer and the participants, as other research methods do, AR focuses on transforming participants into researchers who are more likely to use what they have learned when they try it themselves (O'Brie, 1998), therefore enhancing all participants' skills (Koshy, 2006; De Villiers, 2005). It establishes a framework for tracking and improving a research project's progress through iterative stages (Abdel-Fattah, 2015). This study looked into a topic that has real-world ramifications. However, it did not intend to implement any solution. Hence, action research is not a good match for this study. Moreover, action research implies active participation with the people involved in the problem to provide insights into the potential solutions, which goes beyond the goals of this project.

4.2.7 Functionalism

This paradigm considers the current state of affairs, social order, social integration, consensus, need fulfilment, and rational decision. This experiment was conducted to show how the various parts of a social system interact to make a useful whole (Burrell and Morgan, 1981). In the functionalist view, society is more than the sum of its parts; each portion contributes to the society's general stability. According to Durkheim's model, societal components are interdependent but cannot operate independently. When one part of a system fails, the rest of the system must adapt to fill the gap. According to functionalist theory, societies are primarily social institutions with a definite function. Functionalism inhibits people from improving their social environment, even in their best interests. Instead, functionalism considers social change advocating undesirable because many parts of society appear to compensate for any problems that may come organically (Crossman, 2020). This paradigm is relevant to this issue because it deals

with the status quo, social order, social integration, consensus, needs fulfilment, and rational choice. It aims to describe how the various components of a social system interact to form a cohesive whole (Hirschheim and Klein, 1989).

4.2.8 Social relativism

This paradigm is used to describe theoretical conceptions that are incompatible with absolute or universal norms or criteria. Because there are no objective standards for reality, knowledge, or truth, this viewpoint concludes that they must be socially constructed (Marshall, 1994). As a result, the social relativist paradigm seeks answers in the domains of individual awareness and subjectivity and a social actor's frame of reference rather than in those areas. According to this viewpoint, societal norms and organisations represent men's interpretations of their surroundings (Silverman, 1987).

Rather than relying on mechanical and biological analogies to model civilizations, social relativism emphasises the importance of examining cultures through the eyes of the people who live in them. They view societies as a web of shared assumptions and meanings. It investigates the basics of social reality to explain social actions and the members' consciousness and subjectivity. Human relationships in communities are presumed to be coherent, organised, and interwoven. In other words, it is an ideographic presumption regarding society's nature. Its assumptions are disagreements about whether an organisation operates in more than a conceptual sense (Burrell and Morgan, 1981), for example, if societies exist outside of human consciousness (Maric and Flensburg, 2012). This research aligns nicely with the social relativism paradigm since it assumes that meanings and symbols are socially produced.

Based on the analysis presented above and as summarised in Table 4.3, this research will primarily adopt an interpretive paradigm; and, in the second instance, a pragmatist and functionalist paradigm. Thus, I n the present study, the interpretive research paradigm enables the researcher to obtain a social construct through the experiences of experts and professionals, thereby fulfilling the aim of the study. It is constructivist in nature wherein, through thematic analysis, the knowledge relevant to understanding risk assessment methods in different nations has been extracted and analysed. Through multiple perspectives, the socially constructed reality is framed with the help of this research paradigm (Fleming and Zegwaard, 2018). It also enabled the researcher to understand the concepts, opinions and experiences of professionals from different countries and continents. This study followed a qualitative and inductive method at the research

methods level, in which interviews were used to gather data and grounded theory to analyse it. Data triangulation and researcher triangulation are privileged to increase the quality and rigour.

Table 4.3: Research paradigm, methodology, and quality

Element	Description
Primary paradigm	Interpretivism
Secondary paradigms	Pragmatism and functionalism
Research methodology	- Qualitative and inductive - Interviews - Grounded Theory
Focus	- Risk assessment techniques - Firms
Quality criteria	Data triangulation, researcher triangulation
Ethics	Do not harm, value-neutral

4.3 Ontological, epistemological and ethical considerations of the research.

Ontology, epistemology, and ethics are three concepts related to the concept of paradigm. Ontology is the philosophy of being, or what scientists consider reality, being, and existence (Burgess *et al*., 2013). It focuses on what is real and the nature of the world around us. Epistemology focuses on how ontological knowledge can be obtained. As Johnson and Duberley (2000) explain, epistemology is the study of scientific knowledge frameworks that serve as a guideline for expressing what knowledge claims can be made. Ethics is concerned with research principles (Porra *et al*., 2014). At the ontological level, this research assumed that the reality of risk in significant data contexts depends on the observer; at the epistemological level, this research assumed that it is possible to understand the world around us and search for patterns and relationships to obtain knowledge about risk in big data and cloud computing settings; at the ethics level, the researcher tried to be value-neutral at the moment to compare different risk assessment alternatives in big data and cloud computing settings. In the current study, the researcher used ontology to search for reality through triangulation and qualitative analysis. The researcher investigated how things are in risk assessment processes in different nations; therefore, the ontology was appropriate to use existing evidence in research. Further, the research is also based on epistemology as it helped the researcher to evolve how the information was gathered using interviews, observations and empirical evidence. This approach enabled the researcher to identify the validation of the information through triangulation, and epistemology was applied to understand the possibility, desirability, reliability and generalizability through various sources.

4.4 Research methodology

A research technique is a set of methods and strategies for choosing subjects (and objects), measuring or monitoring events, and collecting, processing, and interpreting data (Cecez-Kecmanovic, 2011). An approach can be defined as a codified collection of goal-oriented procedures in information systems research. These procedures are designed to lead the work and collaboration of the multiple stakeholders participating in the processes (Iivari et al., 1998). In this research, the framing of methodology provides a systematic structure to the data collection process and reaching the study's findings (Fleming and Zegwaard, 2018). The methodology section, in accordance with the aim of the study, makes a planned effort to compare and analyse the risk assessment methods based on different nations in relation to big data using cloud services. For example, question lab experiments and field studies are commonly used in research (Cases-Kecmanovic, 2011; Dwivedi and Kuljis, 2008; Mingers, 2003). The term "research technique" refers to a strategy for organising and performing a study, engaging with the phenomena under inquiry, and developing and confirming knowledge claims. Method assumptions and their linkages to theory and the selection, justification, and usage of method(s) in a specific research context are all topics covered by research methodology (Cecez-Kecmanovic, 2011).

Research methodology emphasises that research methodology encompasses far more than a particular method or collection of procedures. Part of the research technique involves connecting theory to overall research strategy and operations and publicly addressing how knowledge claims are created and confirmed (Cecez-Kecmanovic, 2011). There are three (high-level) basic ways to approach a research process: qualitative, quantitative, and mixed.

4.4.1 Quantitative research

Quantitative research uses statistical methods to investigate numerical data or data converted into numbers. Quantitative research deals with numerical data translated into numbers. The term statistics refers to the most common methods for analysing numerical data. Statistical procedures deal with how numerical data is organised, analysed, interpreted, and presented. Statistics is a broad topic of study with applications in various fields, including information systems and other types of data analysis (Williamson and Johanson, 2018).

4.4.2 Qualitative research

Qualitative research can be described in various ways, with different people interpreting it differently. It is a commitment to multiple forms of naturalistic or interpretive approaches to its subject matter and a constant critique of positivism's politics and methodology (Garcia and Quek, 1997). According to Maanen (1979), qualitative label techniques in the social sciences have no definite meaning. It is best described as a catch-all phrase for numerous interpretive processes aiming at explaining, decoding, translating, and otherwise verbalising the meaning of certain more or less naturally occurring occurrences in the social environment, rather than quantifying their frequency. Qualitative research is based on unquantifiable processes and meanings. Qualitative research refers to the process of collecting, decoding, translating, and comprehending the term's meaning rather than the frequency of naturally occurring events in the social context.

In the present study, qualitative research was conducted to enable the researchers to go deep inside the research topic with the help of detailed interviews wherein the experiences, knowledge and concepts of professionals on risk assessment methods in their specific domains have been highlighted. The qualitative analysis helped understand experts' perceptions of big data and cloud computing issues. The explanatory data helped to draw patterns from the concepts and insights. The interactive interviews and triangulation method helped validate the collected data and provided a piece of well-founded credible information from reliable sources (Yeasmin & Rahman, 2012). This study implemented a four-step process that answers the research questions and meets the objectives through qualitative research. This study first identified the field of research related to risk assessment of big data using cloud services, and based on the research topic; further information was collected. This step consisted of deciding the unit, the topic, and the analysis. The analysis uses different techniques to address risk issues in big data and cloud computing contexts. The firm itself served as a secondary unit of study. Then, the study gathered primary information using responses from participants who were professionals working in close vicinity to cloud platforms. This step collected information about the different risk assessment practices in big data and cloud computing. For this research, the data was collected via interviews from respondents from the four targeted countries working in cloud platform organisations or using services from the cloud, secondary information available on the firms' websites for analysing, and information from the Internet, in general, were also included as detailed in the table below (Table 4.4).

Table 4.4: Data collection framework by researcher

Qualitative research		
Data was collected via interviews, the firm's websites, and information from the Internet.	The focus is on Canada, Jordan, South Africa, and the UK. There are between 3-4 firms per country; 2-3 interviewees per firm (the target is ten participants from each country).	Inductive analysis and finding patterns. Triangulation to add rigour.

After having gathered the information, a review of the evidence was conducted. The study assessed if the collected data was valuable. The gathered information was analysed and reviewed to eliminate conflicting and irrelevant data and to detect incomplete data. After having reviewed the evidence, the study identified patterns. This stage sought to organise the data to discover interesting patterns regarding the problem (i.e., risk factors in big data and cloud computing settings). Grounded theory is based on the systematic collection, analysis, categorisation, and validation (iteratively) of data that can aid in defining an interesting phenomenon.

The study employed data and researcher triangulation to improve the quality and rigour of this research. The researcher considered several elements to ensure that this research meets the specific ethical requirement. This comprised gaining the required permissions to conduct this study, securing participation, protecting the confidentiality of the data at all stages, informing the participants about the potential risks of participating in this research, and following the North-West University ethical process.

4.5 Research design

Research design provides a structure to the study. The researcher obtains information and reaches the findings around the framework. The choice of a specific method would lead to data collection, and a research design must be congruent to the aim of the study. Research design relates to how data collection and analysis are linked to achieving objectives (Almalki, 2016). This research design section shows how the study responded to the research objectives and questions. This section also aims to describe, explain, and justify finding responses to the research questions. Therefore, this study takes a four-step approach to answer the research questions and achieve the goals.

4.5.1 Specify the domain

The unit, the analysis, and the topic were decided at this step. The focal questions presented in the introduction determined the domain of inquiry and certain methodological presuppositions. The study unit of analysis used many approaches organisations use to handle risk issues in the context of big data and cloud computing. An auxiliary unit of research was the firm itself. This study focused on risk assessment implementation, utilisation, and management in big data and cloud environments. The researcher examined businesses in a variety of industries.

4.5.2 Information gathering

This step entailed gathering data on the various risk assessment approaches used in big data and cloud computing. The researcher gathered data from as many primary sources as possible. Mason *et al*. (1997) classified primary source material into four categories: 1) A collection of official documents, such as unpublished documents, diaries, and memoirs; written, such as official documents and letters; 2) material, such as things, artefacts, and visits to real-world locations; 3) traditional, such as historical traditions retold in secondary sources; and 4) eye witness evidence. In the present study, semi-structured interviews were conducted wherein the respondents from the four nations were asked about risks and concerns in their organisations using big data and cloud services; further, information available on the firms' websites concerning cloud platform usage and information security malpractices and concerns were also used to evaluate and validate the information, and information from the Internet in the form of journals and articles was also used to obtain a valid generalization of the findings for this study.

4.5.3 Interviewing

Interviewing is a qualitative research process that collects information through open-ended questions that give an insight into the respondent's experience or opinion on the desired topic. An interview provides in-depth information and is a flexible method to explore ideas and information. Interviews can consist of pre-determined questions designed to obtain opinions and information from respondents. However, questions are asked formally in semi-structured and unstructured interviews, giving a larger space for flexibility and discussion. In this study, a semi-structured interview was conducted in which components of both structured and unstructured interviews were present. The study applied qualitative analysis wherein the interview of respondents from different nations was done based on the perception of big data risks in their nations. Herein, some sets of questions were prepared, and some were left to be asked as per the

replies and content provided by the interviewees. The information was systematically analyzed to find a pattern and generalize the findings. Further, the triangulation method confirmed the validity of the inferences drawn with the help of observation from participants, researching from secondary sources and validating the data

The researcher followed several guides to develop semi-structured interviews as per DeCarlo (2018):

- Make a list of the study's main research topics.
- Describe the general areas of knowledge relevant to answering these questions.
- Create questions for these influential groups tailored to specific types of respondents. The goal was to use their knowledge and experience to assist others.
- Adjust the interview language to the respondent's degree of competence and knowledge.
- Carefully word the questions, so responders feel compelled to respond entirely and honestly.
- Instead of an acceptable "account" of behaviour, ask "how" inquiries rather than "why" questions to acquire stories of the process. "How did you end up being a part of this group?"
- Develop critical questions to elicit more detailed and intricate responses; the more detail, the better.
- Begin the interview with a "warm-up" question that the respondent can quickly and thoroughly answer (though not too long). It is not required to be directly related to the researcher's search criteria (though it may). Nonetheless, making an early connection makes you feel more at ease with each other, making the rest of the interview go more smoothly.
- Considering the order in which the interview unfolded. What should be the top priorities on the agenda? What occurs next, in a "natural" way? The researcher may need to adjust your strategy after a few interviews.
- When rapport has been built, ask difficult or potentially embarrassing questions toward the end of the interview.

- Finally, assuring that the last leave the respondent feeling empowered, heard, or otherwise thankful for their time.

The interviews were transcribed in Atlas.ti software, which was used to codify the answers. Firstly, the word clouds were generated and used to identify the codes. The data was then coded, and Atlas.ti identified the themes relevant to answer the research questions. Atlas.ti was used to generate reports with quotations to support findings in Chapter 5. Only knowledgeable professionals having relevant exposure and experience in information technology and cloud-based platforms were included in the study. Some of the professionals from relevant departments considered inclusive for this study were IT operations and infrastructure, IT operations and access management, enterprise architecture and support, IT lead (helpdesk, system administrators, network operations and security, and infrastructure/network administrators), IT information and security Operations. Following the recommendations provided by DeCarlo (2018), several questions were asked to provide rigour to the interview

- Direct questions: Is it simple for you to keep a smile on your face when serving customers? Are you satisfied with how you and your husband select how to spend money? To avoid overly influencing the interview's direction, such questions should probably be saved for the conclusion.
- Indirect questions: What do most people in this area think of management's treatment of its employees? To connect to the individual's point of view, ask, "Is that how you feel too?"
- Question structure: I want to move on to a new issue now.
- Follow-up questions: Eliciting additional information from the interviewee, such as could you comment on that? What exactly do you mean?
- Probing questions: Direct inquiry to follow up on what has been said.
- More specific inquiries: What were you doing at the time? What did X say in response to what you said?
- Question interpretation: Do you imply that your leadership role has shifted from encouragement to defectiveness? Is it correct to say that you do not mind being pleasant to customers, but it is more difficult when they are rude or demanding?

All relevant data for answering the research questions were preserved in four categories: occurrences primarily impacting security in cloud computing; significant data contexts; company

information; and technology advancements to solve risk issues in cloud computing and big data environments. Large sheets of sketch paper and pencils are the tools of choice when organising the data. Drawings, diagrams, and figures are often used to categorise data. Colour coding is used to distinguish between different data types (i.e., types of risk, potential solutions, and best practices).

4.5.4 Data analysis to determine the risk of big data and cloud computing

Respondents were asked questions about their level of knowledge of big data and cloud computing. These questions were asked on a Likert scale from 1 = basic knowledge to 5 = expert. Similarly, questions were asked about the level of knowledge with risk assessment. These questions were analysed using descriptive statistics (frequencies) presented using graphs.

In addition, the study collected qualitative data to extract experts' knowledge regarding a clear risk assessment plan in their respective countries, what is covered in their current risk plan and their knowledge about the risks related to data protection.

Atlas.ti Software was used to analyse the qualitative interview data, allowing for the creation of word clouds. These word clouds gave useful information on the codes that were utilised to generate the themes. The themes were further investigated in Atlas.ti to find patterns that would respond to the research goals. A thematic analysis was used to classify patterns or themes from the qualitative data. Furthermore, the thematic analysis allowed for identifying noteworthy or intriguing patterns in the data and using the themes to answer research questions or express an opinion on a topic. This was much more than summarizing the material and making sense of it as a sound thematic analysis (Braun and Clarke, 2006).

Atlas.ti was utilized in this study to help organize the data and synthesize the findings. As a result, the software was chosen as the most often used program (Silver and Lewins, 2014). Complete analysis and reporting were possible due to familiarity with the software. The Atlas.ti program was used to analyse the data in the following ways: preparation and importation of data, familiarisation of oneself with the data by creating word clouds, coding of the data, and retrieving and examining codes and quotations. According to Woods et al. (2016), the grounded theory procedure was used to convert all transcripts to the qualitative analysis program Atlas.ti (version 9.1) and narrative analysis. For all transcripts, this entailed producing word clouds, quotations, and codes, as well as looking for conceptual connections, differences, and the most important and relevant information.

The theme determined the length of the quotes (half of a sentence, whole sentence, or paragraph). Each quotation may have several codes. The codes were synthesised and arranged through memoranda to generate the many themes supplied by Atlas.ti, then categorised and analysed to answer the study questions.

4.5.5 Evaluate and justify your method

Qualitative analysis used for the study helped the researcher detail the risks and concerns in different nations and collected information relevant to the topic, including the participants' sentiments. Further, considering respondents from four different countries helped the researcher to understand the approach of IT professionals and experts in different nations and continents and create a comparative analysis. With the help of triangulation of the qualitative data, the detailed information collected from respondents and secondary sources could be validated and generalized to frame an opinion in line with the aim of the study. Thus, it can be said that this study used the triangulation method, enabling a holistic validity of information gathered from qualitative analysis. A detailed fragmentation of the methods used and their justification has been done below:

a) Critique the evidence

The researcher evaluated the value of the data gathered. The information collected was analysed and assessed. This was done as some of it could be erroneous, conflicting, or unimportant, and the majority could be incomplete. With dubious or unproven evidence, several analytical approaches were applied. Examples include basic reasoning, evaluating the sources' credibility, counting the number of times the same observation is made and examining the overall coherence of the evidence. Using these methods, the study assessed the data's internal consistency and agreement with data from the exterior and other sources. The facts were confirmed and condensed to put some meaningful order on the original content (Mason, 1997). Choosing which accounts to use and why required comparing sources and "hearing many voices" (Fritze *et al*., 2002). It was also critical to review the material and ensure it was accurate and reliable.

For a variety of reasons, internal criticism is beneficial. It was first employed to track down trustworthy evidence (Golder, 2000; Gottschalk, 1969). Professional and experts verified documents, interview transcripts, and website information for intentional and unintentional mistakes (Shafer, 1974). Interpretive critique, or determining what the interviewee intended, was the phrase for this method. This stage necessitated an understanding of the interviewees' culture and the events period. Rather than being examined in isolation, all testimonies were considered

entirely (Elton, 2002). Negative internal criticism, or evaluating the integrity of the information's statements, was the second goal of internal critique. The accuracy of the information's claims of beliefs and perceptions about previous events had to be determined during this step. Internal criticism's third goal was to assess the objectivity of observations. Researchers should prefer many independent sources of testimony. Five criteria should be utilised to evaluate such sources:

- Competence (are names, locations, and times reported accurately?)
- Expertise (how well-versed in the subject is the interviewee or source?)
- Objectivity (is the interviewee or source willing to report accurate information?)
- Trustworthiness (does the interviewee or source have a track record of honesty? Is there any self-contradiction in the document?)
- Confirmation (does the evidence come from equally trustworthy interviewees or sources?).

b) Determine patterns

The next step was to organise the data to discover exciting patterns (i.e., risk factors in big data and cloud computing settings) using Atlas.ti. Thus, the meaning of the collection of admissible and arranged facts and information must be understood. The researcher's contribution is interpreting, explaining, and deducing from the evidence. The research went from being primarily empirical to being more inductive and theoretical at this time. Identifying patterns necessitated a creative and poorly understood process of experimenting with hypotheses that generate designs and evaluating data through these lenses using Atlas.ti.

c) Grounded theory

Grounded theory (GT) is a tool that helps to uncover patterns. According to Galal-Edeen (2005), the primary focus of a researcher building a grounded theory is the methodical collection, classification, and validation (iteratively) of data. Let the data speak, but formalise it and use it to guide future data collecting and analysis. To realise that goal, the grounded theory approach serves as a road map and the methods and techniques required (Strauss and Corbin, 1998).

Grounded theory, as the name suggests, is grounded and anchored in the data. Herein, the researcher kept the analysis closely connected to the data gathered and developed a theoretical framework derived from the data. The researcher encompassed the detailed analysis within the frontiers of the data collected, and the analysis was kept closely linked to the immediate data. In

this study, the researcher made a constant comparison of the observations derived from the respondents. Lines of data were appropriately coded, and then the theoretical sampling was followed by data collection. This enabled the researcher to understand the concept of risk assessments in use in nations concerning big data in cloud computing.

The grounded theory technique was used for the interviews, and the relevant material was obtained from the Internet. Grounded theory's application in big data and cloud risk evaluation is justified by the principle that it provides a set of procedures for categorising and analysing data that suit the interpretative approach because it keeps the analysis close to the data and provides inductive discoveries about the phenomena under study (Strauss and Corbin, 1998). Finally, the grounded theory process is iterative, involving frequent movement between concept and data and continuous comparison across various sources of evidence (Abdel-Fattah, 2015).

The researcher used the following steps proposed by Bernard (2013) to carry out grounded theory:

- Compiling all of the data from those categories and comparing them.
    - Create transcripts of interviews and read a small sample of text.
    - Look for potential analytic categories (that is, themes) that emerge.
- Thinking about how categories are related to one another.
- Building theoretical models based on the relationships between categories, regularly evaluating the models against facts, predominantly negative scenarios.
- Using quotes from the interviews to illustrate the idea, provide the findings of the analysis (exemplars)

Essential procedures in grounded theory:

It is challenging to summarise the concepts of GT because there have been several adjustments and adaptations in the 40 years since its inception — and this is one of the key challenges in using it as a research approach. Chun Tie *et al*. (2019) conducted a review of the significant works on GT and claimed the following undisputed principles:

- Continuous comparison
- Exact coding and theory construction
- Theoretical sampling
- Data collecting that is iterative

- Iterative data collection

The way research tasks are carried out differs greatly between ground theory and other sorts of study. In certain research, selecting, collecting, and analysing data are consecutive. In GT, these three steps are iterative. Instead of waiting until all information is recorded, assessment starts as early as the first data is received (Sharma *et al*., 2019).

The approach toward sampling is another key variation between GT and other research methods. Statistical random sampling aims to identify a large enough sample representative of the studied population. There is no such thing as a minimum or maximum sample size in GT research. Researchers frequently choose early "key informants" examples who propose practical study areas (Goulding, 2005). According to Glaser (2008), as Geiger and Turley (2003) quoted, additional samples should be undertaken based on theoretical relevance. The pieces can be chosen to express distinctions from the previously researched group rather than representative. A researcher might, for example, desire to investigate different groups within the same organisation or separate organisations (this is referred to as difference maximisation). A deliberate endeavour is made to find new evidence (disconfirming cases) that contradicts the initial concept to enrich theory development (Sharma *et al*., 2019).

To discover trends, data is compared regularly. There is a careful line-by-by-line reading of the text to identify possible themes (known as categories). This is followed by comparing the differences and similarities between the various categories. It is necessary to revise the definition of a category or create a new variation when contradictions are discovered (Sharma *et al*., 2019).

Researchers should describe data processing, coding, and theory formulation methods in detail. The study's reporting will be affected by this

d) Literature review

This study also used some elements of systematic literature review, which involved:

- selecting preliminary keywords;
- validating keywords;
- selecting inclusion and exclusion criteria;
- selecting indexing databases to analyse;
- screening and scanning documents;

- reviewing the full text of selected documents; and
- analysing and generating the theoretical framework and the conclusions.

This study focused on journal articles, conference papers, and grey literature, complementing high-quality articles (Kitchenham and Charters 2007; Brereton *et al*. 2007).

4.6 Ontological position of the study

This research is built around the ontological position of the interpretative research model (Scotland, 2012). The theory of interpretive knowledge is subjective, seen as personal, and based on phenomena from the real world (Scotland, 2012). The interpretative scholar accepts that their interaction with their environment creates emotional meanings. Interpretative research intends to study social phenomena in their natural habitats. This research considered all participants' different views and refrained from imposing a superficial understanding of the predetermined situation (Orlikowski and Baroudi, 1991).

Based on the interpretative research model, this study involves the following steps —

- Define the objective and issue to analyse.
- Define the methodology to reach the objective.
- Assess if the method is pertinent.
- Collect data and implement the methodology.
- Interpret the phenomenon of avoiding biases.
- Elaborate on a theoretical framework outlining the main interactions among constructs by using grounded theory techniques.

It is impossible to comprehend reality without considering the human actors who form it because interpretation recognises reality as a social product. (Orlikowski and Baroudi, 1991), for example, ethnographic research and case studies learn about reality through social structures, including language, consciousness, meanings, and standard documents (Klein and Myers, 1999).

Grounded theory (GT) approaches are also utilised in interpretive research to systematically collect and evaluate data inductively and develop a set of constructs and relationships between these constructs. Furthermore, GT is a strategy for building a theory from the evidence collected and

examined methodically, as stated by Glaser and Strauss's original authors. Using this procedure, the hypothesis being developed is substantiated (data support it) (Glaser and Strauss, 1697). The GT technique is similar to inductive research methods, in which theory is induced by evidence (that is, it emerges from data and is created from there) (Glaser and Strauss, 1697).

- There are four primary characteristics of GT (Urquhart *et al*., 2009): The primary purpose is to construct a theory.
- Before beginning the study, researchers must guarantee that they have no predetermined theoretical beliefs. Analysis and conceptualisation are created by the primary process of constant comparison, in which each data component is compared to all existing conceptions and constructs.
- Theoretical sampling selects portions of various data kinds, with researchers deciding on the next chunk to be included in the sample based on an analytical foundation.

4.7 Sampling

Sampling is a crucial aspect of research methodology as it helps in obtaining the information from the correct set of respondents as per the aim of the study. In research, sampling is the process of choosing the correct sample from the population to derive inferences. Sampling should be unbiased and free from measurement errors to derive impactful and reliable inferences from the study. The sample size is derived from the target population to obtain critically relevant information.

In this study, the target population is the professionals and experts working in IT-based organisations directly involving big data using cloud services. The sample size for the study includes 40 professionals from four different nations. The experts with sound experience in cloud-based platforms and IT service-providing firms were only selected because it was important for the respondents to be aware of the risks and issues in their organisations and the concerns that have impacted the security and confidentiality of their big data in the cloud. Further, professionals in direct link with IT services could highlight the latest risk assessment methods used in their countries and the seriousness of risk assessment methods taken in their respective firms due to increasing online work practices. Samples from four nations, Canada, Jordan, South Africa and the UK, were selected. Between 3-4 firms per country, 2-3 interviews per firm were conducted. The target was ten participants from each country, so respondents' opinions could be generalized and validated. This study helped involve knowledge of big data in cloud computing services, such

as team leaders and team members working in information technology departments. The participants were selected according to their experience and knowledge of risk assessment, big data, and cloud computing operations.

This research focused on four countries: Canada, Jordan, South Africa, and the United Kingdom. These countries were selected because they have many firms that use cloud services. Canada and the UK have pioneered big data and cloud computing. According to a Frost and Sullivan (2020) analysis, the Middle East and Africa big data analytics market would increase by 28% annually through 2025, hitting $68 billion in sales (Jordan and South Africa are included).

4.8 Data types and data collection methods

Data were primarily gathered through interviews. Interviewing is not only about getting valid answers or questioning whether a specific point is correct or wrong; it is also about getting the information required to reach the research goals, uncover critical information, and better understand a phenomenon (Draganova, 2015). The interviews were complemented from online data sources such as videos, articles, and the websites of the companies selected.

The data used in this study is primarily audio from the interviews that turned into text. The interviews were transcribed to get an accurate account of the participants' points of view. Besides interviews, information was also gathered from the firms' websites and online information.

4.9 Quality and rigour conditions

According to qualitative researchers, one strategy for achieving rigour is triangulating methods. Contrastingly, the employment of various methodologies or triangulation demonstrates an effort to understand the phenomenon under question comprehensively. Triangulation is an option for validation rather than a tool or approach. Multiple points of view, procedures, empirical materials, and observations are thus best understood in a study as a method that can add rigour, breadth, and depth to any investigation (Flick, 1992). The researcher used a variety of factors to ensure quality:

- *Width* refers to the interview quality or observation and the proposed interpretation or analysis. Furthermore, the length includes several quotation citations and recommendations of different views to allow the reader to decide and analyse the information. (Campbell and Machado, 2013).

- *Coherence:* How the many components of the interpretation interact to provide a complete and relevant perception (Lieblich *et al*., 1998). It is beneficial to increase coherence by using the interviewee's voice to ground the study (Campbell and Machado, 2013).
- *Insightfulness*: This refers to how researching a case might help the reader better understand and gain insight into their circumstance (Campbell and Machado, 2013).
- Parsimony is characterised as the ability to provide an analysis based on a small set of concepts and elegance or aesthetic appeal. The study is a reflection of many people's "truths" by delivering a universal story with a single storyline and "weaving in" various theoretical perspectives (Campbell and Machado, 2013).

There were several critical aspects to consider to ensure the quality of research (Radhakrishna *et al*., 2012):

- *Validity:* This phrase refers to the "closeness between the values presented and the true values" (OECD, 2011). The meticulous construction of the questionnaire determines reality. The argument for construct, content, and face validity is made after a thorough analysis of previous studies, ongoing evaluation by an expert panel, and the completion of a field test (Guba, 1981).
- *Reliability:* This is the degree to which measurements are similar (consistent) over multiple measures. The questionnaire's reliability is demonstrated by its careful wording, pilot testing with non-sample people, and high response rate.
- *Data objectivity*: The fact that findings are drawn using statistically valid approaches. Objectivity was demonstrated by applying appropriate statistical procedures and outcomes and thoroughly examining assumptions/hypotheses/objectives/research questions.
- *Integrity:* Is concerned with reducing errors in the data collection, recording, and analysis. Integrity can be improved by adequately training personnel who collect data and double-checking that the data has been recorded correctly.
- *Generalisability:* Refers to sound sampling processes that provide a sample typical of the population on essential variables (Guba and Lincoln, 1983) and non-respondent follow-up (Guba, 1981).
- *Completeness:* The way missing values in a dataset are handled is called completeness. When data goes missing at random, it is because of uncontrollable external events, whereas data that is not lost at random cannot be recovered due to known external events. To better understand the study's limits and generalisability, non-random data must be examined during data analysis.

- *Relevance:* Refers to how essential data is to users and their needs (OECD, 2011). Thorough literature reviews and needs assessments are two ways to ensure a high level of relevance.
- *Utility*: Timeliness (data collected on time so that data remains relevant to users), punctuality (data release), and accessibility are all characteristics of utility (ways in which data are made available to the intended users).

Where possible, the researcher considered several of these aspects to augment the rigour of this study, especially validity, reliability, objectivity, integrity, relevance, and utility.

4.10 Ethical considerations

There are two definitions of ethics: moral principles governing or influencing behaviour and the domain of study concerned with moral principles (Soanes and Stevenson, 2004). There are at least four reasons to consider ethical considerations when performing research: 1) its dual potential; 2) institutional review boards' increased attention to the ethical rules that must be observed in research initiatives, and 3) the diverse ethical priorities researchers have in IS. First, in a fundamental study published 30 years ago, Mason (1986) articulated the dual possibilities of information technology. Mason exemplified how information technology may either enrich or undermine human dignity. Information technology (IT) can improve people's lives but also worsen them.

Information technology can improve patient outcomes in medicine, but increasing surveillance may jeopardise privacy. Given the dual potential of information technology, information system scholars must accept responsibility for the social contract that emerges from the technologies we build and implement (Mason, 1986). Secondly, institutional review boards or human subjects' ethical committees must approve studies involving actual people before they can be conducted. Universities and other organisations need to uphold moral standards when doing IS research.

For example, the Academy of Management's ethical code specifies enforceable norms that all academy workers must follow (Meyerson and Townsend, 2004). The Association for Information Systems (AIS) maintains a code of conduct for researchers that contains both mandated and "recommended ethical behaviour" elements (AIS, 2013). Thirdly, design science researchers' ethical concerns must differ from behavioural researchers in IS. The subjects of social scientists' and behavioural researchers' research are highly valued in most professions. According to the

American Anthropological Association's ethical rules, when there is a conflict of ethical commitments for various stakeholders, anthropologists, like most other social and behavioural researchers, must prioritise the people being researched.

However, engineers and computer scientists do not prioritise the people they are looking for; instead, the public is an engineer's most important ethical duty. The public interest is the first premise of the British Computer Society's Code of Conduct (Venable and Baskerville, 2012). The Association for Computing Machinery (ACM) Code of Ethics' fundamental premise is that members must contribute to society and human welfare. This principle emphasises the need to maintain fundamental human rights and appreciate the diversity of all cultures, which is concerned with everyone's quality of life. One of computer professionals' top concerns is reducing the negative consequences of computing systems, such as health and safety risks. While planning or implementing procedures, computing experts must aim to guarantee that their work products are used in socially acceptable ways, fit societal norms, and have no negative environmental repercussions.

Privacy, accuracy, property, and accessibility are four broad ethical dilemmas essential in the information age and IS (Mason, 1986). Personal information that must be shared with others, as well as the circumstances under which that information must be disclosed and the measures in place to protect that information, is called privacy. Who is responsible for the information's accuracy, validity, and integrity is referred to as accuracy. Data ownership is referred to as property. The term "accessibility" refers to a person's or organisation's ability to access information as well as the security measures in place to protect that access. Table 4.5 summarises some of the ethical difficulties that arise when these four issues mentioned by Mason (1986) are considered.

Table 4.5: Ethical questions for Design Science researchers (Mason, 1986)

Ethical issue	Questions for design science researchers
Privacy	Is the proposed new system going to put your privacy at risk? What are privacy safeguards in place? Who is responsible for protecting personal information? Is the suggested new design in any way a threat to privacy? What are security measures in place to secure personal data? Who is in charge of safeguarding personal information?
Accuracy	Is the data being collected accurately? Who is accountable for maintaining order?
Property	Who is the owner of the artefact's intellectual property rights? Who owns the data that has been gathered?

Access	Who has access to what information? Are some groups of people going to be exempted? Who is accountable for ensuring that the appropriate individuals access the appropriate information?

Methodology-specific ethical considerations: The researcher considered several ethical concerns associated with the methodology chosen. Since this research asks for questions related to risk assessment in big data and cloud computing contexts, ethical issues were not a significant concern. However, since this study involved participants from different backgrounds and experiences, the researcher considered the following elements to guarantee the success of this research:

- Gaining the required written or verbal permission to conduct this study. Written authorisation is privileged.
- Making sure that involvement was entirely optional. Participants are not obligated to contribute information to this study and have the option to withdraw at any time if they believe it is not necessary.
- Ensuring the data's confidentiality at all times. This includes appropriately storing and managing data. It also entails erasing the raw data once the research has been completed satisfactorily.
- Educating participants on the potential dangers of participation in this study.

This research did not involve sensitive material; therefore, the odds were minimal or non-existent. Following the NWU ethical process:

- The researcher adhered to the NWU research ethics policy found at http://www.nwu.ac.za/content/policy rules;
- The researcher signed the code of conduct statement; and
- An NWU Ethics approval letter of the study was issued on 4 January 2021, with a "no risk category" under ethics number 01558-20-A9 and signed by Prof. Roelof Burger (Chairperson Faculty of Natural and Agricultural Sciences Ethics Committee).

This research followed the plan described above by first specifying the domain; this step consists of deciding the unit, the analysis, and the topic. Second, it gathered information. This step collected information about the different risk assessment practices in big data and cloud computing. Third, the study critiqued the evidence; the researcher assessed what the collected data was worth. Fourth, the study determined patterns using Atlas.ti software. The next task was to organise the

information, revealing interesting patterns about the issue analysed (i.e., risk factors in big data and cloud computing settings). Finally, the researcher applied the above quality criteria to add rigour to the study. Overall, it can be said that in the study, ethical considerations abided within the research frontiers. The confidentiality of the personal information of participants was duly maintained, and the information collected was collected in the same form as received. There was no tampering or misinterpretation of the information, and the researcher maintained the integrity of the research significantly.

Further, the participants were informed before the research, and suitable time was taken from them to conduct the research. The respondents were also informed about the purpose of the research, and there was clear communication in the study. Therefore, the research's moral assurance and trustworthiness make it ethically sound and apt.

4.11 Summary

A paradigm is a shared set of viewpoints that binds the work of a group of theorists together. (Burrell & Morgan, 1981). The interpretive paradigm was used to frame this study; it also used elements from pragmatists and functionalists to generate constructive and valuable knowledge (Goldkuhl, 2012). According to (Orlikowski and Baroudi, 1991), interpretive information systems research proposes that the social environment (social interactions, organisations, and labour divisions) is not "given" from an ontological standpoint (1991). On the other side, humans construct and reinforce the world through their activities and interactions.

In an ever-changing world, this research employed pragmatism, which focuses on developing explanations for the status quo, social order, social integration, consensus, need satisfaction, and rational choice, as well as functionalism, which focuses on developing justifications for the status quo, social order, social integration, agreement, need satisfaction, and sensible choice.

This study used mixed methods with qualitative data as the main data. Therefore, the study followed inductive reasoning, focusing on extracting meaning from facts and information (specific-to-general). This study also employed a four-step procedure to answer the research questions and achieve the objectives: define the domain, obtain information, review the evidence, and determine patterns. Grounded theory enhanced the pattern-finding process. To improve the quality and rigour of this study, the researcher used Atlas.ti software and Excel, followed by triangulation.

CHAPTER 5: RESULTS

5.1 Introduction

This chapter presents the results intended to provide a comparative analysis of big data risk assessment techniques based on cloud computing technologies in four countries: Canada, Jordan, South Africa (SA), and the United Kingdom (UK). The findings are presented with quantitative descriptive statistics and then supplemented with qualitative data analysis. The quantitative results draw from the descriptive statistics, frequencies, and the graphical representation of the data based on structured questionnaires administered in four countries to respondents who are industry experts in big data and cloud systems. The qualitative findings draw from the thematic analysis of these key informant interviews analysed using Atlas.ti software.

5.2 Quantitative data based findings:

Risk assessment is an important element in current times due to the rapid adaption of virtual platforms and cloud services. Risk assessment systems vary from country to country, and different approaches to risk assessment depend upon the nation. The study aimed at doing a comparative study of risk assessment methods in different nations and the findings based on questionnaire results collected from different nations have been laid down. The study interpreted the research questions and laid down the critical understanding of processes implemented in the mentioned four nations concerning RA methods in big data using cloud platforms. The quantitative analysis enabled a brief understanding of the risk assessment methods in four nations Canada, Jordan, SA, and the UK. Most of the SA respondents had limited knowledge of big data risk assessment methods. However, the practical application of cloud services is very high in South Africa. Participants possess knowledge of risk assessment methods and an in-house team of professionals for big data management in SA. South Africa runs on a traditional storage medium, and big data technologies are not run in the country's organisations.

Further, the UK revealed practical and ample knowledge of RA methods in big data. Maximum respondents also revealed ample practical knowledge and application of cloud computing services. Further, experts and professionals are present in the nation to apply risk assessment methods. There is also a dedicated team to handle big data risk assessments, and the presence of experts and outsourcing teams creates a holistic approach. In the UK, cloud based storage is in use, and big data technologies are fully implemented for proper assessment and evaluation of the operational processes. Canada showed optimum to low knowledge of RA methods in big data. In Jordan, most

participants had limited and intermediate knowledge of big data. In Canada, cloud computing is also implemented, and there is a presence of expert level knowledge on risk assessment methods. Countries like the UK have a dedicated team of professionals, experts, and outsourced staff for effective network management. Canada uses cloud services in the majority, and traditional storage is losing its space in the nation.

In Jordan, most respondents had medium to low knowledge about big data; the nation is aware and advanced in risk assessment methods and cloud computing services. Further, Jordan has a dedicated team of professionals, experts and outsourced staff to validate their network management. Grounded theory was used to integrate the quantitative and qualitative data, and the findings (quantitative and qualitative) were triangulated along with previous literature. The researcher was able to compare the different constructs and reach the study's aim, i.e. comparative study on risk assessment methods used in big data using the cloud in four nations specified. Data triangulation not only linked and created an integrated construct, but it helped to enhance the quality and vigour of the findings. Thus, quantitative data findings have been elaborated country-wise, and graphical representations have been added.

5.3 Level of knowledge of big data

Big data refers to massive data sets computationally analysed to discover patterns and trends related to data components. The below section reveals respondents' knowledge collected through quantitative analysis based on a questionnaire on various attributes relevant to the study, such as big data, cloud computing, and risk assessment methods. The observations from their responses have been collected and opined below. Figure 5.1 indicates the level of knowledge of big data. There was no minimum amount of data required to be classified as big data for this study, so long as there was enough to draw strong conclusions.

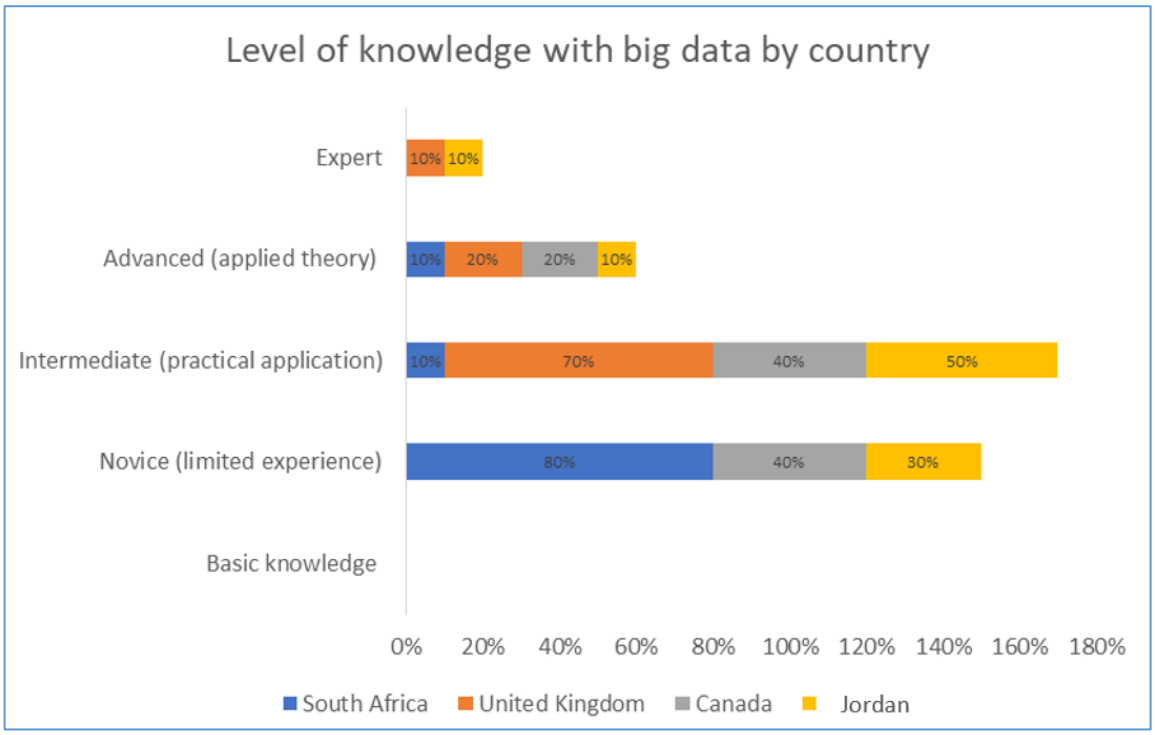

Figure 5.1: Level of knowledge with big data

The majority of respondents reflected novice and intermediate levels of knowledge of big data. South Africa seems to have the majority of respondents (80%) who are novices in knowledge about big data. The UK has the most (70%) practical application (intermediate) knowledge of big data. Only Jordan and the United Kingdom have expert knowledge about big data, while Canada is spread between novice and advanced expertise.

5.4 Level of knowledge of cloud computing

As defined in this study, cloud computing delivers computer services like servers, storage, databases, networking, software, data analytics, and artificial intelligence through the Internet (the cloud) to facilitate faster innovation, more flexible resource allocation, and cost savings. Figure 5.2 shows the knowledge of cloud computing regarding how respondents leverage it to transfer information and computing from home and office computers into external data hubs. Through cloud computing, people can use computing services when and where they need them via the Internet, facilitating the creation of applications in real-time and faster implementation of services.

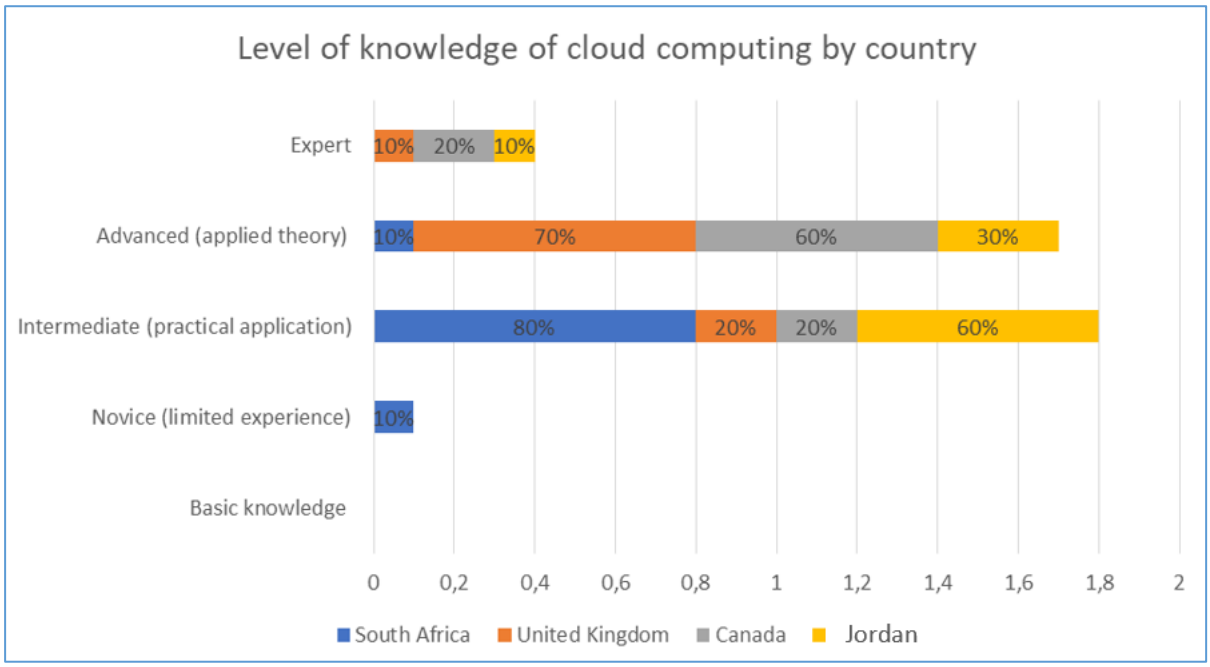

Figure 5.2: Level of knowledge of cloud computing

The majority of respondents reflected the intermediate and advanced levels of knowledge of cloud computing. Only South Africans exhibit a novice level of expertise in this aspect (10%), and most respondents (80%) have intermediate knowledge of cloud computing. The United Kingdom has the biggest number (70%) that have an advanced understanding of cloud computing. However, Canada fares better than all countries in terms of experts in this aspect. Jordan and the United Kingdom also reflect as experts. However, South Africa does not appear in this category.

5.5 Level of knowledge with risk assessment

Figure 5.3 shows the level of knowledge with risk assessment by country. In Figure 5.3, the risk is the probability of a threat of an entity taking advantage of a weakness, which usually impacts businesses and their processes and services. Statisticians, economists, systems analysts, biologists, operations researchers, and others rely on risk assessment. There are two sorts of RA considered in this study. These are quantitative RA, which is objective and quantifiable, and qualitative RA, which is subjective and based on expertise and know-how.

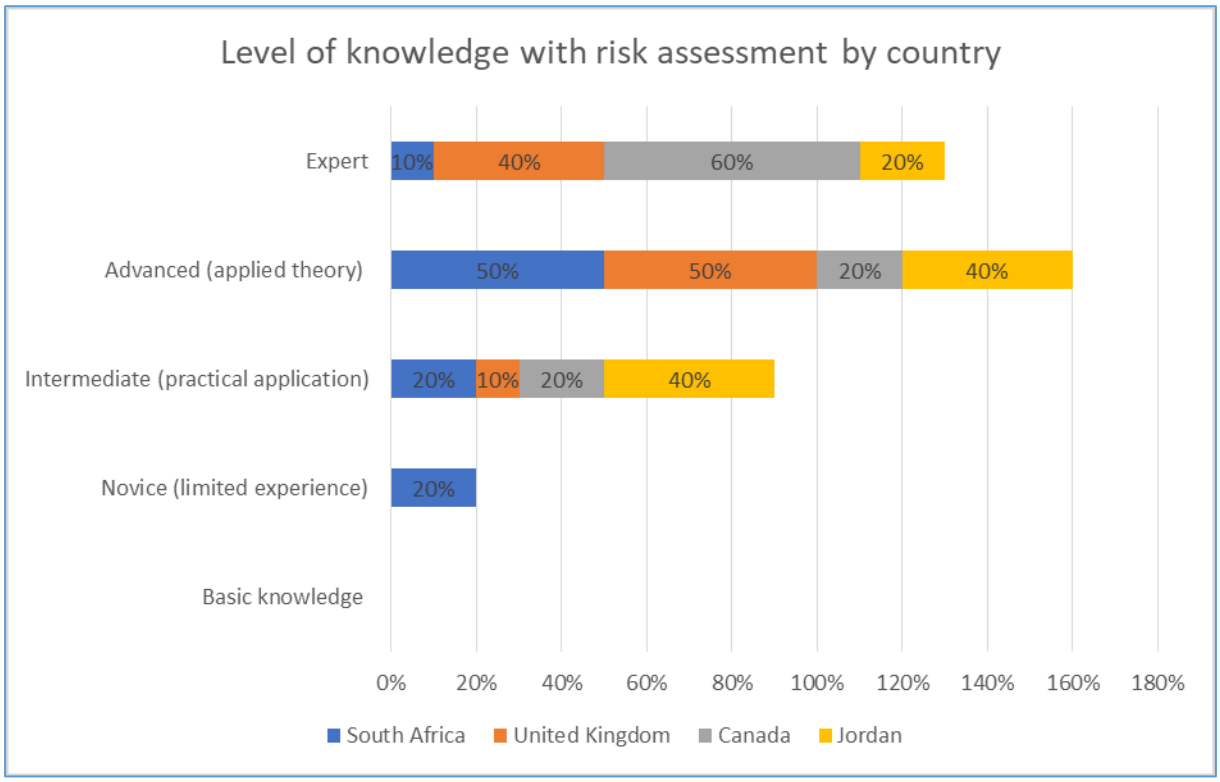

Figure 5.3: Level of knowledge with risk assessment

South Africa is the only country with a challenge with risk assessment, where 20% of the respondents indicate that they are a novice in this aspect. However, most have "advanced to expert" knowledge about risk assessment. Both South Africa and the UK have 50% of their respondents indicating that they have advanced knowledge. Canada tops as an expert in risk assessment, with 60% of the respondents indicating the affirmative, followed by the United Kingdom. Jordan respondents are distributed across intermediate, advanced, and expert levels.

5.6 Management of network by country

Network management refers to the procedures, tools, and software required to administer, maintain, and manage network infrastructure. Simply stated, network management is keeping a company's network healthy. Figure 5.4 demonstrates network management by country.

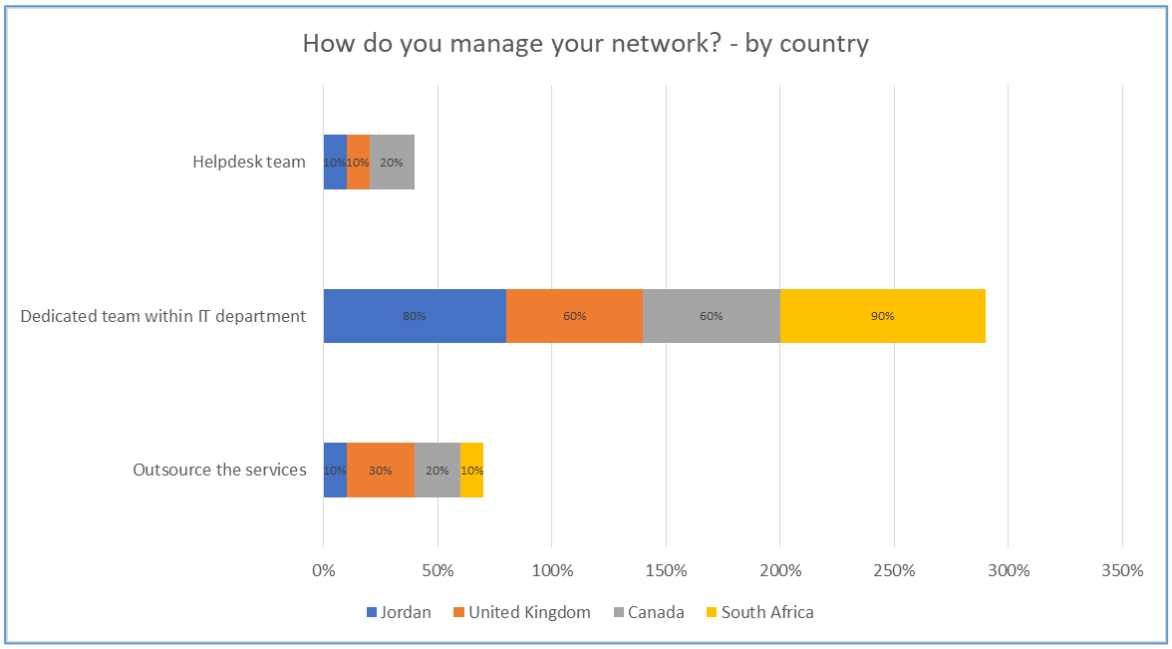

Figure 5.4: Management of network by country

Figure 5.4 indicates that most countries have a dedicated team within their IT departments to manage their networks. South African respondents dominate this category (90%), followed by Jordan at 80% and then the United Kingdom and Canada, each at 60%. The results indicate that help-desk services are not common, particularly in South Africa. Outsourcing services to manage the network is more dominant in the United Kingdom and Canada.

5.7 How countries run big data applications

Figure 5.5 shows how respondents manage data applications in their respective countries. As mentioned in Chapter 1, many big data applications need a set of new platforms in the cloud. In contrast, the cloud's existing security tools and methods may not properly work for these implementations. Therefore, improved tools for safety must be developed to work with these new platforms. These security tools involve encrypting data, authenticating processes, controlling access, detecting interference, and constantly monitoring and tracking events. In addition, cloud storage allows for dynamic file sharing because anyone with Internet connectivity can access it. To share data, traditional storage requires physical drives. In addition, traditional storage is difficult to manage since it requires organisations to run through maintenance tools manually. Cloud storage, on the other hand, is simple to maintain.

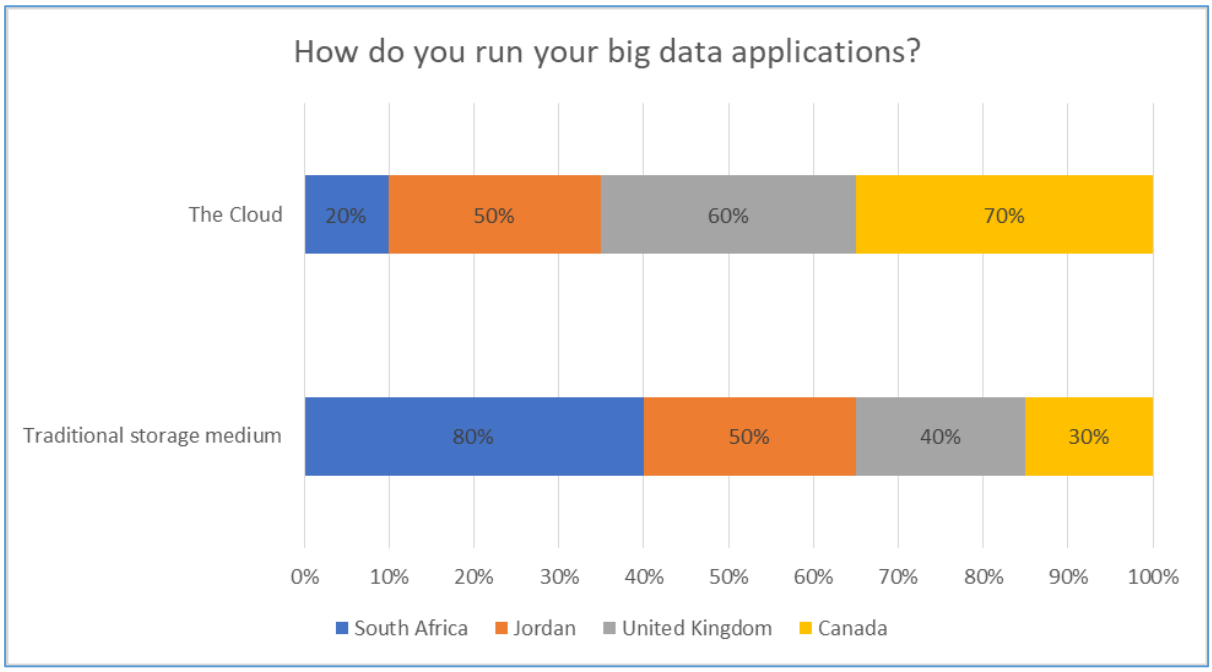

Figure 5.5: Management of network by country

Figure 5.5 shows that South Africa runs its big data applications mainly on the traditional storage system (80%), followed by Jordan (50%). Canada is top in managing the big data applications on the cloud (70%), followed by the United Kingdom. Only 20% of the respondents indicated that they use the cloud in South Africa.

5.8 Implementation of big data technologies within organisations

Figure 5.6 shows the proportionate implementation of the big data technologies in the four countries under study. Big data implementation helps firms across industries adopt a near-scientific approach to what was previously based on guessing. For example, implementing big data technologies can answer the following questions: *What do customers desire to do? How effective is a given course of action? How efficient is a particular marketing project, ad campaign, or staff?*

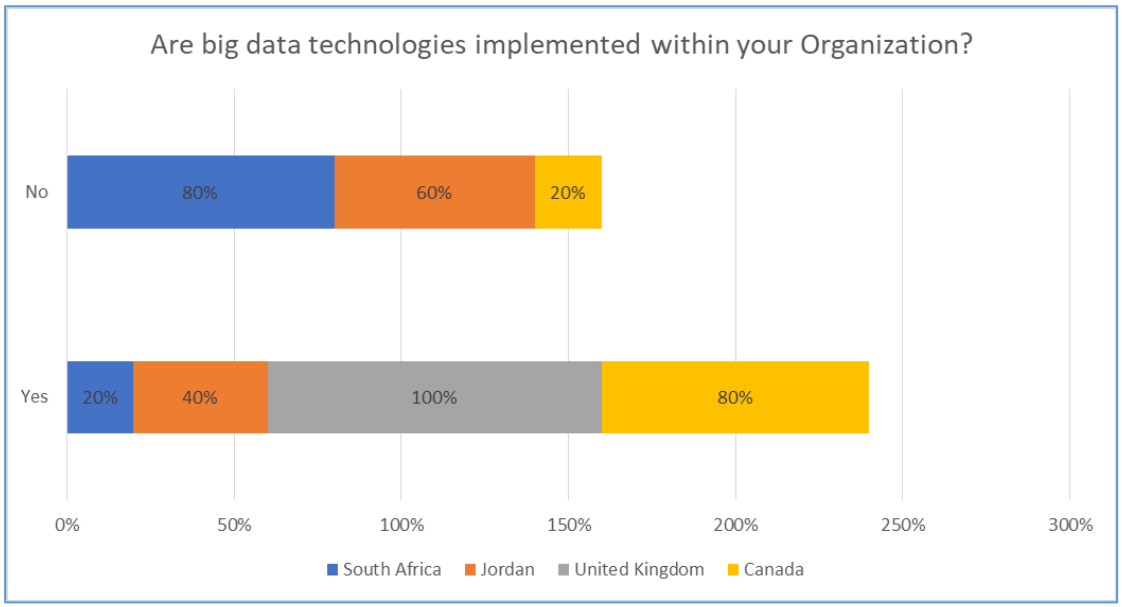

Figure 5.6: Implementation of big data technologies within organisations

Most countries agree that big data technologies are being implemented in their organisations. For example, all respondents in the UK consented to this aspect, followed by Canada. However, 80% of South Africans agree that they are not implementing big data technologies in their organisations, followed by Jordan (60%).

Despite the low adoption of big data in South Africa, many respondents support this technology, as it helps organisations expand. The respondents believe that every improvement to technology improves the employees' ability to work. Big data technologies enhance operations and provide better customer service, giving a better market understanding. In South Africa, big data technology is viewed as helping make smarter decisions based on analytics, helping to know what customers are looking for, helping risk management models work better, and creating smarter plans and policies.

Despite the above, South Africa has the highest number of respondents, indicating that big data was not implemented. Several reasons why big data technologies are not being implemented within South African organisations vary in terms of costs, security issues, and a lack of information. Respondents indicate insufficient information on implementing this technology, so more research is needed. Some respondents highlight employees' lack of required skills and the budget to hire and adapt. In South Africa, the security and privacy issues associated with implementing big data are the main concern, and organisations prefer to learn more about the best practices before implementing such technology. Thus, there is a feeling among respondents that big data will not add value to the current situation, and they will keep evaluating the necessity on an ongoing basis.

Jordan also has many respondents who indicate that big data technologies are not being implemented. For costs and a lack of a qualified workforce, big data technologies are not being implemented within Jordan's organisations. Like in South Africa, in Jordan, there is an indication of the need for more assessment on how big data technology will benefit companies. However, the support for big data technology in Jordan is huge. Big data technology can boost the client base and save time. It also leads to cost savings and is great for customer insight, directing organisations to focus and target customers, resulting in more sales and profits.

In Canada, most respondents indicate that big data is being implemented. However, the few that indicate otherwise raise the issue of costs and the need to study the technology further. However, the perception in Canada is that technology is safe, convenient, and reliable. Big data implementation in the United Kingdom is 100%, and no challenges are raised about the technology.

5.9 Qualitative data based findings:

The study incorporated thematic analysis to interview and analyse the perception of respondents based on different themes identified, such as technology, control, availability, infrastructure, and cost. The qualitative data findings highlighted the importance and advantage in countries and the current state of risk assessment in big data using the cloud in these four nations. Some advantages of the developed nations such as the UK and Canada are the infrastructure support these nations possess and the qualified professionals addressing the IT support dynamics to boost the cloud based set up. Some disadvantages of these developed western nations include extremely competent hijackers and a series of malevolent activities, making the entire cloud system vulnerable. Developing nations such as South Africa and Jordan faced the biggest challenge concerning the cost of implementation and maintaining technology.

Furthermore, there are challenges concerning the shortage of skilled workforce, and therefore, the respondents felt the need for improvement in infrastructure and workforce in South Africa and Jordan. On the other hand, in UK and Canada, the respondents felt the need to strictly implement the legal framework to avoid frauds, hijacking, and other malpractices arising from loopholes. The detailed results concerning thematic analysis and highlights from respondents' interviews have been laid below.

5.10 Advantages and disadvantages of implementing big data

Several themes were highlighted concerning the advantages and disadvantages of using big data.

5.10.1 Theme 1: The technology itself

Theme 1 is more prevalent in Jordan. With big data technologies such as Hadoop and Spark, Jordanian businesses can reduce the cost of storing and analysing massive amounts of data. Big data technology might aid in the discovery of cost-effective and efficient corporate strategies by increasing the accuracy of data, empowering a new generation of employees, reducing costs, and managing online reputation.

The disadvantages noted are related to data quality. For example, one respondent highlights that analysts need to verify that the data they are working with is correct, relevant, and in the right format for analysis. In addition, the technology is still suffering from challenges such as user-level execution that only exists in select channels; user-level results that cannot be presented directly; and user data that is not suited for producing learnings. Jordan also had a special feeling that there is no urgent need for big data technology as the current legacy systems achieve the desired functionalities without using big data technologies. Therefore, decision-makers do not need to pay more money to replace the old technology and replace it with big data technologies.

5.10.2 Theme 2: Control

Theme two is prevalent in the United Kingdom, South Africa, and Jordan

In the UK, the advantages of big data technologies are related to reliability, better data control, high availability, automated updates, easy data backup and recovery, and scalability. In South Africa, the advantage noted is data processing on a large scale.

The disadvantages noted in the UK are limited control of infrastructure, limited flexibility, ongoing costs and security. In SA, it is the limited control of data and security issues. In Jordan, the disadvantage of big data technology is linked to the hidden high cost of the cloud infrastructure in the long term. Also, the limitation of control on systems.

5.10.3 Theme 3: Availability

In Jordan, the advantages are the flexibility and the availability of services. Similarly, in Canada, big data technology ensures the availability of services and flexibility. The technology will guarantee technological independence and flexibility in addition to easy access and flexibility in work handling. In the UK, the advantages are related to ease of use and cost-effectiveness; also, respondents in the UK highlight flexibility in work handling.

5.10.4 Theme 4: Infrastructure

In the UK, advantages relate to reliability, better data control, high availability, automated updates, easy data backup and recovery, and scalability. In addition, respondents indicate that the technology will guarantee technological independence and flexibility. However, some disadvantages are noted: high cost, lack of required experience, risk of being hacked, complicated infrastructure to host the data, and limited control of limited infrastructure flexibility.

One UK respondent had this to say:

> *"Big data is still being used to improve services and build new ones. Using big data, companies can find out what their customers want. In today's market, a company's survival depends on instinct. Thanks to the volumes of available information, thanks to organisations may now develop procedures to track consumer feedback, product success, and what their rivals are doing. All of that will allow companies to access more customers and have a stronger presence in the market."*

South African respondents indicate that the carbon footprint is greatly reduced and the day-to-day costs of the electrical bills; and the staff who need to monitor and maintain that infrastructure if we had it on site. More storage on files is highlighted, keeping the company information safe and in one place and processing data on a large scale. The disadvantage may be limited control of data and security issues. The disadvantages mentioned by South African respondents are that security is one of the biggest concerns as they rely on the user knowledge of the systems in place. When data storage experiences a problem, all the company information is at risk.

Advantages in Jordan are flexibility and the availability of services. The disadvantages are the long-term high cost of the cloud infrastructure and the limitation of control on their systems.

In Canada, respondents raise the positives regarding data availability and flexibility of services. Modern tools such as Microsoft Power BI are available for data management and analysis, which will help improve results. Respondents also note that the technology will guarantee technological independence and flexibility in addition to easy access and flexibility in work handling. The flexibility will gradually save organisations on costs. The highlighted disadvantages in Canada concern data security related to information hijacking. Also, big data technology is perceived as costly if the organisation has no guaranteed outcome.

5.10.5 Theme 5: Cost

This theme is dominated by the high and prohibitive costs of implementing big data. In the UK, the disadvantages are high cost and a lack of required experience, linked to the complicated infrastructure to host the data. Respondents in the UK also highlight that it takes a long period to ensure full implementation of the technology, thereby raising the costs of setting it up. In South Africa and Jordan, the costs of implementing and finding new skilled staff to take on the task are also presented as concerns. In Jordan, this theme is seen as experts and analysts spending time to verify that the data they are working with is correct, relevant, and in the right format for analysis. In Canada, the challenges related to finding qualified people to work in the field, the need for new hardware, which will be an added cost, and cybersecurity are noted. Respondents in Canada also raise the matters of the hidden high cost of the cloud infrastructure in the long term and the limitation of control on their systems.

5.11 Challenges experienced when running big data applications in the cloud

Figure 5.7 shows the respondents' plan to run their big data applications on a traditional storage medium or in the cloud.

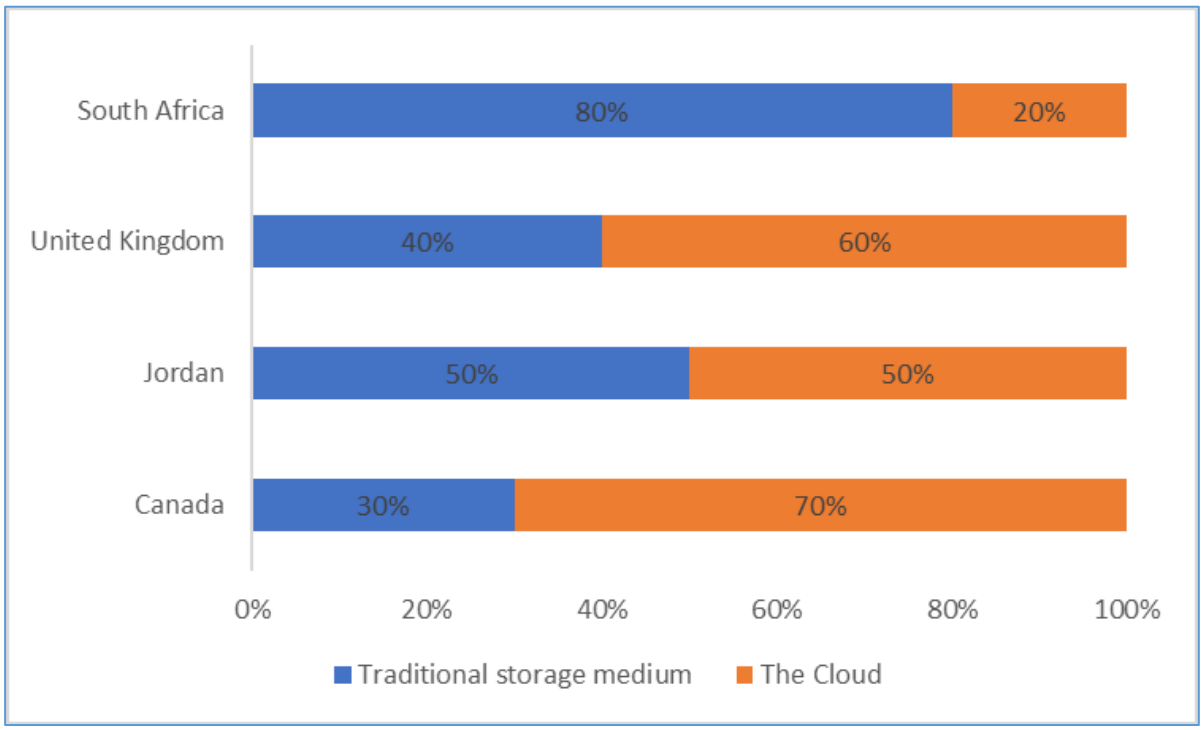

Figure 5.7: Plan to run big data applications on a traditional storage medium or in the cloud

South Africa is the only country with most respondents planning to run their data on traditional storage (80%). In Jordan, the respondents are split equally concerning plans to run big data applications on traditional storage or cloud systems. Consequently, Canada is the only country with most respondents planning to run their data on cloud storage (70%), while the UK has 60% of its respondents planning to run big data on the cloud system.

5.11.1 Secondary data statistics in support of findings:

In South Africa, 90% of African businesses operate without the necessary cybersecurity protocol and post-Covid, the cybercrime ecosystem is accelerated (Interpol, 2021). Nations need to perform risk assessments and evaluations so that high, increasing dependence on online platforms can be safeguarded. Covid-19 dramatically impacted data traffic and global internet bandwidth to 35% in 20202, the largest one-year increase after 2013 (UNCTAD, 2021). An important element of cloud usage is migration to hybrid clouds; therefore, in South Africa, enhanced ICT infrastructure is being integrated (ITU, 2021). In Canada, 88% of Canadian organisations considered cloud platforms and advanced technology to be the future of new business models (Deloitte, 2020). Therefore, cloud-based services are widespread globally, and South Africa is increasing its reach toward big data using the cloud. For the smooth functioning of big data using cloud platforms in global organisations, it is important to address the challenges

Table 5.1 shows the challenges countries face in running big data applications.

Table 5.1: Challenges experienced when running big data applications in the cloud

Country	Experienced when running big data applications in the cloud
Canada	• Scalability • Slow files download • Frequent downtime and many errors – overloaded with error reporting • Data security. Different types of security breaches include malware attacks, cross-site scripting attacks, and phishing attacks.
United Kingdom	• Storage and processing of data are costly. • Security control is too complex. • Expensive cloud storage. • Cost for small businesses and start-ups.
Jordan	• Safety measures and data must frequently be double-checked. • Cost of setting up and maintaining the technology and cloud service charges. • Cloud service reliability, efficiency, and security since the organisations need to integrate parts of their sensitive data into the bigger data. • Huge and complex hardware infrastructure is required to store and analyse a significant volume of data.

	• Big data can only be useful if an information management method is applied to provide high-quality data, which ensures data accuracy and availability. • Too much data means a higher risk for cybersecurity and threats.
South Africa	• Achieving a reliable amount of bandwidth for the operational business requirements. • Companies cannot afford it as it is still too expensive. • Finding trained and skilled individuals with the right knowledge set is a challenge.

5.12 Implementation of big data technologies within organisations

Figure 5.8 shows the degree to which cloud computing is implemented in organisations across all four countries.

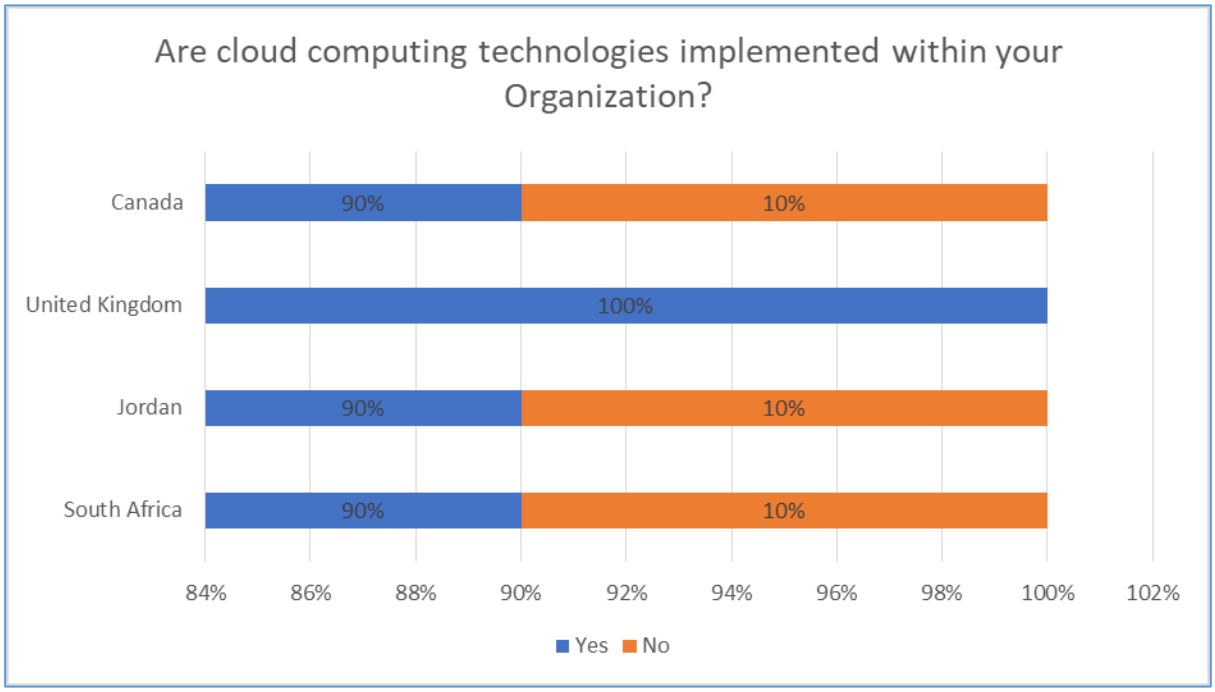

Figure 5.8: Implementation of cloud computing in organisations across all four countries

In Jordan, 90% of respondents confirm that cloud computing is implemented in their organisations. The 10% did not indicate a high cost to implement it, a lack of qualified individuals, and a need for analysis to understand the technology impacts. However, there is an acknowledgement of cloud technology provides better file storage and sharing.

In Canada, 90% of respondents confirm that cloud computing is implemented in their organisations. The 10% did not confirm that they plan to implement the cloud system soon as they are in the early stages of studying it. In Canada, the support for cloud technology is premised on rapid growth

worldwide. Therefore, organisations must cope and prepare for big data handling to accommodate high customer demand.

In South Africa, 90% of respondents confirm that cloud computing is implemented in their organisations. The 10% who did not indicate that they do not have an idea about the decision on why they are not yet adopting this technology. However, if implemented, there is a sense that this technology would improve the overall reliability and uptime.

In the United Kingdom, 100% of respondents indicate that cloud technology is being implemented.

Table 5.2 shows the advantages and disadvantages of implementing cloud computing technologies across all four countries.

Table 5.2: Advantages and disadvantages of implementing cloud computing technologies

Country	Advantages	Disadvantages
Canada	• Management of infrastructure is simplified. • Easy accessibility and management. • Data is safer from loss. • Speed. • Can use PC to access files from anywhere during the pandemic. • Less overhead, scalability, agility, application, and security. • Lower costs in terms of hardware and software.	• Any device can access files from anywhere, which can be considered a risk. • Reliance on internet speed and reliable connectivity. • The challenge of transferring data from on-premises to the cloud due to network bandwidth, hence the reliance on service providers. • Security threat and high training needs.
United Kingdom	• Easier to manage - better scalability, availability and improved turnaround. • Easy workflow, unlimited storage space, efficient and reliable. • There is less need to buy expensive hardware. • Allows customer data analysis, provides better service and support and improves operation and processes. • Speed in uplifting and access across the globe.	• The downtime, since the service is Internet-based, an outage can happen at any time and cause serious issues, and IT is expensive to deploy. • Costly to implement. • Lack of required experience. • Risk of being hacked. • Complicated infrastructure to host the data.

Jordan	• It helps higher education institutions to benefit from computing resources with lower costs. • Cost reduction. • Cloud storage is better to manage and control. • Minimising costs or expenses, allowing files and service access from anywhere at any time. • Enhance security and privacy. • Enhance institutional productivity and make the educational process more efficient. • Boosts collaborative work. • Unlimited storage capacity. • Available all the time. • Easy access and management • Allow hybrid working and save resources, hardware, licensing fees, and DC space. • Flexibility.	• Not all systems can be implemented on the cloud, as high confidential data should be stored and processed on resources owned by the institution. • High costs and harder to control. • If it breaks down, the downtime takes longer to recover. • There is too much trust in service providers, whose longevity cannot be guaranteed. • Suffers from lack of control dependency on network performance. • Lack of required experience • Risk of being hacked. • Complicated infrastructure to host the data. • Service is Internet-based; outages can happen anytime and cause serious issues. • Utilising capacity is low.
South Africa	• More storage resources and backup in case main systems crash or become corrupted. • Turnaround time will improve. • Easy access to files and information is safe. • Files are safe and have more accurate storage efficiency; they can store large amounts of data. • Operational costs are low. • Files can be backed up, and no files get lost. • Cost savings and high-security level. • Fast deployment.	• The cost is too high • Training is unavailable. • Data protection and security - data loss and theft. • Files are hard to sort. • Most of the components are not fully under their control.

5.13 Policies in place in case of termination or transferring data to the cloud service

Figure 5.9 shows, by country, whether or not specific arrangements (policies) are in place concerning organisations' information if they want to terminate or transfer to the cloud service.

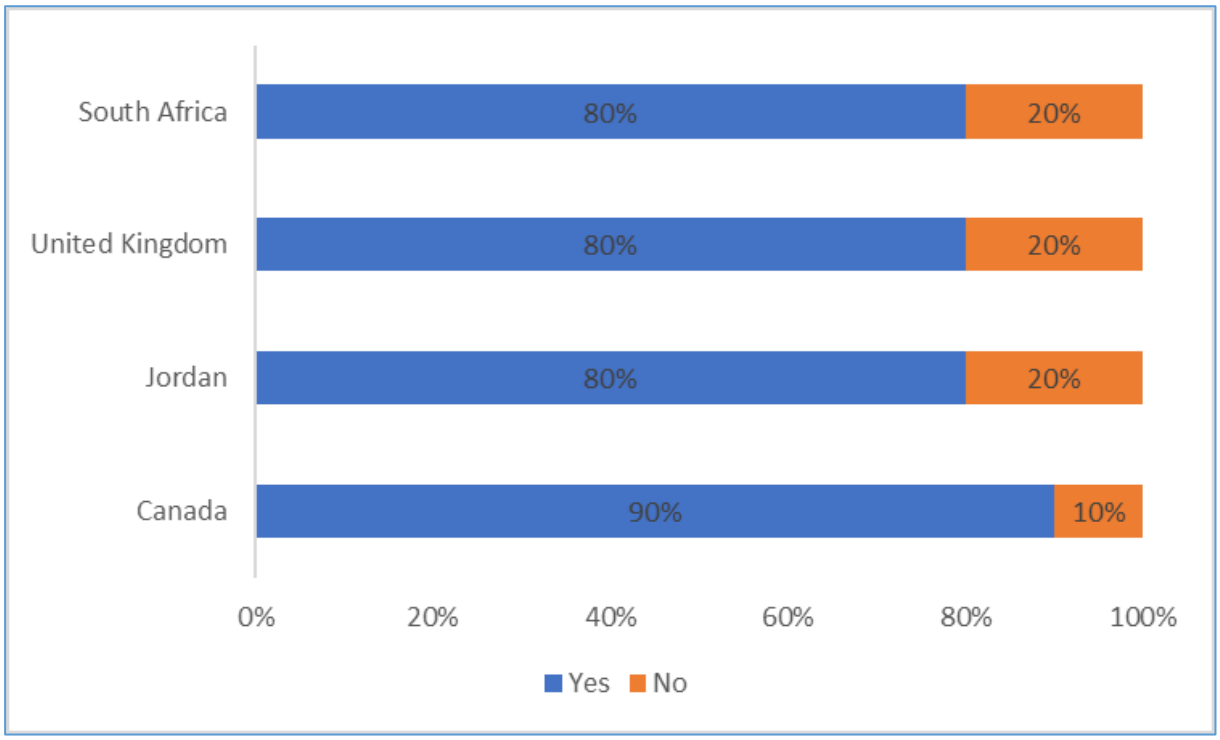

Figure 5.9: Implementation of cloud computing in organisations across all four countries

Figure 5.9 shows that in Canada (90%), the respondents acknowledge the existence of specific arrangements (policies) concerning organisations' information in case they want to terminate or transfer to the cloud service. However, the rest of the other countries (80%) confirmed that they implement cloud computing in organisations, and only 20% did not.

In Canada, when respondents were asked why there are no specific arrangements (policies) in place concerning information in case the organisation wanted to terminate or transfer to the cloud service, responses show that they are in the process of implementing the technology and covering the risk based on a real-life scenario. In South Africa, respondents also indicated that they are still struggling with a clear Information Governance policy in the Governance Risk and Compliance framework universe. South Africa has refrained from moving potential sensitive data to the cloud. South African respondents also indicate that more assessment is needed in this regard. In Jordan, respondents suggest that a plan will be made once costs become worth migration. In the UK, respondents indicate that the plan to put policies in place is still on the cards, but it will not be rushed due to recent challenges, mainly those related to the financial situation due to the COVID-19 pandemic. The respondents in the UK indicate that it will be costly to adopt policies now, but again, it is something under consideration and will be added as soon as they believe it is time.

5.14 Do you have a clear risk assessment plan?

Respondents in South Africa indicate that they have established a well-managed risk management process and system. It forms part of the organisation's overarching GRC (Governance Risk and Compliance framework). It also integrates very closely with a cyber-security strategy with a risk-based approach.

In the United Kingdom, respondents refer to a business recovery plan that includes data and network backup and recovery aspects, which are regularly tested. The plan deals with the following aspects related to big data and cloud computing: applications, data, networks, and service providers. The plan reviews regular reviews and discussions with staff and providers for updates and enhancements. The respondents who are not sure still highlight what the risk assessment plan should entail. For example, one respondent highlights that the following steps must be present: establish a method for regularly reviewing, assessing, and understanding big data in cloud computing threats in an organisation's local and global settings; this entails taking a strategic approach to big data in cloud computing risk management and technological upgrades rather than a tactical approach, therefore, acting appropriately. Another respondent emphasises the need to ensure that the underlying data is reliable, well-defined, and easily understood by stakeholders and decision-makers. It must ensure that data meets all critical quality requirements and clearly articulates data governance responsibilities. The strategy must assure continued investment in cloud computing platforms, analytical big data potentials, and aligning these functions with business objectives. Finally, it is critical to constantly invest in cloud computing infrastructure and big analytical data capabilities to ensure they align with business objectives.

In Jordan, respondents indicate a clear administration level as to who has what level of access to prevent people from performing actions they are unfamiliar with or accidentally acting. However, Jordanian respondents point out that the plan is not mature enough, but meanwhile, they have training sessions available when needed to introduce the risk assessment plans or concepts and to assist the level of knowledge that employees have and what they need to add to the plan based on that. There is a need to identify potential risks and then evaluate and continuously monitor them. Another plan in Jordan specifies the action to take in case of a cyber-attack concerning cloud computing. There is also a policy on the confidentiality of big data. Jordanians constantly assure that all data are safe with offline backup with regular testing and monitoring for suspicious actions. Jordan's data source risk is assessed in four categories: data accuracy, data privacy, data protection, and integrity of data.

In Canadian organisations, executive directors define the risk assessment plans, encompassing cloud computing and big data. Like in Jordan and the United Kingdom, the plans identify potential risks and then continuously evaluate them. Other highlighted plans specify the action to take in case of a cyber-attack concerning cloud computing. There is also a policy on the confidentiality of big data. Canadians are constantly assured that all data are safe with offline backup with regular testing and monitoring for suspicious actions. Some plans in Canada are based on data classification and updated based on observations and testing. This will enable ranking and ordering every risk's likelihood, impact, and severity and determining key indicators (KRIs).

5.15 What is covered in your current risk plan?

5.15.1 Canada

In Canada, respondents point out the following themes in terms of what is covered in their current risk plan: data security, data in transit, data encryption, control systems and data leakage (see Figure 5.10).

Figure 5.10: Themes covered in risk assessment plans in Canada

There are controls to mitigate data leakage and protect data at rest and in transit. Data loss prevention is achieved through education (educating employees about best practices) and robust

cybersecurity controls. The plans state that stored data must be transferred over secure sockets. One respondent had this to say:

> "*When we have data in transit, we encrypt it, so along the way, we make sure that there are no data leaks; if there happens to be any data leakage, we will tree trace the data and encrypt it, therefore protecting it from being leaked again."*

Other respondents indicated that there are controls to mitigate data leakage and protect data at rest and in transit.

In some cases, a compliance team is tasked with this. For example, they use tools in Canada to automate the detection of unintended data access, enforce encryption in transit, implement secure keys, and follow certificate management updates. Respondents cite a policy that defines all types of data they deal with and how data is transmitted electronically. Data encryption is done using Secure Socket Layer (SSL). Another plan highlighted relates to implementing the General Data Protection Regulation (GDPR), processes, protocols, and infrastructure to securely access data. Users must be aware and have some form of training related to cybersecurity and identifying threats. On an ongoing basis, a process is put in place where risks of the inside, outside and third parties are evaluated. Respondents highlight that not all parties might take cybersecurity as seriously. It is, therefore, important to keep evaluating the security posture of all vendors to ensure they are not at risk of suffering a data breach, which will also include monitoring the security posture of all. All network access is monitored to secure all endpoints (firewalls and VPNs).

5.15.2 The United Kingdom

In the United Kingdom, respondents point out the following themes in terms of what is covered in their current risk plan: data security, training, data access, SSL, data transmission, data authentication and end-user (see Figure 5.11).

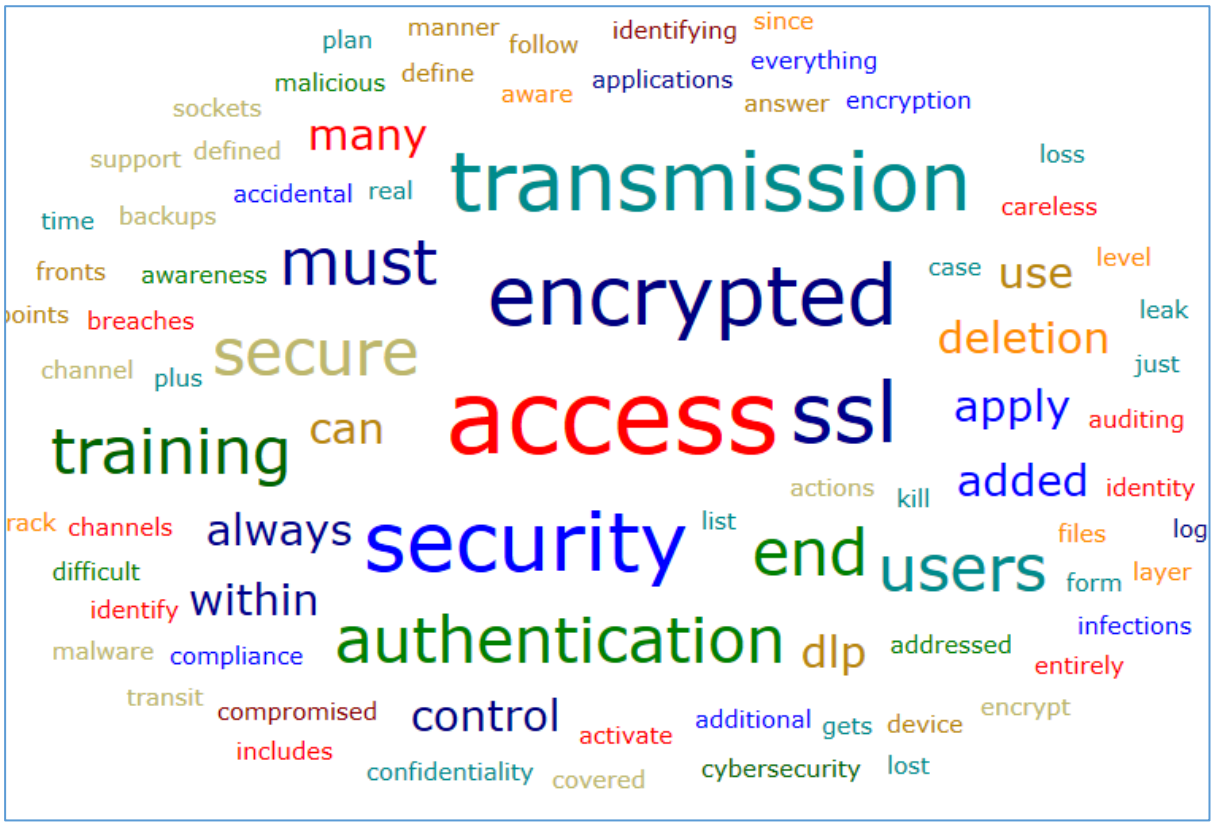

Figure 5.11: Themes covered in risk assessment plans in the United Kingdom

Most risk plans cover everything. In the event of accidental deletion of data, companies have many backups. There is high usage of secure channels for transmission, and a well-defined data level of access exists to secure transmission channels. Respondents provide an example of SSL, security authentication for data transmission. Thus, data in transit is encrypted. Like other countries, the plans also cover security training, awareness sessions and training related to cybersecurity and identifying threats. The United Kingdom companies add TLS (Transport Layer Security) protection when other data confidentiality within their organisation is required and support end-user authentication and access control within applications. If there is a data leak, companies can activate a kill switch. Companies then watch the log files to identify and track any suspicious actions. One respondent had this to say:

> "*To protect our data, which includes protecting data at rest and in transit, we have added data loss prevention to our risk plan. Unfortunately, data is lost to careless, compromised, or malicious users, so DLP must be addressed in a managed manner. It is difficult, if not impossible, for many people to rely entirely on both fronts.*"

Data is also protected by mirroring and restricting access. In addition, the risk plans provide device control since users are used to saving data on phones or tablets in addition to real-time auditing and

reporting. Provision is also made for compliance violations, identity theft, malware infections, and data breaches.

5.15.3 Jordan

Jordan has several themes in data access, control systems, data encryption, security, preventative measures and handling of sensitive data and information (see Figure 5.12).

Figure 5.12: Themes covered in risk assessment plans in Jordan

In Jordan, data that gets stored must be transferred over secure sockets. One highlighted method to protect data in motion is based on its sensitivity; hence, investments in a managed file transfer system are compulsory for all staff members to follow the communication guidelines with business contacts. The use of encryption at all points of the data journey and multifactor authentication is also highlighted. Most risk plans are common to firewalls and network access control to secure networks that transmit data against malware and other malicious threats. It is proactive by protecting information upfront instead of waiting for an issue and implementing different policies controlling all user interactions. Data must be processed legitimately, equitably, and transparently to secure a transmission channel concerning individuals.

Another policy raised during discussions is for all staff members to use VPN when they need to access their files while off-site. Companies do not recommend using public computers and free Wi-Fi in this regard. Risk plans ensure that users have some form of training related to cybersecurity and identifying threats. Also, implementing a data loss prevention (DLP) solution where all staff have to follow guidelines is key. This will be utilised to keep sensitive data from leaking from endpoints (personal computers, laptops, mobile phones, servers). It also automatically blocks quarantine or encrypts sensitive data, leaving an endpoint. Appropriate control is assigned based on data classification, making tracking user interactions with sensitive data easier. Risk plans limit the users accessing sensitive data, reducing the risk of data leakage. The plans also implement email content filtering, and this technology scans emails for sensitive data hidden in the text, graphics, and attachments. If sensitive data is detected, the administrator will be notified so they can validate the transfer.

One respondent had this to say:

> *"Informal data sharing is not acceptable. When employees want access to classified information, they can seek it from their line supervisors. Staff should safeguard all data by taking reasonable steps and following company policies. Strong passwords, in particular, must be used and should never be disclosed. Personal data should not be provided to unauthorised individuals within or outside the organisation. Data should be regularly examined and updated if discovered to be outdated. It should be discarded and disposed of if it is no longer necessary".*

Therefore, the policies limit user access privileges to what is needed. In addition, there are regular cybersecurity awareness training sessions, and access rights are revoked immediately if any suspicious activity is detected.

5.15.4 South Africa

In South Africa, the themes that emerge are data classification, data security, data in transit, data encryption, and control systems (see Figure 5.13).

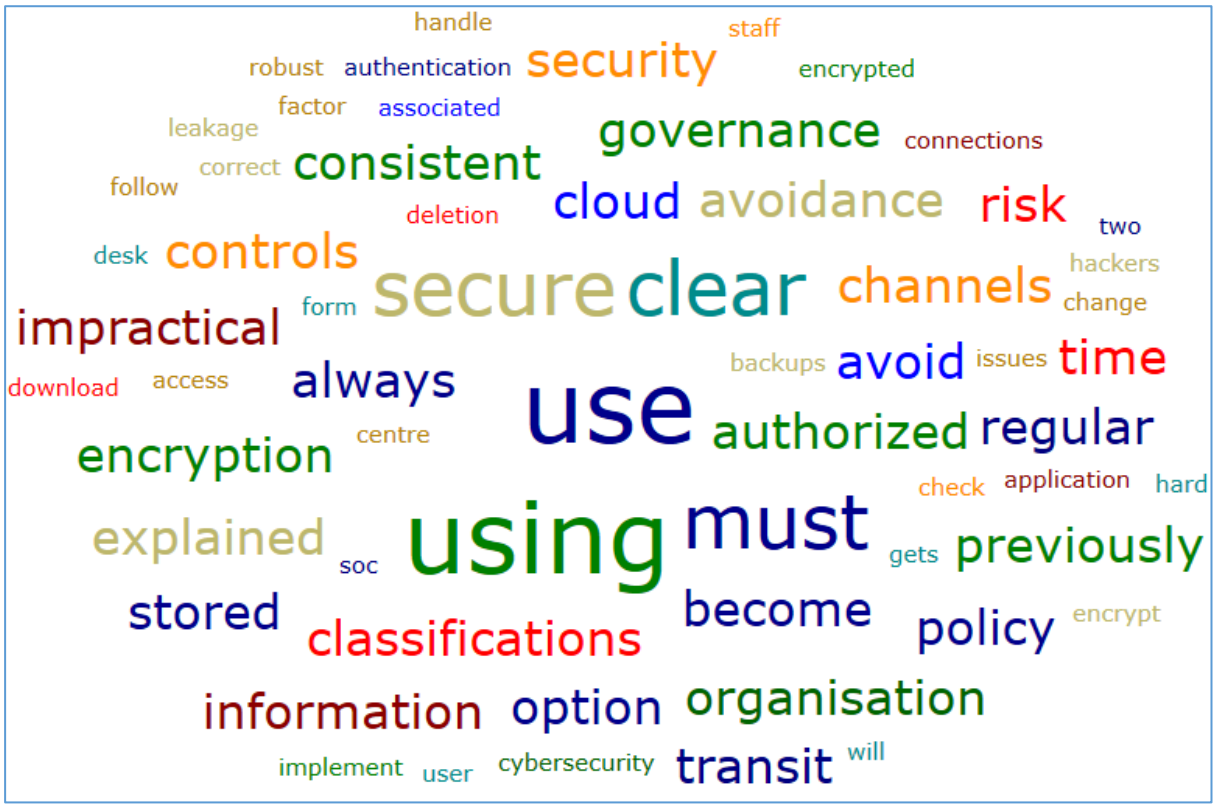

Figure 5.13: Themes covered in risk assessment plans in South Africa

Some organisations were found with no clear and consistent information governance policy. In this instance, it was indicated that data risk classifications become impractical, and the strategy has been avoided for some time. Some respondents identify the implementation of robust network security controls to help protect data in transit, where all staff must follow these controls when connecting to the Internet. There is a need to always use encrypted channels with proper security protocols for transmission. In South Africa, companies avoid data leakage by using high standard encryption on stored data and connections, using two-factor authentication, and keeping backups. Validation is done by authorised persons only. For example, only SOC (Security Operation Centre) handles data in transit. Regular passwords change, and limited data sharing use is also common practice. Similar to other countries, users must be aware and have some form of training related to cybersecurity and identifying threats. Lastly, the following was stated during the discussion:

- Clear guidance on unauthorised data use and privacy issues associated with authorised data processes.
- Restrict access permission to cloud resources prevent deletion of data using cloud lock features.
- Check and download regular operating system and application updates. This will act as a layer of protection against any intruders or hackers.

5.16 Risks related to data protection

5.16.1 Canada

In Canada, themes related to data protection, the federal government's role, and the Personal Information Protection Act emerge (see Figure 5.14).

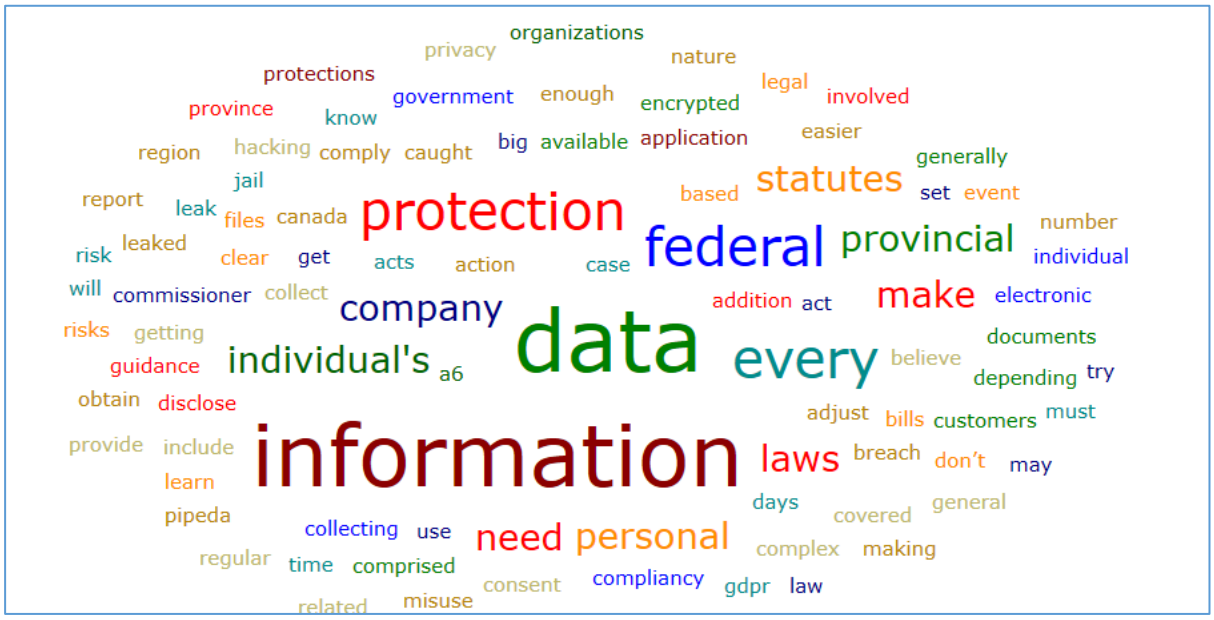

Figure 5.14: Risks related to data protection in Canada

Respondents surveyed said Canada's data privacy laws are a mix of federal and provincial laws. These laws include general federal and provincial data protection regulations. This mix of laws will help everyone concerned to update their risk strategy based on provincial or federal recommendations. Organisations must notify the privacy commissioner depending on the type of data exposed. There are also several requirements to comply with the Personal Information Protection and Electronic Documents Act (PIPEDA). The Personal Information Protection and Electronic Documents Act requires that organisations acquire consent from individuals before collecting, using, or disclosing their personal information.

5.16.2 Jordan

The themes in Jordan relate to legal threats from clients, training, and a lack of authority to deal with the illegal use of data (see Figure 5.15).

Figure 5.15: Risks related to data protection in Jordan

In Jordan, some reported risks include data loss due to malfunction or negligence with potential client lawsuits. Unprotected data can lead to litigation and expose companies to fines and lawsuits. In addition, unauthorised data access can be reported where legal action will be instituted. However, respondents indicate that there is no data protection law in Jordan. Therefore, it is the responsibility of each company to mention all legal actions associated with such risks. Respondents indicate a need to introduce staff training to cover risk matters to mitigate this. In addition, whenever an organisation needs a solution involving data, they should mention that in their agreements as it will be the reference for future issues. However, data must be used fairly, transparently, and adequately for the specific purpose it is intended to.

5.16.3 South Africa

Respondents in South Africa raise several themes relating to the Protection of Personal Information Act (often called the POPI Act or POPIA): legal threats, training, and security (see Figure 5.16).

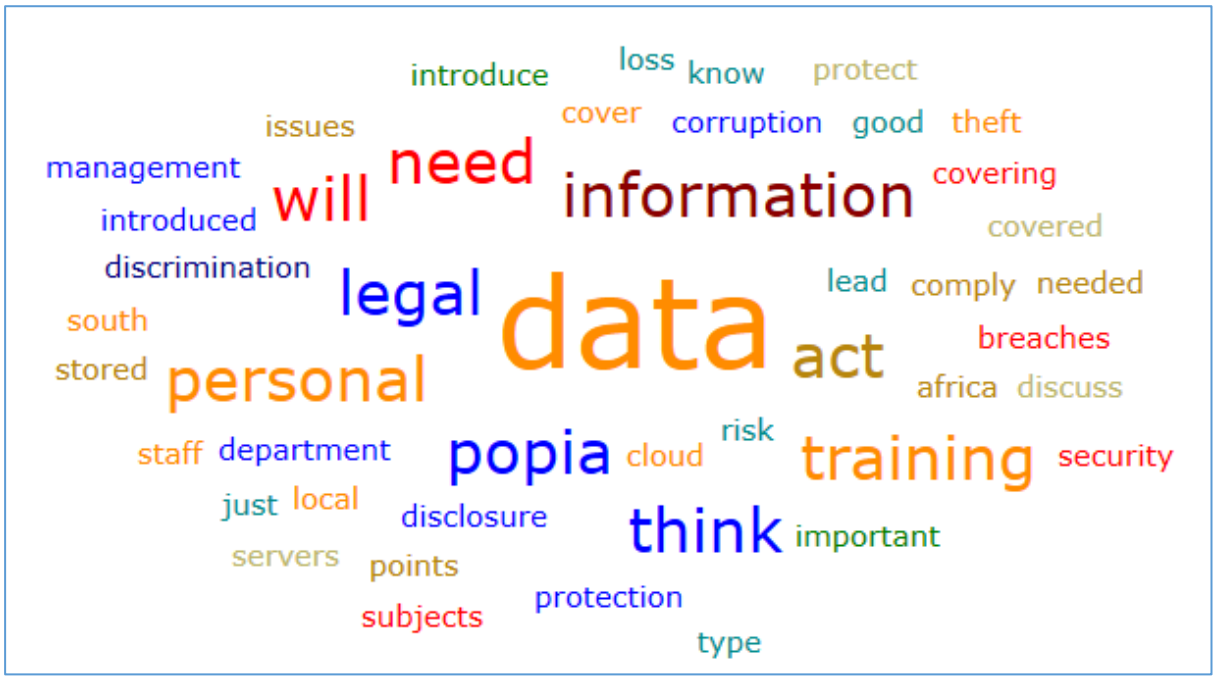

Figure 5.16: Risks related to data protection in South Africa

Data loss, corruption, or disclosure in South Africa leads to legal issues. Thus, good risk management is important when data is stored on local servers or over the cloud. Respondents also concur with those in Jordan that there is a need to introduce some type of staff training to cover data-related risks more. In addition, South Africa introduced POPIA in July 2021, protecting data subjects from security breaches, theft, and discrimination. Therefore, every organisation needs to comply with POPIA covering personal information.

5.16.4 The United Kingdom

Respondents in the United Kingdom raise several themes relating to individual rights, lack of legal threats, data access-related risks, data protection and regulation and rights associated risks (see Figure 5.17).

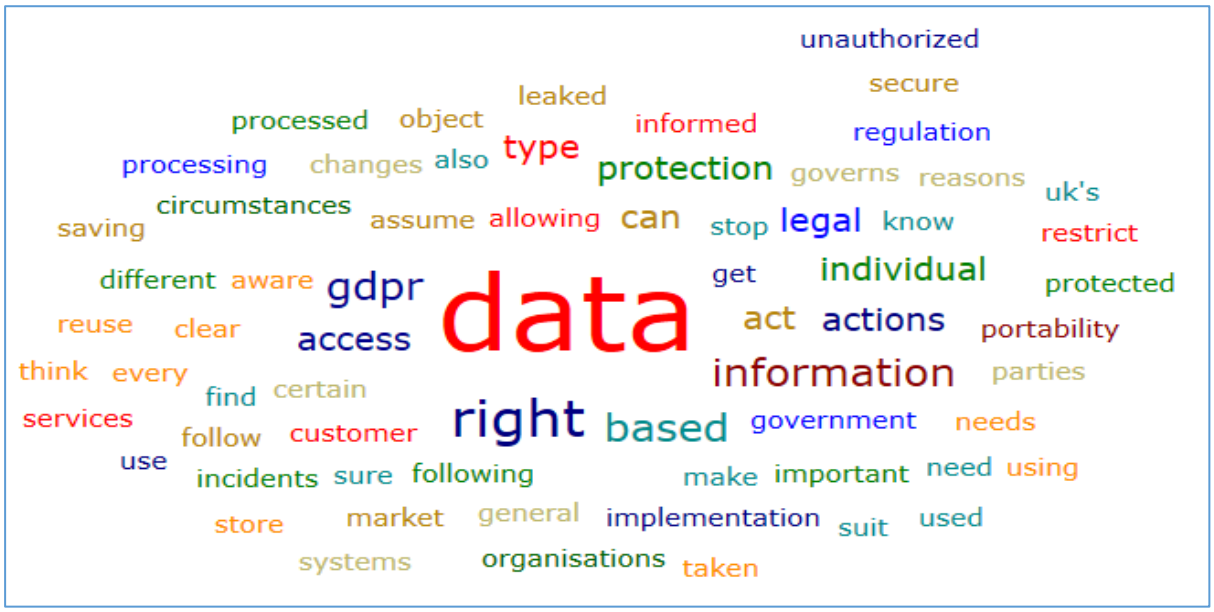

Figure 5.17: Risks related to data protection in the United Kingdom

In the United Kingdom, the Data Protection Act 2018 is the implementation instrument of the General Data Protection Regulation (GDPR). Respondents emphasise the need to make sure that all parties are aware of this act and govern such incidents. Individuals have the right to determine what information the government and other organisations store about them based on GDPR. In addition, companies are following the GDPR with changes that suit their market needs.

In the United Kingdom, individuals have the right to stop or restrict data processing. Legal action can be taken against any unauthorised use of protected data. Every individual (customer) has the right to know what data the organisations are saving or using and how secure their data is on their systems. Individuals have the right to access their data for any reason. One respondent points out that risks are based on the data type, as not all data is important and legal actions should follow based on what information was leaked.

5.17 Summary

This chapter presented the findings of big data risk assessment approaches based on cloud computing technologies in four countries: Canada, Jordan, South Africa, and the United Kingdom. The findings were presented in quantitative descriptive statistics and qualitative data analysis. The quantitative findings were derived from descriptive statistics, frequencies, and graphical representations based on structured questionnaires submitted to industry professionals in the four countries considered big data and cloud systems experts. The qualitative findings were based on a thematic analysis of these key informant interviews done with Atlas.ti software.

CHAPTER 6: DISCUSSION OF RESULTS

6.1 Introduction

This chapter discusses the results presented in Chapter 5, which provided a comparative analysis of big data risk assessment techniques based on cloud computing technologies in four countries: Canada, Jordan, South Africa (SA), and the United Kingdom (UK). The theoretical research objectives of this study were to assess the evolution of Risk Assessment (RA) involving big data in cloud services and the theoretical framework that aggregating different studies could propose to inform the best practices of RA in big data and cloud computing emerging from the scientific literature. The study also aimed at understanding the following empirical research objectives: 1) To analyse the critical features of RA in Canada, Jordan, South Africa, and the UK environments involving big data on cloud services; and 2) to proffering the managerial and policy recommendations that can be developed to mitigate risk on big data in cloud computing environments in these four countries. This chapter provides a brief outlook on the findings per the research questions and objectives. The key findings are listed through empirical; therefore, chapter 6 summarises the results and uses literature to confirm or refute this study's findings. The discussion also proffers an in-depth critical analysis of the study's findings and how the results address the research objectives stated in Chapter 1.

6.2 Outlining the evolution of the research on RA involving big data in cloud services

The virtual dependency increased exponentially over time, and there is a larger need for effective monitoring and management of the data available in the virtual world. Several types of research have emphasized the importance of big data in the cloud and the need for its strict monitoring and articulation. The study aimed at understanding the evolution of risk assessment in big data using cloud services. When machine learning-based algorithms and virtual space gained pace in the industrial world, the malevolent players of information technology started making use of the cloud data that was open, unattended and subject to vulnerability. Therefore, with increasing fraud and data confidentiality issues, risk assessment methods came into existence and western countries adopted the framework on risk assessments due to sound infrastructure making them capable of adapting to IT transformations.

Further, big data using the cloud also gained importance due to high storage capacity, accuracy, just-in-time management and low maintenance compared to real-world data. The study findings

revealed that big data using the cloud evolved and reached widespread due to volume, variety, velocity and veracity. The importance of risk assessment processes in big data and cloud technologies is a ramification of the pressing matter of supporting several incumbents in making insightful decisions, forecasts, and data visualisation. A core element in characterising and reporting threats is implementing the following steps appropriately. Firstly, to analyse the origin of risk events and understand them to visualise their structure correctly. Secondly, to assess the potential losses related to every event in case they materialise. Thirdly, to predict the probability or potential for the event using statistical approaches with probabilistic evaluations or subjective judgments with approximate reasoning (Djemame *et al*., 2016). One critical feature of a risk assessment involving big data on cloud services is the level of knowledge of cloud computing. Thus, the overall exhaustive compilation of protective security and potential threat identification in the four countries was presented in the study. They suggest countermeasures, but the vulnerabilities cannot be avoided without an integrated approach and effective network. Risk reduction can be expected through RA parameters and effective defined mechanisms.

6.3 Compiling the main results of the studies on RA in big data and presenting a theoretical framework aggregating these studies

Risk assessment is still nascent, and there are numerous obstacles to managing confidential and critical big data. Risk assessment methods can quantify the risks corresponding to security, privacy and safety. Several nations have adopted integrated approaches to risk assessment, but in various developing countries, organisational data is subject to risk and threats concerning confidentiality and misuse. Big data usage is abundant, but immersive risk assessment instructions and approaches are lacking in many nations. The linkage of existing distress was with the frequency of threats, duration of fraud and negligence of management towards RA measures. The theoretical framework explains the formation of linkage between the related concepts of big data, cloud and the need for risk assessment. The diagrammatic representation is stated below:

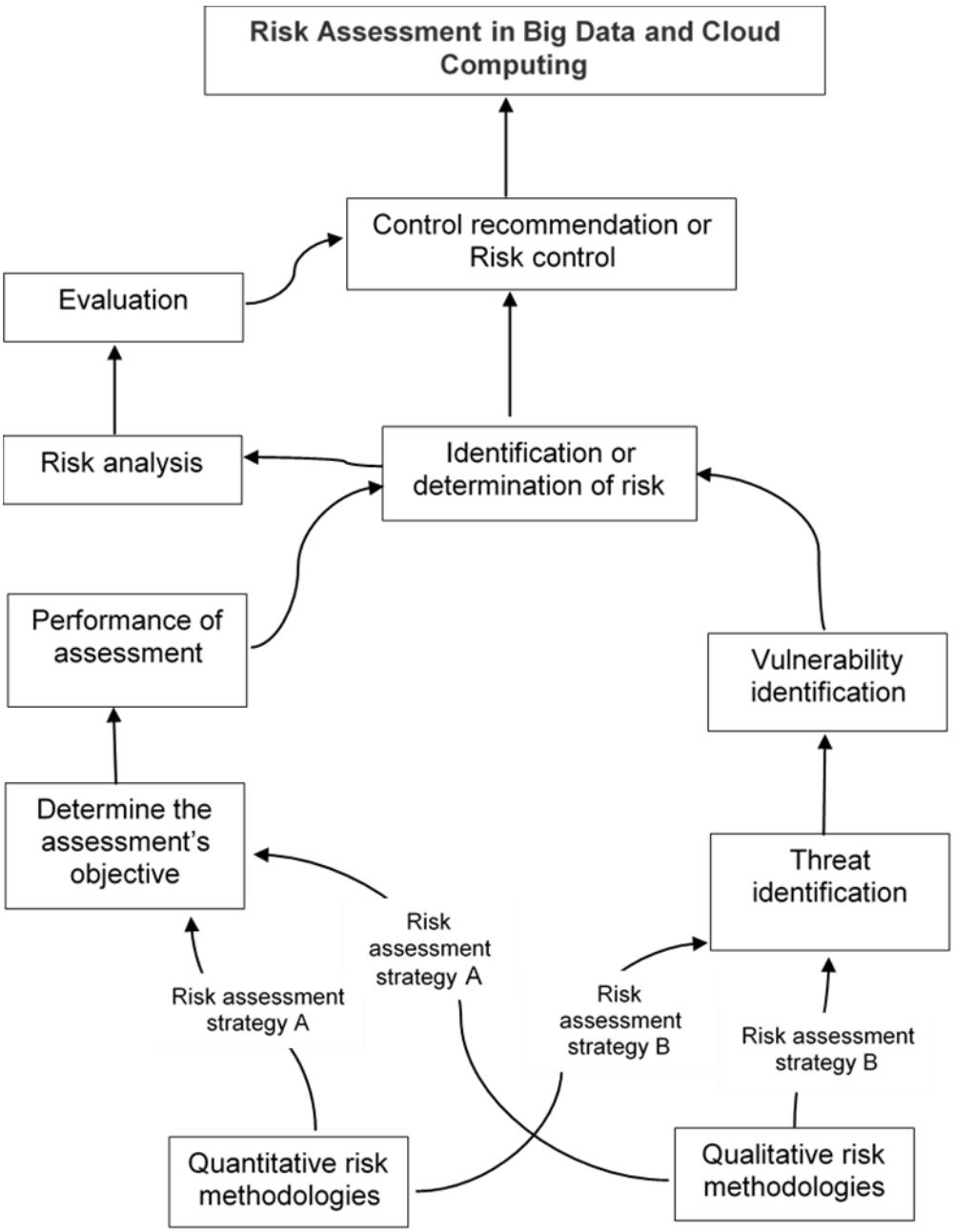

Figure 6.1: Theoretical risk assessment framework for cloud computing and big data

The theoretical framework suggests the holistic approach to risk assessment and creating an integrated environment where big data using the cloud can be secured from external agents. The theoretical framework and its implications are further stated in the chapter below. Cloud platforms are resourceful to enable the users with immense capabilities to have an integrated network where they can consume the big data and successfully implement their operations. However, the data in the cloud is highly vulnerable to various threats, risks, and leakage concerns as virtual data privacy are intricate and advanced. To understand the risk assessment environments, this study evaluated the knowledge of big data across all four countries. The results reflected novice and intermediate

levels of expertise with big data. South African organisations were novice at knowledge about big data. This could be due to a lack of executive understanding of the competitive advantage that big data analytics provides by enabling businesses to respond in real-time to data collected in real-time, even though big data analytics could enable South African businesses to analyse formal and informal data in real-time and thus remain relevant in today's world. Most organisations in the United Kingdom had an intermediate understanding of big data. Jordan and the United Kingdom have expert knowledge about big data. Canada was spread between novice and advanced expertise. Most organisations in South Africa, Jordan, Canada, and the UK conducted risk assessments during planning before deploying their business process to the cloud (Islam *et al*., 2017). Also, Abioye *et al*. (2021) add that risk assessment by companies has led them to enjoy the benefits of cloud-based services. It is important to note that it has become increasingly difficult for companies to survive in a dynamic environment due to limited time to access a market and respond to clients' business demands.

The use of cloud computing has reduced processing time and costs on architectural design. This indicates that Canada, Jordan, South Africa, and the UK have advanced knowledge of these characteristics, which has reduced the risk involved when managing large data. The results have shown that there are common features of risk assessment, and it is critical for countries to assess the risk to ensure that the data available is of quality. Thus, it can be said that big data using the cloud is an imperative and inseparable part of organisational workflow nowadays, and therefore risk assessment methods must be upgraded with time to complement the advancements of malevolent players. The big data of organisations are easily available due to the multi-modal distribution system. Therefore risk assessment frameworks should be implemented at various stages of operations in all countries so that mitigation of potential threats and risks can be assured.

6.4 Compiling the best practices of RA in big data and cloud computing emerging from scientific literature

It is important to prevent undesirable events that could create devastating attack impacts on organisations. The overall security compromise can be reduced by the following analytics, controlling and other practices of RA in big data. In this study, risk assessment involving big data on cloud services considered implementing big data technologies within organisations. While the study found that organisations were implementing big data technologies, low adoption of big data in South Africa was evident. Big data technologies are not being implemented within South African organisations, varying reasons referring to high costs, security issues, and a lack of

information. It was found that there is insufficient information on how to implement this technology, and therefore, more research is needed in this respect. This study highlighted employees' lack of required skills and the budget to hire and adapt. The primary concern was implementing big data's security and privacy issues. South African organisations preferred to learn about best practices before implementing such technology. Thus, there was a feeling that big data does not add value to the current situation, and organisations will keep evaluating the necessity on an ongoing basis. A major drop in the use of cloud computing in South Africa was mainly due to the levels of understanding and maintenance issues of the cloud computing systems.

Similar findings (Abdullahi et al., 2019; Zissis and Lekkas, 2012) indicate that despite efforts done by the South Africans to develop international data transmission links between South African companies with the global market, consultation on the use of cloud computing resources located outside Africa are still high. This indicates that consultation on cloud computing is high, implying the understanding of the software is still low in South Africa. In addition, this clearly shows that there are possibilities that the implementation of big data technologies is still limited in this country. Unlike Canada, Jordan, and the UK, the country still faces resistance from insurance and financial companies that are sceptical of security threats due to hacking. Finally, it is important to note that the level of development of cloud computing is directly proportional to the availability of Internet connectivity.

Despite the low adoption of big data in South Africa, many organisations supported this technology, as it helped expand and improve the employees' ability to work. Generally, results in South Africa indicated that data technologies enhance operations and provide better customer service, giving a better market understanding. For organisations that use big data technology, it helps to make smarter decisions based on analytics, helping to know what customers are looking for, helping risk management models work better, and creating more innovative plans and policies. Financial services corporations are actively leveraging the cloud to increase their business agility, update their operations, save money on procurement, and streamline and accelerate the deployment of new assets and services (Pourghomi and Ghinea, 2012). These advantages contribute to increased flexibility, scalability, and the capacity better to match technological resources to corporate needs and goals, all of which are critical components of a competitive edge. The financial services sector's shift to the cloud has been a long process.

Nevertheless, the nature of the sector's operations has necessitated a cautious embrace of cloud computing (Moonasar and Naicker, 2020). The financial services industry handles some of the

most confidential and sensitive customer information. Individuals can suffer serious repercussions, as can the company itself, when customers' financial and personally identifiable information is breached and exploited. That is why financial institutions are subject to a web of data privacy and security regulations, such as SOX (Sarbanes-Oxley Act), GLBA (Gramm-Leach-Bliley Act), and PCI DSS (Payment Card Industry Data Security Standard), which govern how they store and use data and who can access it, and that dictate stiff fines and penalties.

Jordan also had a significant number that indicated that big data technologies were not being implemented due to costs and a lack of a qualified workforce. There was an indication of the need for more assessment on how big data technology will benefit companies. However, the study found that big data technology can boost the client base and save time. It also leads to cost savings and is excellent for customer insight, directing organisations to focus and target customers, resulting in more sales and profits. Cloud computing is the ideal option for enhancing business performance because it reduces costs, time, and storage (Obeidat and Turgav, 2012). Thus, most cloud computing providers provide public services (Gupta and Thakur, 2014). Thus, transferring operations and services to the cloud reduces energy usage and maintenance costs, making cloud computing a creative technique (Martens and Teuteberg, 2011). This leads to cost-effectiveness, which is why businesses use cloud computing. It also influences people's desire to accept new technology (Odeh et al., 2017; Al-Sharafi et al., 2017).

In Canada, most respondents indicated that big data was being implemented, but a few indicated otherwise, raising the issue of costs and the need to study the technology further. The results showed that the perception in Canada is that technology is safe, convenient, and reliable. Mohammed et al. (2017) found that perceived control in cloud computing technologies influences user intention to adopt the software. However, cloud computing gives third parties, such as institutions, governments, or other communities, control over data (Abu-Shanab and Estatiya, 2017; Lang et al., 2018). Cloud customers will be more inclined to accept cloud computing if they feel in control of their transactions and data (Abu-Shanab and Estatiya, 2017). In summary, cloud customers use cloud applications if they know proper protections over their sensitive data.

In the United Kingdom, no challenges were raised about big data technology. Instead, it was found that the technology was being implemented 100%. The most common factor influenced the adoption of cloud computing technologies in education. The level of education in the UK is approximately 100%, making it easy for most organisations to understand the importance of cloud computing technologies to a business. This implies that most people or organisations are familiar

with Internet technologies, and such individuals can easily use cloud technologies (Gangwar et al., 2015). Most organisations can provide adequate training to service users in the UK, promoting the importance of cloud computing technology and demonstrating how to use it. This has led to the increasing use of technologies in most organisations, and there is a need to improve the functionality of every organisation constantly. Thus, risk assessment practices are effectively adopted in UK and Canada to mitigate the risks and threats. The primary reason behind their efficiency is infrastructure, developed framework and aware community that triggers a sense of responsible big data management. Further, in South Africa and Jordan, the cost of implementing risk assessment methods inhibits the progress of security features and makes the big organisational data vulnerable.

6.5 Explore the critical features of RA in Canada, Jordan, South Africa and the UK environments involving big data and cloud services

The risk assessment framework employs formalized checks, scrutiny and real-time monitoring of big data so that attacks, security breaches and network discrepancies can be minimized. The features concerning extensive applicability and shortcomings of integrating big data, including cloud services and big data, in organisations must also be weighed in a risk assessment. Large-scale data storage, processing, and analysis are becoming increasingly affordable for Jordanian businesses due to advances in big data technology. According to this study's findings, in Jordan, big data technology may be used to identify cost-effective and productive corporate models and practices by providing more accurate data, empowering a new generation of employees, cost-saving, and controlling online reputation. In Jordan, the advantages are the flexibility and the availability of services. In support of this view, Poniszewska-Maranda et al. (2019) argue that cloud computing technologies reduce the need to respond to clients, enhancing an organisation's brand equity, particularly in the banking and insurance companies. Likewise, cloud technologies have also improved companies to respond to the dynamic business environment and the global market's needs (Albelaihi and Khan, 2020) and enhance the educational environment (Backialakshmi and Sumalatha, 2020). The findings from these studies indicate that technologies are useful to modern business, and the business community must adopt these technologies to remain competitive in the market. The organisation creates customer retention with highly satisfied customers, saving advertising costs and reducing high labour turnover. This study found that most organisations reflected the intermediate and advanced levels of knowledge with cloud computing. Results also showed that some South African organisations exhibited a novice level of expertise in the level of knowledge of cloud computing. This could result from security concerns, a lack of data centre

infrastructure, or cost. Also, South Africa being a developing country and a latecomer in big data analytics, there is a possibility of a general lack of strategic vision and drive to support its implementation. However, most organisations in South Africa were found to have intermediate knowledge of cloud computing. The United Kingdom had the most organisations with an advanced understanding of cloud computing. Canada fared better than all countries in terms of experts in the level of knowledge of cloud computing. It is important to add that both Canada and the UK are among the biggest ICT markets worldwide; Jordan and the United Kingdom also reflect high expertise in the level of knowledge of cloud computing.

Moonasar and Naicker (2018) conducted a study on the readiness and maturity of organisations in South Africa on the adoption of cloud services. Their findings indicate that most organisations in South Africa had an intermediate understanding of the knowledge of risks associated with cloud computing and their readiness and maturity for cloud services adoption. Despite the benefits derived from adopting cloud computing and the view that public cloud services are gaining momentum in South Africa, large organisations still have minimal understanding, leading to cautious adoption of cloud services (Moonasar and Naicker, 2020). This implies that most organisations in South Africa perceive challenges associated with cloud computing compared to the Jordan, Canada and UK organisations that fully implement the software. Organisations need to be informed of the software's usefulness since it reduces the costs and time needed to process. Thus, the company will manage more volumes of data and offer quality services to its clients.

South Africa was the only country with some organisations that indicated risk assessment challenges. Unlike in the developed countries (the UK and Canada), this could be partly attributed to most cloud service customers not having adequate knowledge in performing such assessments at a good level in South Africa. Observations during the interviews were that most organisations in South Africa do not employ IT specialists and that there was a lack of transparency, which was intrinsic to the operations of the cloud service provider in the country. A lack of risk assessment may lead to difficulty choosing the right cloud service provider, thus lagging the adoption of cloud-based computing systems. However, most were found to have "advanced to expert" knowledge about risk assessment. The United Kingdom also showed many organisations with advanced knowledge about risk assessment. Canada and the United Kingdom are viewed as experts in risk assessment knowledge. The results indicated mixed risk assessment levels distributed across intermediate, advanced, and expert in Jordan. A study by Alnemr et al. (2016) indicates that UK, Jordan, and Canada have advanced knowledge of risk assessment. However, the Alnemr study's

findings indicated that standard IT infrastructure, data management, and analytical techniques are difficult to adapt to the rapid growth of massive data, resulting in a slew of complications such as data leakage issues. The same study's findings are consistent with Gartner (2013), which shows that South Africa has limited knowledge about risk assessment compared to the UK. The findings of Alnemr's study also indicated that South Africa is the only country lacking knowledge on risk assessment using cloud computing. As a result, the country faces challenges in implementing cloud computing to assess the risk of big data management.

In terms of managing networks by country, the results showed that most countries have a dedicated team within their IT departments to manage their networks. South Africa and Jordan dominated in having a dedicated team within their IT departments to manage their networks. Due to a large number of hiring, especially in the two countries' government sectors, gives them enough workforce to dedicate enough team members. In Addition, South Africa and Jordan do not have the required infrastructure to consider outsourcing. Outsourcing services to operate the network was more dominant in the United Kingdom and Canada. Jordan is a developing country that is fast adopting cloud computing technology in diverse characteristics. The low adoption of CC technology challenges Internet users in Jordan and South Africa. Companies have significantly observed this challenge as they focus on delivering and cost reduction. Jordanian and South African companies have increased investment in networking compared to UK and Canadian companies, mainly outsourcing services (Lang et al., 2018; Santos et al., 2014). In the next five years, most companies in South Africa and Jordan will use cloud computing in their day-to-day business operation, meaning quality service and less time will be needed to offer a service. However, companies in these countries are prone to facing high costs. This is because they have to offer the service themselves, unlike the outsourcing strategy, which is cheap given that a company only pays for the service without considering the servicing costs of the software.

Risk assessment involving big data on cloud services also considers how countries run big data applications (Kaseb et al., 2019). South Africa was the main country running its big data applications, mainly on the traditional storage system. As argued earlier, this could be partly blamed on the costs of adopting the cloud computing system. In addition, observations point to scepticism on adopting the new technology due to data exposure and difficulties in managing the Protection of Personal Information (POPIA) requirements, which commenced in July 2021. In Jordan, the results showed an equal split between the traditional storage system and the applications on the cloud. Canada and the United Kingdom were the main countries that run their big data

applications on the cloud. The findings from prior studies indicate that security challenges companies adopt cloud computing technology (Chang et al., 2016; Odeh et al., 2017). The four countries use cloud computing to secure data (Al-Rousan and Hashish, 2016; Al-Sharafi et al., 2017). This implies that companies in these countries are investing much in big data management to protect their data to ensure confidentiality, integrity, and data availability. Once the company's data is available and protected, companies will address customer needs and respond to market dynamics. However, a lack of strong security procedures and identity management standards adversely impacts organisational adoption of cloud computing (Oliveira et al., 2014). This means data security predicts users' intention to adopt cloud computing

Certain challenges and disadvantages noted in Jordan are related to data quality; for example, analysts need to verify that the data they are working with is correct, relevant, and suitable for analysis. The technology suffered from challenges such as user-level execution that only exists in select channels, user-level results that cannot be presented directly, and user data not suited for producing learnings. There was also a special feeling in Jordan that there was no urgent need for big data technology. Current legacy systems achieve the desired functionalities without using big data technologies. Therefore, decision-makers do not need to pay more money to replace the old with big data technologies. In Jordan, big data technology was associated with cloud platforms' long-term hidden significant expense. The implementation expenses and the expenses of recruiting additional qualified individuals to take on the role were also discussed. Harfoushi et al. (2016) examined cloud computing use in Jordanian hospitals. They assessed 223 IT departments from area hospitals to determine their preparedness for using cloud computing. Their research examined the acceptability of cloud computing by examining technological, organisational, and environmental aspects (the Technology Organisation Environment-TOE framework). The findings suggest that the major components of the TOE framework had a substantial positive effect on Jordanian hospitals' inclination to incorporate cloud computing.

Additionally, they observed that the largest effect on their choice to use cloud computing is due to technical reasons (relative benefit, complexity, and compatibility). Additionally, Hammouri and Abu-Shanab (2020) interviewed 110 students from several Jordanian universities to ascertain the downsides of cloud computing. The results extend the unified theory of acceptance and use of technology (UTAUT) to assess students' adoption of cloud computing services and the extent to which they assist them in completing their assignments. The results indicate that, except for

anxiety, all of the model's primary characteristics are significant predictors of students' desire to utilise cloud computing applications.

This study found that big data technologies are related to reliability, better data control, high availability, automated updates, easy data backup and recovery, and scalability in the UK. Moreover, technology would guarantee technological independence and flexibility. Big data technology is associated with ease of use and cost-effectiveness; also, respondents in the UK highlighted flexibility in work handling. Implementing cloud technologies has assisted most companies in storing and managing data in the health care sector (Kishor et al., 2021). The data was mostly on processed patients with chronic diseases to develop emotion recognition systems and deploy blockchain applications (Tang and Zeng, 2020). Cloud technologies show that service delivery has improved in the UK, and customer satisfaction is highly achieved. It has become easy for most organisations to check on company growth, given that customers can communicate with the company at any given time. This has reduced the rate of errors since consumers now participate in the production department or service delivery by constantly informing the supplier of the service or product they are expecting from the company.

The UK's disadvantages were limited infrastructure control, flexibility, ongoing costs, and security. High cost and a lack of required experience were linked to the complicated infrastructure to host the data. This study found that it takes a long period to ensure full implementation of the technology, thereby raising the costs of setting it up. A lack of required experience, risk of being hacked, complicated infrastructure to host the data, limited control of infrastructure, and limited flexibility featured strongly. According to Taleb et al. (2021), lack of a clear value proposition and standardisation are common challenges when implementing big data technologies in organisations. Value proposition and standardisation affect the implementation of cloud computing technologies in developing countries. It is critical to note that most studies focus on the benefits of cloud computing technologies; however, this study has observed that most organisations face problems they cannot solve to improve implementation. As a result, most organisations face implementation challenges due to the high costs involved since implementation involves monitoring and evaluation as a continuous process. Picciotto (2020) noted that implementation security risks are inherent in the cloud computing service known as infrastructure as a service (IaaS); however, Picciotto's study failed to provide strategies to reduce implementation failure.

In South Africa, the advantage noted was data processing on a large scale. This study pointed out that cloud storage provides more storage on files and keeps the company information safe and in

one place and the processing of data on a large scale. However, in South Africa, most organisations argue a lack of privacy in cloud computing technologies. This is because organisational data can be stored in different locations where the country operates. This is perceived as a bigger risk to confidentiality and privacy breach (Zissis and Lekkas, 2012).

Because each nation has its own rules and jurisdiction surrounding confidentiality and privacy, the cloud data storage problem is exacerbated by legal challenges involving privacy issues (Moonasar and Naicker, 2020). In South Africa, implementation costs were also presented as a challenge in addition to finding new, skilled staff to take up the task. Also, South Africa raised the issue of limited data control and security issues. Chihande and van der Poll (2017) indicate that implementing big data technologies was expensive and difficult for small organisations. The costs factor was mainly due to the initial costs involved, such as training of staff and errors made by staff during implementation. This shows that implementation requires companies to be prepared; hence, there are costs associated with implementing big data technologies. Training, infrastructure development, and consultation costs were additional costs organisations should be prepared to encounter when implementing big data risk assessment software (Chihande and van der Poll, 2017). To guarantee data safety, vigorous training and appropriate technologies are necessary, which normally pushes the costs of the company high.

Similarly, in Canada, big data technology ensures the availability of services and flexibility. The technology will guarantee independence, flexibility, easy access, and flexibility in work handling. However, it is critical to note that risks must be seen as an overall business opportunity. Thus, a business should have a high appetite for risk-taking (Martini and Choo, 2012). This means companies should take an opportunity to invest in big data technologies so that they reduce customer dissatisfaction and also improve design quality. For example, insurance companies will offer services to customers efficiently since the data is easily available.

Conversely, most firms usually have complex IT applications that lack a high degree of data standardisation to fully take advantage of the scalability promised by cloud computing (Yeboah-Boateng and Essandoh, 2014). The highlighted disadvantages in Canada were concern over data security related to information hijacking. Also, big data technology is perceived as costly to implement in Canada if the organisation has no guaranteed outcome. In Canada, the challenges of finding qualified people to work in the field and the need for new hardware will be an added cost. Lastly, this study raised the matter of the hidden high cost of the cloud infrastructure in the long term. Canada and the United Kingdom were the countries with most respondents planning to run

their data on cloud storage. In Jordan, the respondents were split equally concerning plans to run big data applications on traditional storage or cloud systems. South Africa is the only country with most respondents planning to run their data on traditional storage.

However, many challenges are experienced when running big data applications in the cloud. In Canada, the challenges ranged from scalability, slow file download, frequent downtime, and many errors – overloaded with error reporting, data security, and different types of security breaches such as malware attacks, cross-site scripting attacks, and phishing attacks. This normally emanates from divergent consumers' sharing of data resources, preventing service providers from customising their services to suit individual customer needs (Tarkyth, 2019). Firms' big data technology adoption readiness is negatively impacted by customisation because they are unclear about CC's capacity to fulfil their business demands (Wang et al., 2011). They have a distinct perspective on customising and personalising a user's computer environment.

For this reason, organisations must learn more about the available resources and services tailored to their business needs before implementing CC to guarantee that CC offers this customisation. In the United Kingdom, the challenges range from storage and processing of data, which is costly, security control too complex, expensive cloud storage and cost for small businesses start-ups. However, as with other rising economies, the UK has embraced big data technologies, with technology experts extolling the cloud's enormous and untapped business potential (Conway-Smith, 2012). Additionally, cloud computing optimises scarce resources by pooling them into a resource pool that is efficiently provided to customers based on their requirements. Businesses must implement this strategy in IT service delivery (Conway-Smith, 2012). This implies that companies have embraced cloud computing technology, reducing operational costs and the time needed to complete a task. In Jordan, challenges ranged from safety measures, and data that must be double-checked frequently cost of setting up and maintaining the technology and cloud service charges and cloud service reliability, efficiency, and security. Also, storing and analysing a large volume of data requires a vast and complex hardware infrastructure. Too much data also meant a higher risk for cybersecurity and threats.

Alnemr et al. (2016) point out that most countries lack basic infrastructure to effectively implement big data technologies, for example, electricity to power the devices. Any development project relies on these facilities. Organisations fighting to bridge the digital divide are hampered by inadequate infrastructure. Thus, a lack of infrastructure leads to a lack of internet infrastructure, rising service

charges, and rising device costs. Adequate internet coverage is required (Mujinga and Chipangura, 2010).

In South Africa, achieving a reliable bandwidth for the operational business requirements is challenging. Companies cannot afford it as it is still too expensive, and finding trained and skilled individuals with the right knowledge is difficult. On the other hand, cloud service providers are seen as reputable and trustworthy by South African businesses in general. In South Africa, according to the findings of previous research, cloud data technologies are believed to be readily available, and the quality of data is perceived to be increasing. MacVaugh and Schiavone (2010) argue that the dissemination of innovation is influenced by factors such as technology, environment, and social systems in the context of a certain market or sector, backed by research. Cloud computing is a specialised and complicated adoption of information technology that would be exposed to the problem context seen by the relevant stakeholders in addition to other factors (Jokonya et al., 2012). Because of data governance and control, local content security policies are preferred over remote content security policies.

Thus, it can be said that most organisations in all countries confirmed that cloud computing was implemented. The efficiency of cloud computing in Canada was due to simplicity in managing infrastructure and easy accessibility and management. This study found that data is safer from loss, and the organisation can use a PC to access files from anywhere during the pandemic. Lower overhead, scalability, agility, application, and security are advantages. However, any device can access files anywhere, which can be considered a risk. Cloud computing tends to rely on Internet speed and reliable connectivity. There is also a challenge in transferring data from on-premises to the cloud due to network bandwidth, hence the reliance on service providers for the service. Finally, security threats and high training needs are featured prominently. According to Elahi et al. (2021), some organisations have not yet implemented cloud computing due to security and privacy concerns; instead, they evaluate the advantages and disadvantages of various cloud paradigms from afar. As a result, it is necessary to evaluate cloud-based business operations that employ security risk management procedures as a data protection security measure to discover research gaps (Alnemr et al., 2016). This is critical owing to security concerns that impact commercial and organisational objectives.

Advantages in the United Kingdom are that cloud computing is easier to manage — better scalability, availability, and improved turnaround. There is easy workflow, unlimited storage, efficiency, and reliability. Implementation of cloud computing allows customer data analysis,

provides better service and support and improves operation and processes. However, the technology is costly to implement, requiring highly experienced staff and complicated infrastructure to host the data. The downtime, since the service is Internet-based, an outage can happen at any time and cause serious issues, and IT is expensive to deploy. Shifting to cloud computing technologies requires information and experience from the organisation's support engineers. Implementing new technology necessitates a thorough understanding of its existence and usefulness (Heinle and Strebel, 2010). The absence of prior expertise and experience with big data technologies by support engineers can form a risky challenge for the organisation (Khajeh-Hosseini et al., 2010). Organisations must ensure that they have the necessary skills and knowledge for CC implementation, as CC is a technical field that demands specialised resources (Chauhan et al., 2012).

In Jordan, the advantages highlighted were that cloud computing helps higher education institutions to benefit from lower costs, allowing files and service access from anywhere at any time. This study noted that cloud storage is better to manage and control. Other issues highlighted were enhancing security and privacy, enhancing institutional productivity, making the educational process more efficient, enabling collaborative work, and allowing for hybrid working, thus saving resources, hardware, licensing fees, and DC space. However, in Jordan, not all systems can be implemented on the cloud as high confidential data should be stored and processed on resources owned by the institution. Also, in Jordan, the technology suffers from a lack of control dependency on network performance, is associated with high costs, and is harder to control. Security risks of being hacked and lack of required experience are featured prominently. This study also found infrastructure to be complicated to host the data. Since the service is Internet-based, outages can happen anytime and cause serious issues. Cloud computing demands organisations to upgrade their technology, that is, new data structures, to control data and new file systems (Goyal, 2014). This means companies should create an IT department that is immensely involved in big data technologies to ensure that companies are ready to implement the system. In Jordan, most companies failed to effectively implement cloud computing due to a lack of coordination between the IT department and other departments within organisations (Chang et al., 2016). For a company to achieve its objectives, it needs to promote effective implementation of the business.

In South Africa, cloud computing enables more storage resources and backup if main systems crash or become corrupted. Thus, files are safe and have more accurate storage efficiency, storing large amounts of data. There was a notable improvement in turnaround time and easy access to files and

secure information. The issues of cost savings and high-security levels were also noted. However, like all other countries, the cost and the need for highly skilled staff that requires specialised training were highlighted. There are also data protection and security issues — data loss and theft. Lastly, most of the components are not entirely under their control. While the South African information and communication technology (ICT) sector continues to develop rapidly, it has not kept pace with the national goal of affordable access to a complete range of communication services. While access to ICT services continues to expand, broadband connectivity (especially fixed broadband) remains extremely limited in comparison to another low- and middle-income nations. All communication services continue to be expensive by African and global standards. This affects the complete deployment of cloud computing technologies in a growing number of organisations in one way or another (Stork et al., 2013)

6.6 Develop managerial and policy recommendations on RA on big data in cloud computing environments using data from Canada, Jordan, South Africa, and the UK

The scope for web-based applications and operations of companies is vast and exponential. Developing impactful risk assessment measures and creating an infrastructure that could synchronize the flow of functions is highly integral to curbing damage to organisational assets and reputations. While this study found that most organisations in all countries were implementing cloud computing technology, all four countries seemed to struggle to implement policies for terminating or transferring data to the cloud service. In Canada, respondents indicated that they were in the process of implementing the technology and covering the risk involved as a part of their study. In South Africa, the study noted that they were still struggling with a clear information governance policy; hence, South Africa has refrained from moving potential sensitive data to the cloud. This study also indicated that more assessment is needed in this regard. In Jordan, it was revealed that costs are not yet worth migration. In Jordan and the United Kingdom, indications were that the plan to put policies in place is still on the cards, but it will not be rushed due to challenges such as costs (Ali, 2015). Most countries lack a legal framework to support cloud computing compliance. Each nation must develop local legislation and domestic data privacy policies governing the ownership and administration of cloud-based services to ensure successful compliance. Emerging data privacy legislation in the United States, Europe, and other places may contribute to organisations' uncertainties (Lal and Bharadwaj, 2016). As a result, organisations struggle to choose which data should be migrated to the cloud and which should remain on their on-premises system.

The study suggests a need for a clear risk assessment plan and framework based on which companies and cloud service providers can be made resourceful. In South Africa, organisations have established a well-managed risk management process and system, which forms part of the organisation's overarching Governance Risk and Compliance framework. It also integrates very closely with a cyber-security strategy with a risk-based approach. IBM opened a data centre in Johannesburg in 2009 to provide cloud services to South African businesses, while VMware also developed vSphere, a cloud operating system, in 2009 to serve the South African market (Kshetri, 2010). South Africa's bandwidth has increased significantly (Mohlameane and Ruxwana, 2014). South Africa's bandwidth challenges are improving due to the development of new undersea telecommunications cables (Grobler and Dlamini, 2010). South Africa's overall online activity was predicted to account for 67% of all African countries' combined online activity.

Furthermore, SA is one of Africa's top Internet countries, which is advantageous for CC because it indicates that the quality of internet services available to support CC will be adequate (Grobler and Dlamini, 2010). Moonasar and Naicker (2020) argue that the risk assessment model for South Africa is too abstract; this means the model still has leakages making it difficult for organisations of different sizes to implement. Despite the availability of the framework, companies are failing to implement it due to limited levels of awareness. Low levels of understanding by most firms on cloud computing technologies have made it difficult for companies to implement risk assessment plans.

Further, the study suggests necessitating towards business recovery plan so that the damage caused due to threats can be recovered and built up again. In the United Kingdom, respondents indicated a business recovery plan that includes data and network backup and recovery aspects, tested regularly. Other highlighted programs specify the action to take if there is a cyber-attack concerning cloud computing and a policy on the confidentiality of big data. Most European countries have established well-documented policies that guide companies in implementing risk assessment. This has reduced risks for companies since cloud computing is very diverse. For example, the European Union has developed mandatory risk assessment methods to mitigate threats (Mackita *et al*., 2019). This has reduced cyber-crime incidents and increased e-commerce activities, resulting in market opportunities such as the export market. Other authors focus on risk management methods in e-commerce contexts (Li *et al*., 2019), risk management methods in resource provisioning (Halabi and Bellaiche, 2019), and risk management in logistic processes based on the cloud (Maniah and Milwandhari, 2020). However, loss of control and lack of data ownership were risks that firms might face when implementing cloud applications. Although these

works provide important insights on risk assessment in cloud environments, they mostly focus on quantitative methods and not on several countries, a restriction that this research tries to overcome.

Furthermore, there is a need for effective training as per the advancements so that the workforce and the community are prepared for the essential drive that could safeguard their organisation, system, or mobile-based data. In Jordan, respondents indicated a precise administration level as to who has what level of access to prevent people from doing things they do not know or doing it accidentally. However, this study pointed out that the plan is not mature enough. Meanwhile, they have training sessions available to introduce the risk assessment plans or concepts and assist employees' level of knowledge and what they need to add to the plan based on that. Another plan in Jordan specifies the action to take in case of a cyber-attack concerning cloud computing. There is also a policy on the confidentiality of big data. Jordanians regularly test and monitor their systems for any suspicious activity to ensure that all data is safe. Another set of policies ensures that a data source's quality, privacy, security, and integrity risk is assessed in four categories. Developing countries like Jordan use cloud computing in various methods (Hammouri and Abu-Shanab, 2020; Yeboah-Boateng and Essandoh, 2014). Most Internet users in these nations face difficulties because of the low adoption rate of CC technology. This study's findings confirmed the significance of factors such as service quality, cost reductions, IT expertise, and a feeling of competence. Users' level of control over their sensitive data, as well as their background in information technology (IT), service costs, and quality of cloud services, all significantly influence the decision to utilise online CC services. Finally, Jordanian users believe that purpose is more strongly influenced by their sense of control over the situation.

Canadian organisations' executive directors define the risk assessment plans, encompassing cloud computing and big data. Like in Jordan and the United Kingdom, the Canadian plans identify potential risks and continuously evaluate them. Canadians are constantly assured that all data are safe with offline backup with regular testing and monitoring for suspicious actions. Some plans are based on data classification and are updated based on observations and testing. This will enable ranking and order of every risk's likelihood, impact, and severity and determine key indicators (KRIs). (Drissi *et al*. (2013) indicate that a risk assessment plan is lacking in most organisations. Since most organisations perceive cloud computing technologies as expensive, they are resistant to design frameworks to achieve their objectives. However, companies in Jordan need to learn to balance risk's possible negative consequences against its associated opportunity's potential benefits. The standard ISO 27005 states that information security risk may benefit from hardware and software flaws, negatively affecting the institution and its processes (ISO, 2018).

These policies and recommendations cover a series of risks and potential threats. The results showcased a wide variety of threats to be mitigated due to effective and planned risk assessment. Firstly, data encryption is done using SSL. There are controls to mitigate data leakage and protect data at rest and in transit. Data loss prevention is achieved through education (educating employees about best practices) and robust cybersecurity controls. The plans state that stored data must be transferred over secure sockets. Also covered is the use of tools to automate the detection of unintended data access, enforce encryption in transit, implement secure keys, and follow certificate management updates. According to Turskis *et al*. (2019), the most common risk assessment tools and techniques used at the industry level are the Failure Modes and Effects Analysis (FMEA) (Baynal *et al*., 2018); Bow-Tie Analysis (Muniz *et al*., 2017); Fault Tree Analysis (FTA) (Giraud and Galy, 2018); Layer of Protection Analysis (LOPA) (Yan and Xu, 2018); and Hazards and Operability Studies (HAZOP) (Taylor, 2017). The most common factors involved in risk management include scanning the business environment, identifying vulnerabilities, and carrying out SWOT analysis. The purpose is to assess the ability of the company to assess risk. This is important because it enables companies to adopt a risk assessment framework that they can manage.

Another plan highlighted in Canada relates to implementing GDPR, processes, protocols, and infrastructure to securely access data. Users must be aware and have some form of training related to cybersecurity and identifying threats. On an ongoing basis, a process is put in place where risks of the inside, outside and third parties are evaluated. Not all parties might take cybersecurity as seriously. Therefore, it is essential to keep evaluating the security posture of all vendors to ensure they are not at risk of a data breach. That also will include monitoring the security posture of all. All network access is monitored to secure all endpoints (firewalls and VPNs). Before introducing CC, organisations should ensure that the appropriate infrastructure is in place to support it. This includes the availability of necessary technologies and the knowledge to run those (Al-Somali *et al*., 2011). An organisation's expertise and awareness of technology are critical for implementation success. If an organisation is aware of technology, it will undertake essential preparations before adoption, increasing the probability of successful integration (Carroll *et al*., 2011). Organisations require individuals familiar with all relevant security processes (McKendrick, 2013). Individuals with negotiation experience will be critical to the organisation's success when negotiating SLAs with cloud providers.

Additionally, they will engage with team members to address any outstanding difficulties. Additionally, they must negotiate with cloud providers if the service goes down or the supplier

does not meet its obligations (McKendrick, 2013; Al-Somali *et al*., 2011). They must possess the necessary expertise to quickly fix any issues with the cloud provider while maintaining a positive connection with the service provider. Negotiating with internal workers is also critical to ensuring that internal change management is successful.

In the UK, secure channels for transmission are highly used, and a well-defined data level of access exists to secure transmission channels. An example is the SSL, security authentication for data transmission. Thus, data in transit is encrypted. Like the other three countries, the plans also cover security training, awareness sessions and training related to cybersecurity and identifying threats. This study found that companies added TLS (Transport Layer Security) protection in the UK when other data confidentiality within their organisation is required and support end-user authentication and access control within applications. Organisations need personnel who know all security protocols (McKendrick, 2013). People who understand cloud computing laws and regulations must ensure that their organisation always follows the law (Chang *et al*., 2018). There is a need for a clear legal framework that guides companies. Companies need to have laws in cloud computing as this gives them patent rights and privacy on their data.

The risk plans in the UK also provide device control since users are used to saving data on phones or tablets, plus real-time auditing and reporting. If there is a data leak, companies can activate a kill switch. Companies then watch the log files for any suspicious actions to identify and track suspicious activity. As a result, organisations require expert project managers to accomplish their targeted outcomes (Dioubate *et al*., 2015). Skills in change management are also required, as cloud computing will force modifications in company operations, the IT department, and even the entire organisation's structure. Scalability in cloud computing necessitates project and capacity management in addition to monitoring tools to determine how much service is needed without paying for extras that are not needed (McKendrick, 2013).

Jordan's risk plan covers a variety of issues. Data that gets stored must be transferred over secure sockets. The use of encryption at all points of the data journey and multifactor authentication was also highlighted. Also, implementing a data loss prevention (DLP) solution where all staff have to follow guidelines is critical. This protects sensitive data from being leaked from endpoints (personal computers, notebooks, portable devices, and servers).

Additionally, the plans include email access control lists, which employ a method known as a deep content inspection to locate sensitive data in emails' text, photos, and attachments. If sensitive data

is discovered, a warning is sent to the administrator, who can check the transfer's legality. Therefore, the policies limit user access privileges to what is needed. In addition, there are regular cybersecurity awareness training sessions, and access rights are revoked immediately if any suspicious activity is detected. Schneir (2010) suggests that security psychology is centred on security as both a sensation and a fact. However, security as a sensation and as a fact is not similar. Schneir (2010) argues that security is contingent on the probability of various risks and the effectiveness of various mitigation measures in mitigating those risks. Security is also a psychological sensation triggered by both risks and measures. Policymakers can use massive data to examine government policies and plans methodically, while businesses mine big data for profit/earnings, useful consumer insights, and marketing shares (Xu and Shi, 2015).

The risk plans Jordan also ensures investments in a managed file transfer system are compulsory for all staff members to follow the communication guidelines with business contacts. Most risk plans are common to firewalls and network access control to secure networks that transmit data against malware and other malicious threats. The organisation requires individuals capable of integrating on-premises and cloud data. Additionally, they will require data analytic expertise to evaluate which data is suited for the cloud and which should remain on-premises for confidentiality and data security reasons (Taivalsaari and Mikkonen, 2015; McKendrick, 2013). This is crucial for financial institutions that handle sensitive consumer data. Cybercrime is a serious threat to the economy; a crime committed using computers and the Internet. Financial services organisations risk cybercrime if they do not develop sufficient measures to protect clients and improve data security.

Another policy raised in Jordan is for all staff members to use VPN when they need to access their files while off-site. Companies do not recommend using public computers and free Wi-Fi in this regard. Risk plans ensure that users have some form of training related to cybersecurity and identifying threats. As a result, the number of individuals accessing sensitive data is constrained, minimising the risk of data leakage. Certain concepts, such as 'Bring Your Own Device' (BYOD), have gained traction over the last decade as more businesses see the cost savings and enhanced staff productivity associated with them (Hamel, 2012). With more companies becoming mobile, there is a need for financial systems to move toward the cloud as organisations migrate IT assets from their data centres to the cloud. Businesses that utilise the cloud must host at least some of their data on the cloud, including mission-critical workloads (Furht, 2010). Training is one of the aspects that contribute to the concept's effective implementation, and as such, organisations must train their employees on how to install software efficiently.

In South Africa, some organisations had no clear and consistent information governance policy. In this instance, it was indicated that data risk classifications become impractical, and the strategy has been avoided for some time. Financial institutions are moving to the cloud by increasing business flexibility, modernising processes, cutting inventory costs, and simplifying and expediting the deployment of infrastructure investment resources (Pourghomi and Ghinea, 2012). The financial services industry handles some of the most sensitive and personal client information. Losing a customer's personal and financial information can devastate the individual and the business. As a result, financial institutions are subject to stringent data storage and security rules, including SOX, GLBA, and PCI DSS, which restrict who has access to and how the firms can retain and utilise customer data.

The development of complex network security rules to assist in the protection of data in transmission, with all employees required to adhere to these policies when connecting to the Internet. Thus, there was a need to always use encrypted channels with proper security protocols for transmission. Companies avoid data leakage by using high standard encryption on stored data and connections, using two-factor authentication, and keeping backups. Validation is done by authorised persons only. For example, only SOC (Security Operation Centre) handles data in transit. The majority of big data is unstructured and unorganised. While MapReduce (Hadoop) technology may be used to collect huge amounts of data, computer science's traditional data acquisition and management techniques must be augmented with management science skills. For example, the organisational strategy of using big data must be addressed before gathering the big data. A large database's basic design and administration should be based on data capacity, value, ethics, ownership, policy, and quality guarantee (Laundon and Laundon, 2018). Big data may play a vital part in making a practical choice with management science. However, big data mining discoveries must be assessed against the user's judgment, as knowledge varies by person and circumstance (Yu *et al*., 2014).

Regular passwords change, and limited data sharing use is also standard practice. Similar to other countries, users must be aware and have some form of training related to cybersecurity and identifying threats. Financial organisations must also consider legal considerations when it comes to data ownership and access. A court order or government officials' request to get data for inquiry raises several complex legal concerns, and financial institutions need to be prepared for these scenarios to maintain the credibility and confidence of their clients (Sharma, 2012). Finally, financial institutions value the information they collect about their customers and their

transactions. To protect the institution's reputation and boost client confidence, data must be safeguarded from leakage, purposeful or unintentional loss (Sharma, 2012).

Canada identified several threats to data security. Canada's data protection legislation comprises numerous federal and provincial statutes. Federal and provincial data protection legislation of general application are included in this set of laws. This will simplify the process for all parties revising their risk plans in light of provincial and federal government instructions. Organisations must notify the privacy commissioner of data breaches depending on the type of information compromised. As an added precaution, the Personal Information Protection and Electronic Documents Act must be adhered to (PIPEDA). The PIPEDA requires organisations covered by the act to seek an individual's permission before collecting, using, or disclosing that individual's personal information. There must be a clear separation between the sensitive and confidential data stored on traditional systems and the cloud-hosted data. Rather than using a public cloud, they might use one for less sensitive information and a private cloud for more critical information (Misra and Mondal, 2011). They will be able to keep their data safe using this. Resources, platforms, applications, and security must be in place for highly essential operations (Sharma, 2012). Organisations should avoid using the cloud for crucial work since service level agreements (SLAs) have not yet been standardised for CC.

In Jordan, some risks highlighted were data loss due to malfunction or negligence with potential client lawsuits. Unprotected data can lead to litigation and expose companies to fines and lawsuits. In addition, unauthorised data access can be reported where legal action will be instituted. However, respondents indicated that there was no data protection law in Jordan. Therefore, it is the responsibility of each company to mention all legal actions associated with such risks. In addition, whenever an organisation needs a solution involving data, they should mention that in their agreements as it will be the reference for future issues. Data and access control, identification management, and intrusion prevention can all be protected using a multi-layered approach incorporating convergent encryption (Chang *et al*., 2016). In the absence of effective security measures and access control standards, cloud computing adoption in organisations suffers (Oliveira *et al*., 2014). Most researchers have concluded that security predicts users' propensity to use cloud computing (Qasim and Abu-Shanab, 2015).

Data loss, corruption, or disclosure in South Africa leads to legal issues. Therefore, organisations in South Africa concurred with Jordan that there is a need to introduce some type of staff training to cover data-related risks more. In addition, South Africa introduced POPIA in July 2021,

protecting data subjects from security breaches, theft, and discrimination. Security measures recommended by POPI Bill sections 19 to 22 must be followed when personal information is processed by service providers and organisations (Netshakhuma, 2019). Protecting the personal information as instructed in section 19(2) by assigning responsibilities such as identifying and assessing risks, developing and maintaining appropriate safeguards, and monitoring the efficacy of safeguards (Netshakhuma, 2019). A written agreement between an organisation and its service provider should be formed to ensure that personal information is protected under the POPI Bill's Section 21.

Hence, every company must adhere to POPIA's requirements for protecting personal information. A private cloud may be the most cost-effective solution for large organisations with their own data centre. They will have more control over their cloud infrastructure, and their confidentiality and security will be protected. When negotiating a cost with a service provider, it is crucial to consider how many resources an organisation utilises. In the long run, this will help them save costs. The data lock-in issue should be considered because it will allow the organisation to renegotiate with the service provider early. This will also assist in resolving issues with integration. The lack of adaptability of the user interface is still another crucial consideration. This should be included in the organisation's selection criteria for the service provider. It is important to address implementation concerns to develop strategies to handle them if and when they happen. To limit the perceived risk related to CC, organisations must conduct a risk assessment and governance procedures (Zissis and Lekkas, 2012; Armbrust *et al*., 2010).

The General Data Protection Regulation (GDPR) is the Data Protection Act 2018 in the UK (GDPR). According to GDPR, citizens have a right to know what information the government and other organisations collect about them. Citizens have a right to request that data processing be terminated or restricted. Any illegal use of sensitive information may result in legal action. Each individual (client) has a right to know what data an organisation saves or uses, as well as the level of security the organisation's systems provide for their data.

Individuals have the right to request access to their personal information for any reason. Regulations, legislation, and jurisdictional concerns should be addressed at this stage, as they will have an impact on the success of the adoption of cloud computing. If there are regulations, legislations, or jurisdictional issues that can assist in regulating the computing cloud, organisations should be aware of them and implement plans to ensure compliance with those legislations and regulations (Moonasar and Naicker, 2020; Netshakhuma, 2019). Jurisdictional difficulties should

also be thoroughly investigated since they will impact their cloud computing adoption. There may be legal, regulatory, and legislative differences between countries where the cloud provider's data centre is located and those where the user's data resides. The organisations must understand which laws, regulations, and legislation apply and how those laws influence who controls information.

6.7 Conclusion

As the results of this study show, the UK is leading primarily in all aspects among the other three, and that is a result of different elements, such as being the most advanced and the biggest cloud market in Europe and the large majority of UK businesses understand the worth of investing in ICT.

Market overview:

- There are more than 100,000 software registered businesses.
- It is the second-largest ICT market in ICT investing per head (the United States is number 1).
- London is the second most connected location for technology, just behind Silicon Valley.

According to the American International Trade Administration report, cloud computing supplies substantial development possibilities: Almost all UK software providers use the cloud, and opportunities are available in both the public and private sectors for cloud-related businesses. In 2020, the UK will have made close to $12 billion from public cloud services, which dominates the industry. With Amazon Web Services (AWS) being ahead of other competitors, Microsoft Azure and Google Cloud are the other market leaders. While some IT demand has decreased due to the pandemic, particularly in standard day-to-day services, demand for public cloud solutions has surged in the UK technology business. There is a clear divide between large and small UK service providers in cloud adoption, although IaaS products are gaining endorsement among SMEs. Certainly, UK services are increasingly accepting a cloud-first strategy, with an increasing number of businesses preparing for a day when most of their IT will be shifted to the cloud.

Because of the pandemic, the continuous transformation has supported the UK's growing need for public cloud services for remote working. Cloud solutions are being utilised and considered by enterprises across the globe.

The UK government has also been a vocal proponent of cloud computing. For example, the UK Government's Cloud (G-Cloud) programme is revolutionising government IT procurement. The

G-Cloud architecture allows governmental agencies to purchase asset-based cloud services via a government-approved electronic database called the Digital Marketplace. This accelerated procurement approach is consistent with the UK government's "Cloud First" policy and is a critical component of the government's objective of managing a "cloud-native" digital platform.

The UK Government is still exploring new and cost-effective ways to benefit from and utilise big data, as well as integrate big data sets into a multi-structured data ecosystem in the field of big data. As a result, linking big data has risen to the top of the priority list. To succeed, the government needs in-depth industry knowledge, consumer and regulatory insights, and multidisciplinary expertise combining technological, analytical, Promotional, and Legal disciplines to produce unique concepts and accelerate possibilities.

The Canadian government launched the Cloud Adoption Strategy. They saw a surge in the number of Canadian organisations contributing to cloud computing and the necessity for a comprehensive plan encompassing all technology areas. In terms of risk assessment, this method shows how such risks can be systematically handled while allowing departments and groups to respond according to their threat tolerance adoption. Cloud computing in Canada closely tracks global developments, with all big global corporations reappearing in the Canadian market. However, several specialised Canadian cloud start-ups attempt to deliver custom-tailored cloud services in Canada rather than competing with the major giants such as Amazon, Microsoft, and Google.

Canada is moving in the right direction in terms of big data implementation and integration, as it recognises that applying big data to core sectors means lowering prices and lowering environmental impact, which will help to reduce the amount of waste and help the environment, pollution, and land degradation while creating significant new economy employment that will define Canada's prosperity over the next century.

Although Canadian colleges create a disproportionate share of the world's big data professionals, Canada scores low in terms of big data possibilities. Because of the country's low population density, the consumer-related potential is lower than predicted. Many professionals depart for the United States since finding employment in Canada is difficult.

Moving to Jordan, it is obvious from the results that Jordan is on the correct track toward more big data and cloud computing acceptance and implementation for a variety of reasons:

- The government encourages corporations to engage in this industry through business opportunities and assistance packages such as tax breaks, free zones, and subsidies for small and medium-sized businesses (SMEs).
- The ministry of commerce launched several projects related directly or indirectly to big data or cloud computing.
- More colleges and universities offer courses and degrees linked to big data and cloud computing technology.
- Jordan's central location assists the ICT sector and government in taking the necessary steps to position it as a regional digital centre.

While the government is on pace to provide more digital services, various obstacles and hurdles are predicted, including:

- Legal obstacles: Jordan has a defined data protection legislation, and more legalisation is needed to provide a safe, secure, and trustworthy environment.
- Strategy obstacles: Complicated data and shared access to records will need clear regulations; the Jordanian government is developing such a policy.
- Technological barriers: New technology utilisation may necessitate the purchase of new technological equipment, which will demand a dedicated budget; however, some government funds may be available to assist cover some of these expenses.
- Organisational barriers: SMEs still see this as a major challenge, owing to the need for new qualified hires and new infrastructure, but they should also see it as a great opportunity to discover and expand great opportunities that big data technologies can provide, which will be very beneficial to their development and keeping up with economic growth in the economy.

South Africa has one of the largest ICT industries in Africa, leading in several areas, and it will only be a matter of time until enterprises see big data applications and cloud computing as the new technological norm. Big data and cloud computing are among the initiatives included in the government's aim to teach more than one million young people by 2030.

The POPIA is another example of how South Africa is setting the path for the future of technology by controlling and legalising data access. However, this is also a barrier since, according to data specialists, this act will not allow big data to grow because personal information is obtained exclusively for a specified purpose. Furthermore, this study's findings indicate that further work is needed to encourage using big data and cloud computing technologies. Another issue is the

scarcity of talent and data scientists since it is still difficult for young people to obtain high-quality education and training. Furthermore, infrastructure must be established, restructured, and new systems developed to use and manage big data to reap the full benefits of big data.

6.8 Summary

This chapter discussed the results offered in Chapter 5 and offered a comparative analysis of big data risk assessment techniques based on cloud computing technologies in four countries: Canada, Jordan, South Africa (SA), and the United Kingdom (UK). The discussion summarised the results and used the literature to confirm or refute the findings. Furthermore, this chapter comprehensively evaluated the findings and how the results covered the research objectives in Chapter 1.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

7.1 Introduction

This study compared cloud-based big data risk assessment approaches in Canada, Jordan, South Africa (SA), and the United Kingdom (UK). These four countries illustrate different continents and how they handle big data and cloud computing threats. This study stated that big data and cloud computing interact more since the latter provides an optimal online working environment. According to another argument, big data is characterised by large-scale data's quick and analytical utilisation. The report said the volume of data created by present devices is staggering. Every internet-connected device creates a digital trace.

This study defined risk as the possibility of a threat entity exploiting a flaw that affects firms' processes and services. Risk management involves the activities and procedures that help organisations in the four countries mitigate or eliminate risks and attain their goals. Risk assessment is a key element in risk management, too. An organisation can recognise hazards and assess their likelihood, consequences, and mitigating actions through risk assessment. As explained, risk assessment aims to identify and implement effective measures to limit or eliminate risks. Information systems, software, hardware, management, and environmental security are important for risk assessment.

Big data and cloud computing are being produced, expanded, and consolidated. Due to its open environment, little user control, and high level of automation, cloud services have security and anonymity issues. Big data security is important in cloud computing because it is employed in fundamental company operations and procedures.

This study compared cloud-based large data RA protocols in Canada, Jordan, South Africa, and the UK. This literature review identified big data and cloud computing risk assessment. The literature on risk assessment, cloud computing, and big data was analysed. A deeper theoretical understanding of these mechanisms helped risk assessment minimise huge data in cloud services. The theoretical objectives included outlining the evolution of research on risk assessment involving big data in cloud services, compiling the main results of studies on risk assessment in big data, and presenting a theoretical framework aggregating these studies and compiling best practices of risk assessment in big data and cloud computing from the scientific literature. The framework was developed based on qualitative analysis and triangulation method where the stated

findings were checked for validity, and due to repeated occurrences, statements depicting risk assessment methods in the respective nations of the respondents were undermined.

Furthermore, the issues and threats revealed through the analysis were synchronized with nation-wise inferences drawn from various sources. The results helped to create a systematic structure of risk assessment methods in use in the four mentioned countries for big data using cloud services. The analysis highlighted the big data vulnerability in cloud platforms and the emphasis on risk assessment modes and was conducted to decrease the threat patterns and risks.

This qualitative study followed both theoretical and empirical interpretive research paradigm objectives. The empirical aims were to study the important aspects of risk assessment in Canada, Jordan, South Africa, and the UK using big data on cloud services and to provide managerial and policy suggestions utilising data from the same nations.

The interpretative research paradigm was well-suited for this study since it offered a deeper understanding of the issues and the relationships between constructs. Interpretivism helped cover a research vacuum because RA in big data cloud services is understudied. This interpretive study produced more meaningful and grounded outcomes by analysing issues in context. Cloud computing and big data made the results more relevant, leading to better policy and management recommendations. It used grounded theory. This study used quantitative and qualitative methodologies to triangulate findings by simultaneously focusing on both information forms and previous literature. Four countries' specialists completed a standardised questionnaire to evaluate big data use and storage. Quantitative data supplemented qualitative data in the analysis. Atlas.ti was used for qualitative analysis, which followed a four-step process to address the study's inquiries: Define the domain, collect data, assess the evidence, and draw conclusions about trends. SPSS and Excel iterations were used to analyse descriptive statistics in quantitative analysis. These descriptive statistics describe big data, cloud computing, and risk assessment management knowledge levels.

7.2 Major findings

7.2.1 Research objective 1: outlining the evolution of the research on RA involving big data in cloud services

7.2.1.1 Evolution of big data

According to this study's literature evaluation, big data has been a topic of discussion since 2001. Big data lacks a standardised definition and is interpreted differently by academics and businesses. Big data is defined in the literature review as large data sets derived from various digital sources such as instruments, sensors, Internet transactions, email, video, click streams, and other digital sources currently accessible or developed. Furthermore, big data is defined by high velocity, volume, and diverse information assets that require innovative processing forms to improve decision-making, insight finding, and operational efficiencies. There are diverse opinions on big data among academics, businesses and policymakers. Big data is a term used to describe complicated, diverse, variable, and potentially valuable data collection. In the digital age, policymakers see it as a new strategic resource that spurs innovation and reshapes human productivity and way of life.

Big data became a new trend in the internet and IT business in 2009. Initially, big data uses were mostly in the internet industry. Also, data on the internet doubles every two years, according to the literature. Big data is becoming increasingly important to most global internet companies' competitiveness.

7.2.1.2 Evolution of cloud computing

The study found that big corporations like Google and Amazon started embracing cloud computing around 2006. Cloud computing is a new paradigm in which individuals access software, computing power, and files via the Internet rather than their Desktop computers. A technical evaluation traced the term's origins to the late 1990s at a Houston-area commercial park. At the time, the Netscape Web browser was the hot new technology. A group of technology executives at Compaq Computer was planning the future of the internet industry, called cloud computing. Their foresight was precise. Not only would all commercial software be Web-based, but so would cloud computing-enabled applications like consumer file storage. George Favaloro, a Compaq marketing officer, and Sean O'Sullivan, a young technologist, predicted cloud computing. Compaq capitalised on the opportunity to sell servers to Internet service providers, but O'Sullivan's investment led to

disillusionment and bankruptcy. The phrase has become controversial due to billions of dollars in IT spending. Programmers were outraged when Dell tried to trademark cloud computing in 2008. Others, including IBM and Oracle, have been accused of cloud washing or misusing the phrase to represent outdated product lines. Like Web 2.0, cloud computing has become an industry buzzword that many in the industry despise yet cannot escape.

Finally, according to the literature, information technology trends move computation and data from desktop and portable computers to large data centres. Thus, cloud computing allows anyone to use programs, applications, and computing resources on-demand, anywhere, anytime. As a result, cloud computing is a set of services that do not require end-user awareness of the system's physical location or configuration.

7.2.2 Research Objective 2: compiling the main results of the studies on RA in big data and presenting a theoretical framework aggregating these studies.

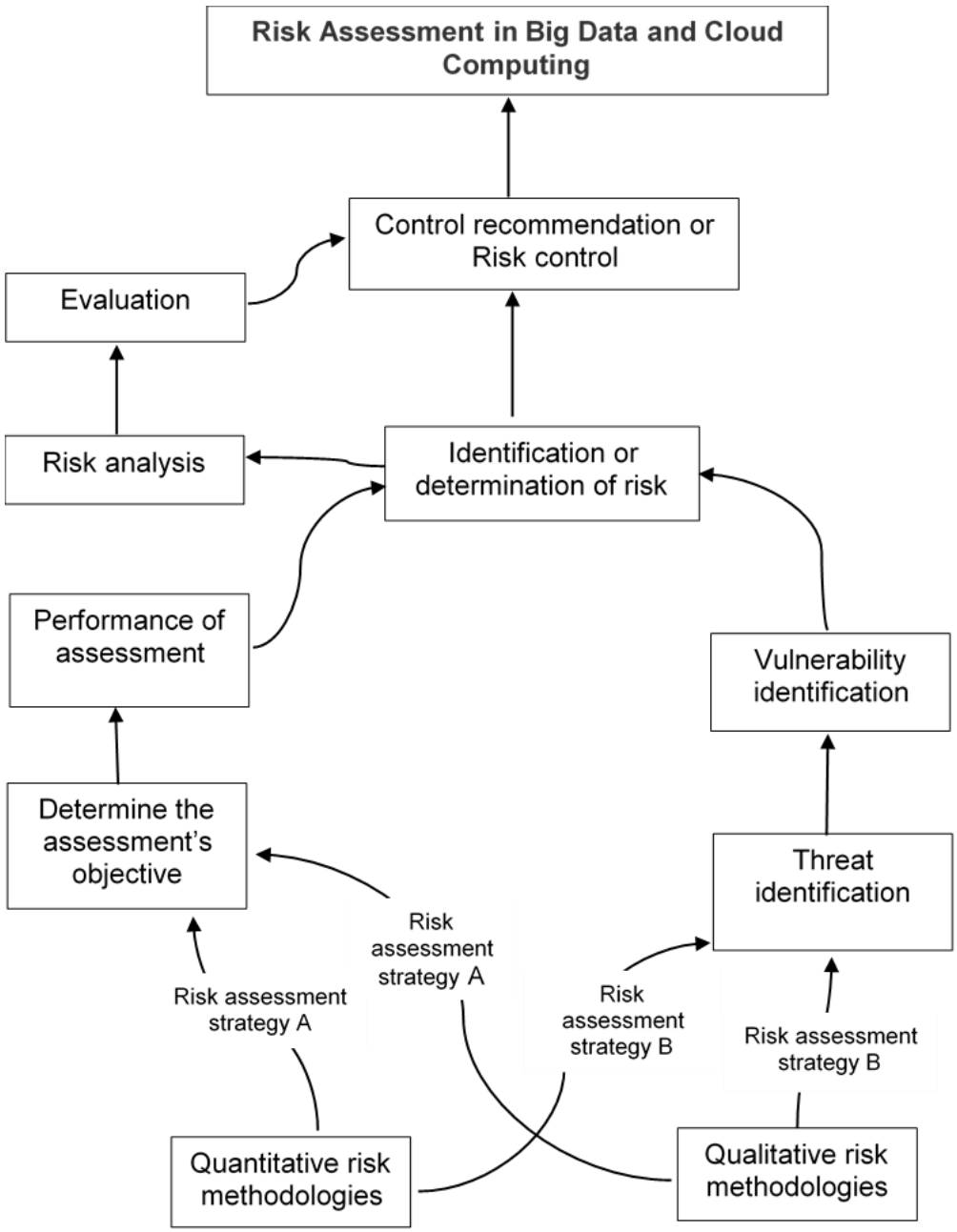

Figure 7.1: Theoretical risk assessment framework for cloud computing and big data

This study found risk assessment to be an essential step in risk management, which defines suitable control methods for reducing or eliminating risks. Figure 7.1 shows that the risk assessment starts with deciding whether to adopt the qualitative or the quantitative methodology. Quantitative risk is employed in many well-developed businesses but not in information technology. The qualitative risk analysis model is employed when expressing risk numerically is challenging. Rather, it details the possibility of results. The theoretical framework was formulated and developed through an understanding of concepts relevant to the study's aim, and guided by the existing literature, articulation of theoretical assumption was made. By virtue of revelations on the isolated and important variables, the salient findings were noted down. Furthermore, the key points derived and

the relationship between them were understood, stated and with reference to research objectives, it was constructed

7.2.2.1 Risk assessment strategy A

As Figure 7.1, the risk assessor may choose the risk assessment strategy A. As shown in Figure 7.1, this strategy entails six steps that the risk assessor must undertake below:

1) Determine the assessment's objective. This step defines the information system data, hardware and software assets.
2) Performance of assessment. To improve the evaluation plan, the assessor must, at this stage, determine the assessment process and select proper assessment methods and tools.
3) Identification of risk. This step identifies the assets within the assessment scope, the threat, asset vulnerability, and existing security measures.
4) Risk analysis. This step involves analysing the possibility and consequences of threats and vulnerabilities.
5) Evaluation. At this stage, the assessor must evaluate the result to prepare a risk assessment report to be assessed by an expert.
6) Risk management. This final step entails implementing suitable risk transfer mechanisms and avoiding or mitigating risk to maintain system control.

7.2.2.2 Risk assessment strategy B

Alternatively, the risk assessor may choose risk assessment strategy B. Strategy B is quicker than strategy A. As indicated in Figure 7.1, the risk assessor must undergo four steps as below.

1) Identification of potential threats. The initial step is to identify all of the system's vulnerabilities. In addition, it makes it possible to determine the system's serious risk sources.
2) Identification of vulnerabilities. This step is intended to generate a list of system and security breaches (flaws or weaknesses) that possible threat sources may exploit.
3) Identification (assessment) of risk. This stage aims to determine the level of risk associated with the system.
4) Proposal of control. Controls that are in line with the organisation's operations and goals are suggested in step four, which is the last phase. The recommended controls are designed to lower the system's level of risk.

7.2.3 Research objective 3: explore the critical features of RA in Canada, Jordan, South Africa, and the UK environments involving big data on cloud services.

Several features are noted in the environments involving big data. This study found that most countries planned to run their big data on cloud systems except for South Africa, whose respondents indicated that traditional storage would continue into the near future. However, most reflected novice and intermediate knowledge of big data, particularly in South Africa. Only Jordan and the United Kingdom have expert knowledge about big data, while Canada is spread between novice and advanced expertise. Many reflected the intermediate and advanced levels of expertise in cloud computing. Only South Africa exhibited a novice level of knowledge in this aspect. In addition, South Africa was the only country challenged with risk assessment. South Africa and Jordan tend to run their big data applications on the traditional storage system. At the same time, Canada and the United Kingdom were top in managing the big data applications on the cloud.

This study revealed that big data technologies help organisations cut expenses by saving, analysing, and assessing massive amounts of data. More accurate data, motivating a new generation of employees, cost savings, and control of digital reputation are all benefits of big data technologies. The study also found that big data technologies are related to reliability, better data control, high availability, automated updates, easy data backup and recovery and scalability.

While most countries agree that big data technologies are being implemented in their organisations, the disadvantages vary from country to country. For example, in the UK, the disadvantages were high cost and a lack of required experience, which was linked to the complicated data infrastructure. It also takes a long time in the United Kingdom to fully implement the technology, thereby raising costs. In South Africa and Jordan, implementation costs were highlighted as a challenge and finding new skilled staff to take the task. Canada's Personal Information Protection and Electronic Documents Act imposes some restrictions on electronic documents, increasing the potential for data breaches. In Jordan, risks were data loss due to malfunction or negligence with potential client lawsuits. Data loss, corruption, or disclosure in South Africa leads to legal issues. In the United Kingdom, individuals have the right to stop or restrict data processing. They can also take legal action against any unauthorised use of protected data.

The study also noted several aspects to mitigate the risk by organisations. In Canada, controls are in place to mitigate data leakage and protect data at rest and in transit through educating employees

about best practices and robust cybersecurity controls. In the United Kingdom, companies have many backups in the event of accidental deletion of data. There is also high usage of secure channels for transmission and a well-defined data level of access to secure transmission channels. In Jordan, data that get stored is transferred over secure sockets. Thus, to protect data in motion based on its sensitivity, investments in a managed file transfer system are compulsory for all staff members to follow the communication guidelines with business contacts. Organisations in Jordan also highlighted encryption at all points of the data journey and the use of multifactor authentication. The study found organisations with no clear and consistent information governance policy in South Africa. In this instance, this study indicated that data risk classifications become impractical, and the strategy has been avoided for some time.

7.2.4 Research Objective 4: develop managerial and policy recommendations on RA on big data in cloud computing environments using data from Canada, Jordan, South Africa, and the UK.

There are several implications from the findings of this study. Organisations need to empower a new generation of employees and improve reliability, data control, availability, automated updates, and easy data backup, recovery, and scalability. The recommendations for risk assessment are reflected in the framework in Figure 7.2. Several themes are proffered as recommendations are described as technological, organisational and environmental.

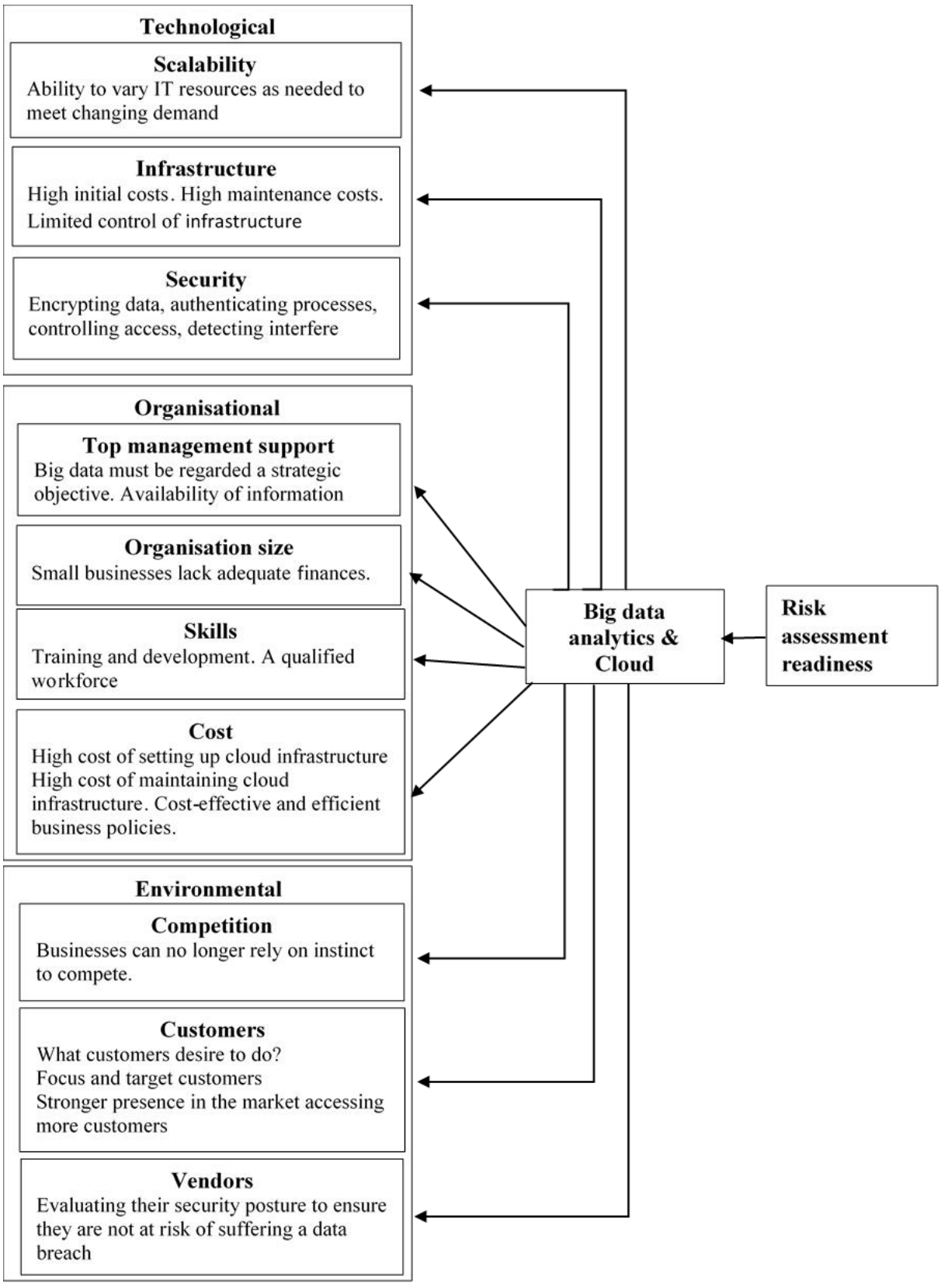

Figure 7.2: Recommendation framework for risk assessment in cloud computing and big data

7.2.4.1 Technological aspects of risk assessment and big data analytics

(a) Scalability

The organisation must have the capacity and ability to vary IT resources as needed to meet changing demands. Recent disruptions from shocks such as the COVID-19 pandemic have reinforced the notion that the cloud must not be treated as a future aspiration. It is a commercial necessity. Cloud computing is a key facilitator of modern digital technologies that open the door to new products, services, and revenue streams, promising more efficiency and accelerating innovation. However, a rushed move from old systems without a defined strategy may cost the firm, leaving existing outdated applications racking up usage and expenditures.

(b) Infrastructure

The cloud system and cloud computing are generally associated with high initial costs. Therefore, organisations need to conduct a cost-benefit analysis before choosing the vendor. Even when cloud computing is being utilised, there are often associated with high maintenance costs that affect particularly small businesses. Organisations also need to manage infrastructure control, especially outsourcing, as control can easily be lost and data security risk.

(c) Security

There is a need for continuous improvement in encrypting data, authenticating processes, controlling access, and detecting interference. Thus, organisations need to invest in infrastructure that enhances encryption at all points of the data journey and ensure the usage of secure channels for transmission.

7.2.4.2 Organisational aspects of risk assessment and big data analytics

(a) Top management support

It was found that organisations must regard and elevate big data and cloud computing as strategic objectives. This way, they can ensure that all the business processing systems use the infrastructure and benefit from the economies of scale in terms of initial costs and costs of maintenance. In addition, the management must ensure that information is available at all levels regarding the infrastructure, its benefits and advantages for individual personal development.

(b) Organisation size

This study found that small businesses lack adequate finances and are the most vulnerable to cloud computing and the effects of big data. The government can help small enterprises by subsidising cloud services and big data technologies. The smaller businesses use the cloud system and big data, the better growth prospects. In the long term, more people are employed, and the tax base increase for the government.

(c) Skills

This study found training and development to be imperative. Organisations need to educate employees about best practices and robust cybersecurity controls. This will ensure a qualified workforce using big data and infrastructure maintenance. Qualified personnel also reduce the rate of outsourcing, therefore saving on costs.

(d) Costs

The study found that there are associated high costs of setting up cloud infrastructure and the cost of maintaining cloud infrastructure. Organisations must find a way to reduce big data storage costs, particularly in the cloud system. In addition, ways to benefit from economies of scale, such as linking all the business processes to the cloud system, are imperative. Small businesses can also consider sharing the platform with similar organisations, but they have to ensure they do not compromise their data security. Otherwise, they can easily lose competitiveness.

There are also opportunities and benefits in ensuring cost-effective and efficient business policies. For example, organisations need to invest in big data and cloud storage systems to help discover cost-effective and efficient business policies by providing more accurate data. Policies force the entire organisation to harmonise its activities towards cloud computing and big data, reducing costs.

7.2.4.3 Environmental aspects of risk assessment and big data analytics

(a) Competition

Findings indicated that businesses could no longer rely on instinct to compete. Instead, organisations need to keep watch on their competitors and adapt accordingly. Cloud computing and big data can edge the competitors. Still, the organisations need to ensure that the service

providers are innovative and continuously strive to improve the service for this to happen. Big data is moving quickly, and for organisations to survive, they need to keep their systems on their toes.

(b) Customers

The study was clear that customers matter. The organisation needs to be aware of what customers desire to do and focus on and target its customers. This way, they can respond to the customers' needs accordingly. A stronger presence in the market will also ensure that the organisations access more customers by taking advantage of big data and cloud computing.

(c) Vendors

The study found that vendors are vital in big data and cloud computing risk assessment. Therefore, it is imperative to evaluate their security posture to ensure they are not at risk of a data leak. The consequences of the data leak are born by the organisation and not the Vendor. Therefore, it is imperative that correct vetting is done before choosing the vendor.

7.3 Contribution to the knowledge

This study's findings will help cloud service providers to improve risk assessment during the construction, deployment, and operation of cloud-integrated big data services. In addition, a comparative analysis conducted in this study will allow the development of strategies for improving the risk management of the cloud services companies offer. Small, medium, and large companies, in one way or another, will use these services to monitor their resources, carry out planning activities, monitor their processes, make payments, buy supplies and sell their product online. This study identified critical elements to be considered when safely deploying these services.

This research study generated new knowledge that contributed to cloud computing, big data, and risk assessment literature by understanding the interactions between these concepts. As a result, a better understanding of these concepts is expected to facilitate the scholarly debate and discussion on risk assessment in the cloud and big data environments, taking more and more relevance in academic discussions.

7.4 Study limitations

One limitation of this study was finding professionals to participate and accept the invitation to be interviewed. The most difficult were Jordan and South Africa, as the researcher had to contact too

many individuals, and it was not a success in most cases. To mitigate this, the researcher followed two different plans. The first was creating a Google Form with the questions from the questionnaires and posting it on different social media platforms related to the study subjects. The link was posted to a LinkedIn profile, and network connections shared the post; the second contingency plan involved the researcher contacting a firm in the USA that helped share the questionnaires on their platform and target a specific group.

A further limitation was that due to time limitations and to schedule a time that works for all in different time zones. Then with the pandemic, potential respondents became unwilling to spare half an hour to answer the questions. Whenever the researcher had confirmed respondents and could not have an interview, the researcher offered other options, like recording answers and submitting them in an audio or video format.

Finally, due to the Covid-19 pandemic, it was also impossible to conduct face-to-face interviews, and for the most, potential respondents did not want to participate in a video call. Covid-19 prompted most to drastically alter daily routines over the last few months, including avoiding people as much as possible, working from home rather than in the office, and virtually attending school. Again, the researcher offered other options for respondents to answer the questions, like audio or video recording or typing the answer in any text format.

7.5 Areas for future research

The phenomenon of risk assessment for cloud computing and big data is becoming critical to organisations worldwide. This thesis contributes to knowledge creation in risk assessment for cloud computing and big data, mainly focusing on Canada, Jordan, South Africa, and the UK. This study recognises that new technology is no longer a luxury but a necessity. Some countries have neglected using the 4th IR in their risk assessment plans. Issues such as the role of big data and how to capitalise on it for competitiveness have been neglected. However, some of the cloud computing and big data concepts described in this thesis will require more development and extension to be fully realised. The researcher recommends the following areas for additional investigation, believing that they will contribute to the body of knowledge that will help organisations better identify risks when implementing cloud computing and big data technologies.

First and foremost, the researcher asserts that additional research is required to fully comprehend the evolution of analytics and information management in cloud-based analytics. Secondly, the study found that cloud computing and big data were being used at a low rate in less developed

countries such as Jordan and South Africa, which is concerning. Thus, there is a need to research the adaptation and strategies to improve efficiency and mitigate risks in cloud computing and big data. Thirdly, this study found that the fear of cyber security threats is one of the main stumbling blocks to adopting cloud computing and big data. In this regard, the study suggests future research in formulating strategies and techniques to deal with privacy and security concerns. However, the areas for future research are not limited to the points mentioned earlier, so long as it aims to transform the cloud system from merely a data management platform to a scalable data analytics platform.

7.6 Summary

This study aimed to provide a comparative analysis of big data risk assessment techniques based on cloud computing technologies in four countries: Canada, Jordan, South Africa (SA), and the United Kingdom (UK). The study achieved this by reviewing existing literature to understand better research methodologies, big data, and cloud computing as required to answer the first two theoretical research questions. This study analysed the primary data across the four countries to answer the empirical research questions. The results discussed in Chapter 6 indicate where the four countries stand regarding big data and cloud computing and how they can manage the risk associated with both technologies.

The UK and Canada both showed an advanced level of adoption and recognised the importance of big data technologies and cloud computing. Both Canada and the UK, especially the UK, are investing big in developing new strategies to govern the technologies and mitigate the risks, while Jordan and South Africa need more regulations and adoption for both technologies; still, the future is promising. Both Jordan and South Africa promote big data technologies among the young generation by introducing new funded courses and new undergraduates and graduates' courses. In addition, strategies have been developed and adopted to open new investment opportunities in these developing countries. Overall, there is an intrinsic need to identify the advanced risk assessment methods concerning confidentiality, privacy and integrity of big data, and by regular and planned series of assessments and monitoring, and enabled and integrated infrastructure for big data using the cloud can be created. With rigorous technological advancements, the threats and risks have increased, and therefore organisations and cloud service providers should enhance their risk assessment modules on severity level so that the cloud computing environment can be made safe, trustworthy and controllable.

REFERENCES

Abdalkafor, A., Abdalqahar Jihad, A. & Tariq Allawi, E. 2021. A cloud computing scheduling and its evolutionary approaches. Indonesian Journal of Electrical Engineering and Computer Science, 21(1):489-493.

Abdel-Fattah, M.A. 2015. Grounded theory and action research as pillars for interpretive information systems research: a comparative study. Egyptian Informatics Journal, 16(3):309-327.

Abdullahi, M., Ngadi, M.A., Dishing, S.I., Abdulhamid, S.M. & Ahmad, B.I. 2019. An efficient symbiotic organisms search algorithm with chaotic optimization strategy for multi-objective task scheduling problems in cloud computing environment. Journal of Network and Computer Applications, 133:60-74.

Abioye, T., Arogundade, O., Misra, S., Adesemowo, K. & Damaševičius, R. 2021. Cloud-Based Business Process Security Risk Management: A Systematic Review, Taxonomy, and Future Directions. Computers, 10(12):5-23.

Addakiri K., Khallouki H., Bahaj M. 2020. Big data for context-aware computing. In: Ezziyyani, M. ed. Advanced intelligent systems for sustainable development (AI2SD'2019). AI2SD 2019. Advances in Intelligent Systems and Computing, vol 1105. Springer, Cham. https://doi.org/10.1007/978-3-030-36674-2_18.

Agarwal V., Kaushal A.K., Chouhan L. (2020) A Survey on Cloud Computing Security Issues and Cryptographic Techniques. In: Shukla R., Agrawal J., Sharma S., Chaudhari N., Shukla K. (eds) Social Networking and Computational Intelligence. Lecture Notes in Networks and Systems, vol 100. Springer, Singapore. https://doi.org/10.1007/978-981-15-2071-6_10.

Ahmed, E., Naveed, A., Gani, A., Hamid, S.H.A., Imran, M. & Guizani, M. 2019. Process state synchronization-based application execution management for mobile edge/cloud computing. Future Generation Computer Systems, 91:579-589.

Ahmad, W., Rasool, A., Javed, A. R., Baker, T., & Jalil, Z. (2022). Cyber security in IoT-based cloud computing: A comprehensive survey. *Electronics (Switzerland)*, *11*(1), 1–34. https://doi.org/10.3390/electronics11010016

Alali, M., Almogren, A., Hassan, M.M., Rassan, I.A.L. & Bhuiyan, M.Z.A. 2018. Improving risk assessment model of cyber security using fuzzy logic inference system. Computers & Security, 74:323-339.

Albelaihi, A. & Khan, N. 2020. A brief study of top benefits and hindrances to cloud computing adoption in Saudi Arabia. Journal of Information Technology Management, 12(2):107-122.

Ali, H.A. 2015. Cloud computing security: an investigation into the security issues and challenges associated with cloud computing, for both data storage and virtual applications. International Research Journal of Electronics and Computer Engineering, 1(2):15-20.

Almalki, S. 2016. Integrating quantitative and qualitative data in mixed methods research—challenges and benefits. Journal of Education and Learning, 5(3):288-296.

Al-Marsy, A., Chaudhary, P. & Rodger, J.A. 2021. A model for examining challenges and opportunities in use of cloud computing for health information systems. Applied System Innovation, 4(1):4-18.

Alnemr R. et al. (2016) A Data Protection Impact Assessment Methodology for Cloud. In: Berendt B., Engel T., Ikonomou D., Le Métayer D., Schiffner S. (eds) Privacy Technologies and Policy. APF 2015. Lecture Notes in Computer Science, vol 9484. Springer, Cham. https://doi.org/10.1007/978-3-319-31456-3_4

Alrawais, A. 2021. Parallel Programming Models and Paradigms: OpenMP Analysis. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). (April, 8).

Al-Rousan, T. & Hashish, N. 2016. Resistance factors influencing the adoption of cloud computing in middle east government sector. International Journal of Computing Academic Research, 5:215-219.

Alshammari, A., Aldribi, A. Apply machine learning techniques to detect malicious network traffic in cloud computing. Journal of Big Data 8, 90 (2021). https://doi.org/10.1186/s40537-021-00475-1.

Al-Sharafi, M.A., Arshah, R.A. & Abu-Shanab, E.A. 2017. Factors Influencing the

Continuous Use of Cloud Computing Services in Organisation Level. *Proceedings of the* International Conference on Advances in Image Processing. (August, 25).

Al-Somali, S.A., Gholami, R. and Clegg, B., 2008, September. Internet banking acceptance in the context of developing countries: an extension of the technology acceptance model. In European Conference on Management of Technology (Vol. 12, No. 9, pp. 1-16).

Altaee, M. & Alanezi, M. 2021. Enhancing cloud computing security by paillier homomorphic encryption. International Journal of Electrical and Computer Engineering, 11(2):1771-1779.

Andreadis, G., Fourtounis, G. & Bouzakis, K.-D. 2015. Collaborative design in the era of cloud computing. Advances in Engineering Software, 81:66-72.

Araz, O.M., Choi, T.M., Olson, D.L. and Salman, F.S., 2020. Role of analytics for operational risk management in the era of big data. Decision Sciences, 51(6), pp.1320-1346.

Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. and Zaharia, M. 2010. A view of cloud computing. Communications of the ACM, 53(4), pp. 50-58.

Association for Information Systems (AIS). 2013. Code of Research Conduct

Azumah, K.K., Kosta, S. and Sørenson, L.T., 2018, December. Scheduling in the hybrid cloud constrained by process mining. In 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom) (pp. 308-313). IEEE.

Azumah, K.K., Sørensen, L.T., Montella, R. and Kosta, S., 2021. Process mining-constrained scheduling in the hybrid cloud. Concurrency and Computation: Practice and Experience, 33(4), pp. 2-18.

Backialakshmi, B. & Sumalatha, V. 2020. a novel approach for cloud-based e-learning system. http://www.ijstr.org/final-print/apr2020/A-Novel-Approach-For-Cloud-based-E-learning-System.pdf Date of access 4 Jun. 2019.

Baesens, B. 2014. Analytics in a big data world: the essential guide to data science and its

applications. Hoboken, New Jersey: John Wiley & Sons, Inc.

Bai, H. 2014. Overview of Cloud Computing. Zen of Cloud. (August, 12):18–35.

Baig, R., Freitag, F., Khan, A.M., Moll, A., Navarro, L., Pueyo, R. and Vlassov, V., 2015, October. Community Clouds at the Edge deployed in Guifi. net. In 2015 IEEE 4th International Conference on Cloud Networking (CloudNet) (pp. 213-215). IEEE.

Banchhor, C. and Srinivasu, N., 2021. Analysis of Bayesian optimization algorithms for big data classification based on Map Reduce framework. Journal of Big Data, 8(1), pp.1-19.

Baskerville, R. and Myers, M.D., 2004. Special issue on action research in information systems: Making IS research relevant to practice: Foreword. MIS quarterly, pp.329-335.

Baynal, K., Sari, T. & Akpinar, B. 2018. Risk management in automotive manufacturing process based on FMEA and grey relational analysis: a case study. Advances in Production Engineering & Management, 13(1):69-80.

Belgaum, M.R., Alansari, Z., Musa, S., Alam, M.M. and Mazliham, M.S., 2021. Role of artificial intelligence in cloud computing, IoT and SDN: reliability and scalability issues. International Journal of Electrical & Computer Engineering (2088-8708), 11(5):4458-4470.

Bello-Orgaz, G., Jung, J.J. & Camacho, D. 2016. Social big data: recent achievements and new challenges. Information Fusion, 28:45–59.

Bermani, A.K., Murshedi, T.A.K. & Abod, Z.A. 2021. A hybrid cryptography technique for data storage on cloud computing. Journal of Discrete Mathematical Sciences and Cryptography, 1–12.

Bernard, H.R. 2013. Social research method: qualitative and quantitative methods. Thousand Oaks, California: Sage Publications.

Bifulco, I., Cirillo, S., Esposito, C., Guadagni, R. & Polese, G. 2021. An intelligent system for focused crawling from Big Data sources. Expert Systems with Applications. 184(ISSN 0957-4174):4–12.

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

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

Burgess, K., Kerr, D.V. & Houghton, L. 2013. Paradigmatic approaches used in enterprise resource planning systems research: a systematic literature review. Australasian Journal of Information Systems, 18(1):5-24.

Burrell, G. & Morgan, G. 1981. Sociological paradigms and organisational analysis: elements of the sociology of corporate life. The British Journal of Sociology., 32(3): pp. 455-459.

Butler, T. 1998. Towards a hermeneutic method for interpretive research in information systems. Journal of Information Technology, 13(4):285-300.

Cains, M.G., Flora, L., Taber, D., King, Z. & Henshel, D.S. 2021. Defining Cyber Security and Cyber Security Risk within a Multidisciplinary Context using Expert Elicitation. Risk Analysis. (February, 14).

Campbell, D.F. & Machado, A.A. 2013. Ensuring quality in qualitative inquiry: using key concepts as guidelines. Motriz: Revista de Educação Física, 19(3):572-579.

Cao, R., Tang, Z., Liu, C. & Veeravalli, B. 2019. A scalable multi-cloud storage architecture for cloud-supported medical internet of things. IEEE Internet of Things Journal, 7(3):1641–1654.

Carroll, M., Van Der Merwe, A. and Kotze, P., 2011, August. Secure cloud computing: Benefits, risks and controls. In 2011 Information Security for South Africa (pp. 1-9). IEEE.

Carvalho, G., Cabral, B., Pereira, V. & Bernardino, J. 2021. Edge computing: current trends, research challenges and future directions. Computing, 103(5):993-1023.

Cayirci, E., Garaga, A., Santana de Oliveira, A. & Roudier, Y. 2016. A risk assessment model for selecting cloud service providers. Journal of Cloud Computing, 5(1):1-12.

Cecez-Kecmanovic, D. 2011. Doing critical information systems research – arguments for a critical research methodology. European Journal of Information Systems, 20(4):440-455.

Cecez-Kecmanovic, D., Klein, H.K. & Brooke, C. 2008. Exploring the critical agenda in information systems research. Information Systems Journal, 18(2):123-135.

Chang, V., Kuo, Y. H. & Ramachandran, M. 2016. Cloud computing adoption framework: a security framework for business clouds. Future Generation Computer Systems, 57:24–41.

Chaudhary, K., Alam, M., Al-Rakhami, M.S. & Gumaei, A. 2021. Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. Journal of Big Data, 8(1):1-120.

Chauhan, V., Bansal, K. & Alappanavar, P. 2012. Exposing cloud computing as a failure. International Journal of Engineering Science and Technology, 4(4).

Chihande, M.K. and van der Poll, J.A., 2017, March. Post cloud computing implementation benefits and challenges realised for a South African technology company. In 2017 Conference on Information Communication Technology and Society (ICTAS) (pp. 1-6). IEEE.

Chun Tie, Y., Birks, M. and Francis, K. (2019) 'Grounded theory research: A design framework for novice researchers', SAGE Open Medicine. doi: 10.1177/2050312118822927.

Cohen, L., Manion, L. & Morrison, K. 2018. Research methods in education. 8th ed. New York: Routledge.

Cole, B. 2019. What is risk assessment? - Definition from WhatIs.com. SearchCompliance. https://searchcompliance.techtarget.com/definition/risk-assessment Date of access: 20 Mar. 2022.

Conway-Smith, B. 2012. South Africa: Striking miners "thought they were invincible" after taking "muti." The World from PRX. https://theworld.org/stories/2012-08-21/south-africa-striking-miners-thought-they-were-invincible-after-taking-muti Date of access: 27 Apr. 2019.

Cook, T.D. 1980. Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand Mcnally.

Creswell, J.W., 2002. Educational research: Planning, conducting, and evaluating quantitative (Vol. 7). Prentice Hall Upper Saddle River, NJ.

Creswell, John W. (1994). Research design: qualitative & quantitative approaches. Thousand Oaks, California: Sage Publications.

Crossman, A. 2020. Understanding functionalist theory. ThoughtCo. https://www.thoughtco.com/functionalist-perspective-3026625 Date of access: 2 Jun. 2019.

Darwish, A., Hassanien, A.E., Elhoseny, M., Sangaiah, A.K. & Muhammad, K. 2017. The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. Journal of Ambient Intelligence and Humanized Computing, 10(10):4151-4166.

Dataprise. 2017. IT Terms Glossary | Information Technology Definitions | Dataprise. Dataprise.com. https://www.dataprise.com/it-glossary Date of access: 15 Feb. 2022.

Dauriz, L., Remy, N. & Sandri, N. 2014. Luxury shopping in the digital age. McKinsey & Company. https://www.mckinsey.com/industries/retail/our-insights/luxury-shopping-in-the-digital-age Date of access: 4 August 2019.

Deb, D. & Fuad, M. 2021. Integrating big data and cloud computing topics into the computing curricula: A modular approach. Journal of Parallel and Distributed Computing, 157:303-315.

DeCarlo, Matthew. Scientific inquiry in social work. 2018.

de Lanerolle, I., 2018. How ICT Policy and Regulation Is Failing the 'Less Connected'in South Africa. Available at SSRN 3269698.

Deloitte. (2020). Accelerating to the cloud Breaking through the cloud adoption plateau.

De Villiers, M. 2005. Three approaches as pillars for interpretive information systems research: development research, action research and grounded theory. (In. Proceedings

of the 2005 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries organised by: Citeseer. p. 142-151).

Dioubate, B.M., Molok, N.N.A., Talib, S. and Tap, A.O.M., 2015. Risk assessment model for organisational information security. Asian Research Publishing Network Journal for English and Applied. Science, 10(23):17607-17613.

Djemame, K., Armstrong, D., Guitart, J. & Macias, M. 2016. A risk assessment framework for cloud computing. IEEE Transactions on Cloud Computing, 4(3):265-278.

Draganova, A. 2015. Book Review: Irving Seidman, Interviewing as Qualitative Research: A Guide for Researchers in Education & the Social Sciences. Qualitative Research, 15(3):411-412.

Drissi, S., Houmani, H. and Medromi, H., 2013. Survey: risk assessment for cloud computing. International Journal of Advanced Computer Science and Applications, 412.

Dwivedi, Y.K. & Kuljis, J. 2008. Profile of IS research published in the European Journal of Information Systems. European Journal of Information Systems, 17(6):678-693.

Eichstädt, J., Vymazal, M., Moxey, D. & Peiró, J. 2020. A comparison of the shared-memory parallel programming models OpenMP, OpenACC and Kokkos in the context of implicit solvers for high-order FEM, Computer Physics Communications, Volume 255, 2020, 107245, ISSN 0010-4655, https://doi.org/10.1016/j.cpc.2020.107245.

Elahi, H., Wang, G., Xu, Y., Castiglione, A., Yan, Q. & Shehzad, M.N. 2021. On the Characterization and Risk Assessment of AI-Powered Mobile Cloud Applications, Computer Standards & Interfaces, Volume 78, 2021, 103538, ISSN 0920-5489, https://doi.org/10.1016/j.csi.2021.103538.

Elhoseny, M., Abdelaziz, A., Salama, A.S., Riad, A.M., Muhammad, K. & Sangaiah, A.K. 2018. A hybrid model of Internet of Things and cloud computing to manage big data in health services applications. Future Generation Computer Systems, 86(17):1383-1394.

Elomari, A., Hassouni, L. & Maizate, A. 2021. DFS response time prediction using the techniques of “deep learning.” Artificial Intelligence and Industrial Applications, 144:26-

35.

El-Seoud, S. A., El-Sofany, H. F., Abdelfattah, M. A. F. and Mohamed, R. (2017). Big Data and Cloud Computing: Trends and Challenges, International Journal of Interactive Mobile Technologies (iJIM), 11(2), pp. pp. 34–52. doi: 10.3991/ijim.v11i2.6561.

Elton, G.R. 2002. The practice of history. Malden, Ma: Blackwell Publishers.

Elzamly, A., Hussin, B. & Basari, A.S.H. 2019. Classification of critical cloud computing security issues for banking organisations: a cloud Delphi study. International Journal of Grid and Distributed Computing, 9(8):137-158.

European Network and Information Security Agency. 2009. Cloud computing risk assessment [WWW Document]. ENISA. The European Union Agency for Cybersecurity. URL https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment. Accessed: May 15, 202.

Fleming, J., & Zegwaard, K. E. (2018). Methodologies, methods and ethical considerations for conducting research in work-integrated learning. International Journal of Work-Integrated Learning, *19*(3), 205–213.

Flick, U. 1992. Triangulation revisited: strategy of validation or alternative? Journal for the Theory of Social Behaviour, 22(2):175-197.

Frankenfield, J. 2020. Cloud Computing. Investopedia. https://www.investopedia.com/terms/c/cloud-computing.asp Date of access: 31 Jan. 2022.

Fritze, R. H. (2002). Review of From Reliable Sources: an Introduction to Historical Methods, by M. Howell & W. Prevenier. The Sixteenth Century Journal, 33(4), 1247–1249. https://doi.org/10.2307/4144236.

Furht, B. 2010. Cloud computing fundamentals. Handbook of Cloud Computing, 3-19.

Gadia, S. 2018. How to manage five key cloud computing risks. KPMG LLP (Canada). https://assets.kpmg/content/dam/kpmg/ca/pdf/2018/03/cloud-computing-risks-canada.pdf Date of access 14 Jan 2019.

Galal-Edeen, G.H. 2005. Information systems requirements engineering: An interpretive approach. The Egyptian Informatics Journal, 6(2):154-174.

Gangwar, H., Date, H. & Ramaswamy, R. 2015. Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. Journal of Enterprise Information Management, 28(1):107-130.

Qualitative research in information systems: time to be subjective?. In Information systems and qualitative research (pp. 444-465). Springer, Boston, MA.

Gartner. 2013. Definition of Big Data - Gartner Information Technology Glossary. Gartner, Inc. http://www.gartner.com/it-glossary/big-data/ Date of access: 18 Jun 2020.

Geiger, S. & Turley, D. 2003. Grounded theory in sales research: an investigation of salespeople's client relationships. Journal of Business & Industrial Marketing, 18(6/7):580-594.

Giraud, L. & Galy, B. 2018. Fault tree analysis and risk mitigation strategies for mine hoists. Safety Science, 110:222-234.

Glaser, B.G. 2008. Doing quantitative grounded theory. Mill Valley, Ca: Sociology Press.

Glaser, B.G. & Strauss, A.L. 1967. The discovery of grounded theory: strategies for qualitative research. Oxon, London: Routledge.

Golder, P.N. 2000. Historical method in marketing research with new evidence on long-term market share stability. Journal of Marketing Research, 37(2):156-172.

Goldkuhl, G. 2012. Pragmatism vs interpretivism in qualitative information systems research. European Journal of Information Systems, 21(2):135-146.

Goswami, P., Mukherjee, A., Maiti, M., Tyagi, & Yang, L. 2022. A neural-network-based optimal resource allocation method for secure IIoT network. IEEE Internet of Things Journal, 9:2538-2544.

Gottschalk, L.R. 1969. Understanding history: a primer of historical method. New York: Knopf.

Goulding, C. 2005. Grounded theory, ethnography and phenomenology. European Journal of Marketing, 39(3/4):294-308.

Goyal, S., 2014. Public vs private vs hybrid vs community-cloud computing: a critical review. International Journal of Computer Network and Information Security, 6(3), pp.20-29.

Grobler, M.M. & Dlamini, I.Z. 2010. Managing digital evidence - the governance of digital forensics. Journal of Contemporary Management, 7(1):1-21.

Grybauskas, A., Pilinkienė, V. & Stundžienė, A. 2021. Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. Journal of Big Data, 8(1):1-20.

Guba, E.G. 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. Educational Communication and Technology Journal, 29(2):75-91.

Guo, S., Liu, J., Yang, Y., Xiao, B. & Li, Z. 2019. Energy-efficient dynamic computation offloading and cooperative task scheduling in mobile cloud computing. IEEE Transactions on Mobile Computing, 18(2):319-333.

Halabi, T. and Bellaiche, M., 2019, June. Security risk-aware resource provisioning scheme for cloud computing infrastructures. In 2019 IEEE Conference on Communications and Network Security (CNS) (pp. 1-9). IEEE.

Hamel, D.W., 2012. Supervisors and subordinates reactions to performance appraisals (Doctoral dissertation).

Hammouri, Q. & Abu-Shanab, E.A. 2020. major factors influencing the adoption of cloud computing in Jordan. International Journal of Technology and Human Interaction, 16(4):55-69.

Han, B., Chen, Z., Liu, C. and Shang, M., 2020, September. Design and Implementation of Big Data Management Platform for Android Applications. In Proceedings of the 2020 3rd International Conference on Big Data Technologies (pp. 36-40).

Harfoushi, O., Akhorshaideh, A.H., Aqqad, N., Janini, M.A. & Obiedat, R. 2016. Factors affecting the intention of adopting cloud computing in Jordanian hospitals.

Communications and Network, 08(02):88-101.

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A. and Khan, S.U., 2015. The rise of “big data” on cloud computing: Review and open research issues. Information systems, 47, pp.98-115.

Heinle, C. and Strebel, J., 2010, August. IaaS adoption determinants in enterprises. In International Workshop on Grid Economics and Business Models (pp. 93-104). Springer, Berlin, Heidelberg.

Hirschheim, R. & Klein, H.K. 1989. Four paradigms of information systems development. Communications of the ACM, 32(10):1199-1216.

IBM Cloud Team. 2020. Top 7 Most Common Uses of Cloud Computing. www.ibm.com. IBM. https://www.ibm.com/cloud/blog/top-7-most-common-uses-of-cloud-computing Date of access: 15 Aug 2019.

Iivari, J., Hirschheim, R. & Klein, H.K. 1998. A paradigmatic analysis contrasting information systems development approaches and methodologies. Information Systems Research, 9(2):164-193.

Interpol. (2021). African cyberthreat assessment report. Interpol, October, 1–34.

Islam, S., Fenz, S., Weippl, E. & Mouratidis, H. 2017. a risk management framework for cloud migration decision support. Journal of Risk and Financial Management, 10(2):3-22.

ISO. 2019. ISO/IEC 27005:2018. ISO - International Organisation for Standardization. https://www.iso.org/standard/75281.html Date of access: 7 Jun. 2019.

ITU. (2021). Digital trends in Africa 2021: information and communication technology trends and developments in the Africa region 2017-2020. In *International* Telecommunication Union Publication (Vol. 2, Issue 1). https://elibrary.acbfpact.org/acbf/collect/acbf/index/assoc/HASH0146/faa3ee36/f864265c/9547.dir/Digital trends in Africa 2021.pdf

Jain, A., & Mahajan, N. (2017). Introduction to Cloud Computing. In *The Cloud DBA-Oracle*. https://doi.org/10.1007/978-1-4842-2635-3_1

Jingrui, H. 2017. Learning from Data Heterogeneity: Algorithms and Applications. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. (August).

Joe, V., Raj, J.S. & Smys S. 2021. Towards efficient big data storage with MapReduce deduplication System. International Journal of Information Technology and Web Engineering, 16(2):45-57.

Johnson, P. & Duberley, J. 2000. Understanding management research. London; Thousands Oaks; New Delhi: Sage Publications, Cop.

Jokonya, O., Kroeze, J.H. & van der Poll, J.A. 2012. Towards a framework for decision making regarding IT adoption. Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference on - SAICSIT '12.

Jones, S., Irani, Z., Sivarajah, U. & Love, P.E.D. 2017. Risks and rewards of cloud computing in the UK public sector: A reflection on three organisational case studies. Information Systems Frontiers, 21(2):359-382.

Karimi-Alaghehband, F. and Rivard, S., 2019. Information technology outsourcing and architecture dynamic capabilities as enablers of organisational agility. Journal of Information Technology, 34(2), pp.129-159.

Kaseb, M.R., Khafagy, M.H., Ali, I.A. & Saad, E.M. 2019. An improved technique for increasing availability in Big Data replication. Future Generation Computer Systems, 91:493-505.

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

Khajeh-Hosseini, A., Greenwood, D. and Sommerville, I., 2010, July. Cloud migration: a case study of migrating an enterprise it system to iaas. In 2010 IEEE 3rd International Conference on cloud computing (pp. 450-457). IEEE.

Kishor, A., Chakraborty, C. & Jeberson, W. 2021. Intelligent healthcare data segregation using fog computing with internet of things and machine learning. International Journal of Engineering Systems Modelling and Simulation, 12(2-3):188-194

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J. & Linkman, S. 2009. Systematic literature reviews in software engineering – A systematic literature review. Information and Software Technology, 51(1):7-15.

Klein, H.K. & Myers, M.D. 1999. A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Quarterly, 23(1):67-93.

Knutsson, O., Sneiders, E. & Alfalahi, A. 2012. Opportunities for improving eGovernment. Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance - ICEGOV ’12. 495–496.

Konietschke, F., Schwab, K. & Pauly, M. 2020. Small sample sizes: a big data problem in high-dimensional data analysis. Statistical Methods in Medical Research, 30(3):687-701.

Koshy, V. 2006. Action research for improving practice a practical guide. London: Paul Chapman Pub.

Kotval, X.P. & Burns, M.J. 2013. visualization of entities within social media: toward understanding users’ needs. Bell Labs Technical Journal, 17(4):77-101.

Kshetri, N. 2010. Cloud computing in developing economies. Computer, 43(10):47-55.

Kshetri, N. 2015. Recent US cybersecurity policy initiatives: challenges and implications. Computer, 48(7):64-69.

Ku, C.-H. & Leroy, G. 2014. A decision support system: automated crime report analysis and classification for e-government. Government Information Quarterly, 31(4):534-544.

Kuhlman, C., Jackson, L. & Chunara, R. 2020. No computation without representation: avoiding data and algorithm biases through diversity. arXiv:2002.11836 [cs]. (February, 26). http://arxiv.org/abs/2002.11836 Date of access: 19 Feb. 2022.

Kuhn, T.S. 1962. The structure of scientific revolutions. Chicago: University of Chicago Press.

Kumar, P.R., Raj, P.H. & Jelciana, P. 2018. Exploring data security issues and solutions in cloud computing. Procedia Computer Science, 125(2009):691-697.

Kusi-Sarpong, S., Orji, I.J., Gupta, H. & Kunc, M. 2021. Risks associated with the implementation of big data analytics in sustainable supply chains, Omega, Volume 105, 2021, 102502, ISSN 0305-0483, https://doi.org/10.1016/j.omega.2021.102502.

Langford-smith, J. (2020). Information Systems Audit Report 2020 – Local Government Entities. June.

Lal, P. & Bharadwaj, S.S. 2016. Understanding the impact of cloudbased services adoption on organisational flexibility. Journal of Enterprise Information Management, 29(4):566-588.

Lampos, V. & Cristianini, N. 2012. Nowcasting events from the social web with statistical learning. ACM Transactions on Intelligent Systems and Technology, 3(4):1-22.

Laney, D. 2001. Application Delivery Strategies. META Group. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf Date of access: 10 Oct 2019.

Lang, M., Wiesche, M. and Krcmar, H., 2018. Perceived Control and Privacy in a Professional Cloud Environment Hawaii International Conference on System Sciences (HICSS) Big Island. Hawaii, USA.

Laudon, K.C. & Laudon, J.P. 2018. Management information systems: managing the digital firm. 15th ed. New York, New York: Pearson.

Leung, P.P.L., Wu, C.H., Kwong, C.K., Ip, W.H. & Ching, W.K. 2021. Digitalisation for optimising nursing staff demand modelling and scheduling in nursing homes. Technological Forecasting and Social Change, 164:5-15.

Li, C., Liu, J., Li, W. & Luo, Y. 2021. Adaptive priority-based data placement and multi-task scheduling in geo-distributed cloud systems. Knowledge-Based Systems, 224:1-16.

Lieblich, A., Tuval-Mashiach, R. & Zilber, T. 1998. Narrative research: reading, analysis, and interpretation. Thousand Oaks; London; New Delhi: Sage Publications, Corp.

Lin, B., Zhu, F., Zhang, J., Chen, J., Chen, X., Xiong, N.N., et al. 2019. A time-driven data placement strategy for a scientific workflow combining edge computing and cloud computing. IEEE Transactions on Industrial Informatics, 15(7):4254-4265.

Liu, S., Wu, J., Lu, Z. and Xiong, H., 2013, June. Vmras: A novel virtual machine risk assessment scheme in the cloud environment. In 2013 IEEE International Conference on Services Computing (pp. 384-391). IEEE.

Li, Y., Zhao, H. and Zhu, L., 2019, April. Research on the construction of e-commerce security risk assessment model based on cloud computing. In 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 589-592). IEEE.

Maanen, J.V. 1979. Reclaiming qualitative methods for organisational research: a preface. Administrative Science Quarterly, 24(4):520.

Machuga, R. 2020. Factors determining the use of cloud computing in enterprise management in the EU (considering the type of economic activity). Problems and Perspectives in Management. 18(3):93–105.

Mackenzie, N. & Knipe, S. 2006. IIER 16: Research dilemmas: Paradigms, methods and methodology. Issues in Educational Research. 16. http://www.iier.org.au/iier16/mackenzie.html Date of access: 18 Jun. 2019.

Mackita, M., Shin, S.-Y. & Choe, T.-Y. 2019. ERMOCTAVE: a risk management framework for IT systems which adopt cloud computing. Future Internet, 11(9):1-19.

MacVaugh, J. & Schiavone, F. 2010. Limits to the diffusion of innovation. European Journal of Innovation Management, 13(2):197-221.

Mailavaram A., Padmaja Rani B. (2019) Big Data: Scalability Storage. In: Saini H., Sayal R., Govardhan A., Buyya R. (eds) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol 74. Springer, Singapore. https://doi.org/10.1007/978-981-13-7082-3_54.

Malathi, L., Malathi, S. & Nasurdeen Ahamed, N. 2019. Hybrid Cloud Storage for Secure Authorization and Information Hiding. International Journal of Engineering and Advanced Technology (IJEAT). 8(6S). https://www.ijeat.org/wp-content/uploads/papers/v8i6S/F10820886S19.pdf Date of access: 06 2022.

Maniah, Abdurachman, E., Gaol, F.L. & Soewito, B. 2019. Survey on threats and risks in the cloud computing environment. Procedia Computer Science, 161:1325-1332.

Maniah & Milwandhari, S. 2020. Risk Analysis of Cloud Computing in the Logistics Process. In 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE) (pp. 1-5). IEEE.

Manogaran, G., Thota, C. & Kumar, M.V. 2016. MetaCloudDataStorage Architecture for Big Data Security in Cloud Computing. Procedia Computer Science. 87:128–133.

March, S.T. & Smith, G.F. 1995. Design and natural science research on information technology. Decision Support Systems, 15(4):251-266.

Maric, B. & Flensburg, P. 2012. Enhanced four paradigms of information systems development in network societies. http://urn.kb.se/resolve?urn=urn:nbn:se:bth4287 Date of access: Sept 19 2020.

Marshall, G. 1994. The concise Oxford dictionary of sociology. Choice Reviews Online. 32(03):32–126132–1261.

Martens, B. & Teuteberg, F. 2011. Decision-making in cloud computing environments: a cost and risk based approach. Information Systems Frontiers, 14(4):871-893.

Martini, B. & Choo, K.-K.R. 2012. An integrated conceptual digital forensic framework for cloud computing. Digital Investigation, 9(2):71-80.

Mason, R.O. 1986. Four ethical issues of the information age. In Computer ethics 10(1): 41-48. Routledge. Vancouver.

Mason, R.O., McKenney, J.L. & Copeland, D.G. 1997. An historical method for MIS research: steps and assumptions. MIS Quarterly, 21(3):307-320.

Mastorakis, G. (2018). Human-like machine learning : limitations and suggestions. 1–24

Ma, X., Gao, H., Xu, H. & Bian, M. 2019. An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing. EURASIP Journal on Wireless Communications and Networking, (1):1-19.

McKendrick, J. 2013. In the rush to cloud computing, here’s one question not enough

people are asking. Forbes. https://www.forbes.com/sites/joemckendrick/2013/02/19/in-the-rush-to-cloud-computing-heres-one-question-not-enough-people-are-asking/?sh=141827727194 Date of access: 26 Dec 2019.

Mertens, D.M. 2019. Research and evaluation in education and psychology: integrating diversity with ... quantitative, qualitative, and mixed methods. Thousand Oaks, California: Sage Publications.

Meyerson, B. & Townsend, J. 2004. Revised Code of Ethics for Commercial Arbitrators Explained - ProQuest. Dispute Resolution Journal. 59(1):10–17. https://www.proquest.com/openview/0446f718195267acb8d97d8717a0c26d/1?pq-origsite=gscholar&cbl=25210 Date of access: 05 Feb 2019.

Mingers, J. 2003. The paucity of multimethod research: a review of the information systems literature. Information Systems Journal, 13(3):233-249.

Misra, S.C. & Mondal, A. 2011. Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment. Mathematical and Computer Modelling, 53(3):504-521.

Mohammed, F., Ibrahim, O., Nilashi, M. & Alzurqa, E. 2017. Cloud computing adoption model for e-government implementation. Information Development, 33(3):303-323.

Mohlameane, M. & Ruxwana, N. 2014. The awareness of cloud computing: a case study of South African SMEs. International Journal of Trade, Economics and Finance, 5(1):6-11.

Mondragón-Ruiz, G., Tenorio-Trigoso, A., Castillo-Cara, M., Caminero, B. & Carrión, C. 2021. An experimental study of fog and cloud computing in CEP-based Real-Time IoT applications. Journal of Cloud Computing, 10(1):1-17.

Moonasar, V. & Naicker, V. 2018. Cloud adoption: a conceptual model to assess the maturity of south African large enterprises. In: ICEME 2018: Proceedings of the 2018 9th International Conference on E-business, Management and Economics. pp. 15–21.

Moonasar, V. & Naicker, V. 2020. Cloud capability maturity model: a study of South African large enterprises. SA Journal of Information Management, 22(1):1-12.

Mthunzi, S.N., Benkhelifa, E., Bosakowski, T., Guegan, C.G. & Barhamgi, M. 2020. Cloud computing security taxonomy: From an atomistic to a holistic view. Future Generation Computer Systems, 107:620-644.

Mujinga, M. & Chipangura, B. 2011. Cloud computing concerns developing economies. 9th Australian Information Security Management Conference. 127.

Mukred, M., Yusof, Z.M., Al-Moallemi, W.A., Mokhtar, U.A. & Hawash, B. 2021. Electronic records management systems and the competency of educational institutions: Evidence from Yemen. Information Development Vol. 38(1) 125–148.

Muniz, M.V.P., Lima, G.B.A., Caiado, R.G.G. & Quelhas, O.L.G. 2017. Bow tie to improve risk management of natural gas pipelines. Process Safety Progress, 37(2):169-175.

Myers, M.D. & Venable, J.R. 2014. A set of ethical principles for design science research in information systems. Information & Management, 51(6):801-809.

National Science Foundation (NSF). 2014. Core Techniques and Technologies for Advancing Big data Science & Engineering (BIGDATA). www.nsf.gov. https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf14543 Date of access: 18 Jun. 2021.

Netshakhuma, N.S. 2019. Assessment of a South Africa national consultative workshop on the Protection of Personal Information Act (POPIA). Global Knowledge, Memory and Communication, 69(1/2):58-74.

Neuman, W.L. 2002. Social research methods: qualitative and quantitative approaches. Teaching Sociology, 30(3):380-389.

Niehaves, B. & Stahl, B. 2006. Criticality, Epistemology, and Behaviour vs. Design-IS Research across different sets of paradigms. In 14th European Conference on Information Systems (ECIS 2006), Göteborg..

Obeidat, M.A. & Turgay, T. 2012. Empirical analysis for the factors affecting the adoption of cloud computing initiatives by information technology executives. Journal of Management Research, 5(1):3-13.

Odeh, M., Garcia-Perez, A. & Warwick, K. 2017. Cloud computing adoption at higher education institutions in developing countries: a qualitative investigation of main enablers and barriers. International Journal of Information and Education Technology, 7(12):921-927.

OECD. 2011. Quality Framework for OECD Statistical Activities - OECD. Oecd.org. https://www.oecd.org/sdd/qualityframeworkforoecdstatisticalactivities.htm Date of access: 14 May 2019.

Oktian, Y.E., Lee, S.-G. & Lee, B.-G. 2020. Blockchain-Based continued integrity service for IoT big data management: a comprehensive design. Electronics, 9(9):4-20.

Oliveira, T., Thomas, M. & Espadanal, M. 2014. Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. Information & Management, 51(5):497-510.

Open Group. 2012. *Service-Oriented Cloud Computing Infrastructure (SOCCI)* Framework – Viewpoints. www.opengroup.org. shorturl.at/cDRX1 Date of access: 16 Jul. 2022.

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

O'Brie, R. 1998. An overview of action research methodology. Richardson R. Theory and Practice of Action Research. Universidade Federal da Paraíba, João Pessoa.

O'leary, Z. 2004. The essential guide to doing research. London: Sage Publications.

Paltrinieri, N., Comfort, L. & Reniers, G. 2019. Learning about risk: machine learning for risk assessment. Safety Science, 118:475-486.

Pandey, V. & Saini, P. 2020. A heuristic method towards deadline-aware energy-efficient MapReduce scheduling problem in Hadoop YARN. Cluster Computing, 24(2):683-699

Paquet, C., Coulombier, D., Kaiser, R. & Ciotti, M. 2006. Epidemic intelligence: a new framework for strengthening disease surveillance in Europe. Eurosurveillance, 11(12):5-6.

Pańkowska, M., Pyszny, K. & Strzelecki, A. 2020. Users' adoption of sustainable cloud computing solutions. Sustainability, 12(23):1-17.

Picciotto, R. 2019. Evaluation and the big data challenge. American Journal of Evaluation, 41(2):166-181.

Poniszewska-Maranda, A., Matusiak, R., Kryvinska, N. & Yasar, A. U. H. 2019. A real-time service system in the cloud. Journal of Ambient Intelligence and Humanized Computing, 11(3):961-977.

Porra, J., Hirschheim, R. & Parks, M. 2014. The historical research method and information systems research. Journal of the Association for Information Systems, 15(9):536-576.

Potey, M., A Dhote, C. & H. Sharma, D. 2013. Cloud computing understanding risk, threats, vulnerability and controls: a survey. International Journal of Computer Applications, 67(3):9-14.

Pourghomi, P. & Ghinea, G. 2012. Challenges of managing secure elements within the NFC ecosystem. In 2012 International Conference for Internet Technology and Secured Transactions (pp. 720-725). IEEE.

Power, R. & Naysmith, J. 2005. Action research: a guide for associate lecturers. Milton Keynes: Open University. shorturl.at/dquH7.

Qasem, Y.A.M., Abdullah, R., Jusoh, Y.Y., Atan, R. & Asadi, S. 2019. cloud computing adoption in higher education institutions: a systematic review. IEEE Access, 63(3):63722-63744.

Qasim, H. & Abu-Shanab, E. 2015. Drivers of mobile payment acceptance: the impact of network externalities. Information Systems Frontiers, 18(5):1021-1034.

Qin, J., Xi, Y. & Pedrycz, W. 2020. Failure mode and effects analysis (FMEA) for risk assessment based on interval type-2 fuzzy evidential reasoning method, Applied Soft Computing, Volume 89, 2020, 106134, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2020.106134.

Radhakrishna, R., Tobin, D., Bressman, M. & Thomson, J. 2012. ensuring data quality in

extension research and evaluation studies. The Journal of Extension, 50(3). https://tigerprints.clemson.edu/joe/vol50/iss3/61/.

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

Raja, K. & Hanifa, S.M. 2017. Bigdata Driven Cloud Security: A Survey. IOP Conference Series: Materials Science and Engineering. 225(1):012184.

Rajen, R. & Prasad, R.V.V.S.V. 2015. Cloud Computing Research: Challenges and Security Issues. International Journal of Computer Trends and Technology. 30(3):157–161.

Rao, V.N. & Rao, T.N.P. 2021. Big Data and Its Applications. International Journal of Advanced Research in Science, Communication and Technology. 2(3):149–157.R.

Rayala, V. & Kalli, S.R. 2020. Big data clustering using improvised fuzzy C-Means clustering. Revue d'Intelligence Artificielle, 34(6):701-708.

Reena, M. & Nargunam, A.S. 2019. Secured storage of big data in cloud. International Journal of Recent Technology and Engineering, 8(2S3):6-10.

Regalado, A. 2011. Who Coined "Cloud Computing"? Technology Review. MIT. https://www.technologyreview.com/2011/10/31/257406/who-coined-cloud-computing/ Date of access: 4 Jun. 2019.

Richter, A.N. & Khoshgoftaar, T.M. 2019. Efficient learning from big data for cancer risk modeling: a case study with melanoma. Computers in Biology and Medicine, 110:29-39.

Rjoub, G., Bentahar, J. & Wahab, O.A. 2020. BigTrustScheduling: trust-aware big data task scheduling approach in cloud computing environments. Future Generation Computer Systems, 110:1079-1097.

Sabahi, F., 2011, May. Cloud computing security threats and responses. In 2011 IEEE 3rd International Conference on Communication Software and Networks (pp. 245-249). IEEE.

Sadowski, A., Galar, Z., Walasek, R., Zimon, G. & Engelseth, P. 2021. Big data insight on global mobility during the Covid-19 pandemic lockdown. Journal of Big Data, 8(1):1-33.

Saeedi, S., Khorsand, R., Ghandi Bidgoli, S. & Ramezanpour, M. 2020. Improved many-objective particle swarm optimization algorithm for scientific workflow scheduling in cloud computing, Computers & Industrial Engineering,Volume 147, 2020, 106649, ISSN 0360-8352, https://doi.org/10.1016/j.cie.2020.106649.

Saez, J.A. & Corchado, E. 2019. KSUFS: a novel unsupervised feature selection method based on statistical tests for standard and big data problems. IEEE Access, 7:99754-99770.

Sahu, Y. & Pateriya, R.K. 2013. Cloud computing overview with load balancing techniques. International Journal of Computer Applications, 65(24). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.302.9773&rep=rep1&type=pdf Date of access: 15 Oct. 2019.

Saif, S. & Wazir, S. 2018. Performance analysis of big data and cloud computing techniques: a survey. Procedia Computer Science, 132(1877-0509):118-127.

Santos, D.R. dos, Westphall, C.M. & Westphall, C.B. 2014, May. A dynamic risk-based access control architecture for cloud computing. In 2014 IEEE Network Operations and Management Symposium (NOMS) (pp. 1-9). IEEE.

S.A.S. 2016. What is Big Data and why it matters. Sas.com. https://www.sas.com/en_ca/insights/big-data/what-is-big-data.html Date of access: 15 Feb. 2022.

Schneier, B. 2010. Essays: The Psychology of Security (Part 1) - Schneier on Security. www.schneier.com. https://www.schneier.com/essays/archives/2008/01/the_psychology_of_se.html Date of access: 25 May 2019.

Scotland, J. 2012. exploring the philosophical underpinnings of research: relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. English Language Teaching, 5(9):9-16.

Sellami, R., Zalila, F., Nuttinck, A., Dupont, S., Deprez, J.-C. & Mouton, S. 2020. FADI - A deployment framework for big data management and analytics. In 2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE) (pp. 153-158). IEEE.

Shafer, R.J. 1974. A Guide to historical method. Illinois Dorsey Press.

Shafiu, I., Wang, W.Y.C. & Singh, H. 2016. Drivers and barriers in the decision to adopt IaaS: a public sector case study. International Journal of Business Information Systems, 21(2):249-267.

Shameli-Sendi, A. & Cheriet, M. 2014. Cloud Computing: A Risk Assessment Model. Boston, MA: IEEE International Conference on Cloud Engineering. http://www.synchromedia.ca/system/files/RAaaS.pdf Date of access: 15 Jun. 2019.

Sharma, S.K., Sengupta, A. & Panja, S.C. 2019. Grounded theory: a method of research inquiry. Methodological Issues in Management Research: Advances, Challenges, and the Way Ahead, 11:181-201.

Shen, J., Liu, D., Bhuiyan, M.Z.A., Shen, J., Sun, X. & Castiglione, A. 2020. Secure verifiable database supporting efficient dynamic operations in cloud computing. IEEE Transactions on Emerging Topics in Computing, 8(2):280-290.

Shen, J., Zhou, T., He, D., Zhang, Y., Sun, X. & Xiang, Y. 2019. Block design-based key agreement for group data sharing in cloud computing. IEEE Transactions on Dependable and Secure Computing, 16(6):996-1010.

Shi, Y. 2014. Big data history, current status, and challenges going forward. The Bridge. 44(4). https://www.nae.edu/129211/Big-Data-History-Current-Status-and-Challenges-going-Forward Date of access: 8 Jun 2019.

Shu, W., Cai, K. & Xiong, N.N. 2021. Research on strong agile response task scheduling optimization enhancement with optimal resource usage in green cloud computing. Future Generation Computer Systems, 124:12-20.

Silver, C. & Lewins, A. 2014. Using software in qualitative research: A step by step guide. London: Sage.

Silverman, D. 1987. The theory of organisations: a sociological framework. Aldershot, England: Gower.

Singh, N.P. & Singh, S. 2019. Building supply chain risk resilience. Benchmarking: An International Journal, 26(7):1-16.

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

Sivasubramanian, Y., Ahmed, S.Z. & Mishra, V.P. 2017. Risk Assessment for Cloud Computing. International Research Journal of Electronics and Computer Engineering. 3(2):7.

Soanes, C. & Stevenson, A. 2004. Concise Oxford English dictionary. 11th ed. Vol. 42. Choice Reviews Online. Oxford: Oxford University Press.

Sohaib, O., Naderpour, M., Hussain, W. & Martinez, L. 2019. Cloud computing model selection for e-commerce enterprises using a new 2-tuple fuzzy linguistic decision-making method. Computers & Industrial Engineering, 132:47-58.

Somekh, B. & Lewin, C. 2005. Research methods in the social sciences. Thousand Oaks, Calif.: Sage Publications.

Sorensen, L.C. 2018. “Big data” in educational administration: an application for predicting school dropout risk. Educational Administration Quarterly, 55(3):404-446.

Stergiou, C.L., Psannis, K.E. & Ishibashi, Y. 2020. green cloud communication system for big data management. 2020 3rd World Symposium on Communication Engineering (WSCE). (October, 9).

Stergiou, C., Psannis, K.E., Gupta, B.B. & Ishibashi, Y. 2018. Security, privacy & efficiency of sustainable cloud computing for big data & IoT. Sustainable Computing: Informatics and Systems, 19:174-184.

Stork, C., Calandro, E. & Gillwald, A. 2013. Internet going mobile: internet access and use in 11 African countries. Emerald insight, 15(5):34-51.

Strauss, A., & Corbin, J. 1998. Basics of qualitative research: techniques and procedures

for developing grounded theory. 2nd ed. Thousand Oaks, California: Sage Publications.

Sundararaj, V. 2018. Optimal task assignment in mobile cloud computing by queue based ant-bee algorithm. Wireless Personal Communications, 104(1):173-197.

Swathy Akshaya, M. & Padmavathi, G. 2018. Taxonomy of security attacks and risk assessment of cloud computing. Advances in Intelligent Systems and Computing, 12:37-59.

Taivalsaari, A. and Mikkonen, T., 2015, August. Cloud technologies for the Internet of Things: Defining a research agenda beyond the expected topics. In 2015 41st Euromicro Conference on Software Engineering and Advanced Applications (pp. 484-488). IEEE.

Taleb, I., Serhani, M.A., Bouhaddioui, C. & Dssouli, R. 2021. Big data quality framework: a holistic approach to continuous quality management. Journal of Big Data, 8(1):1-41.

Tang, G. & Zeng, H. 2020. Collaborative management and control of blockchain in cloud computing environment. Journal of Intelligent & Fuzzy Systems, 40(4):1-11.

Tarkyth, D. 2019. The challenges of big data in expanding geoscience: embracing new initiatives to untangle our world. Geoscience Canada, 18:159-162.

Tashakkori, A. & Teddlie, C. 2003. Handbook of mixed methods in social & behavioral research. Thousand Oaks, California.: Sage Publications.

Taylor, J.R. 2017. Automated HAZOP revisited. Process Safety and Environmental Protection, 111:635-651.

Tchernykh, A., Schwiegelsohn, U., Talbi, E. & Babenko, M. 2016. Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability. Journal of Computational Science, 36:1-9.

Thakur, N., Bisen, D., Rohit, V. & Gupta, N. 2014. Review on cloud computing: issues, services and models. International Journal of Computer Applications, 91(9):34-39.

Tian, W. 2021. Personalized emotion recognition and emotion prediction system based on cloud computing. Mathematical Problems in Engineering, 2021:1-10.

Trinath Basu, M. & Sastry, J. 2020. Enhancing data security under multi-tenancy within open stack. International Journal of Advanced Trends in Computer Science and Engineering, 9(1):533-544.

Turskis, Z., Goranin, N., Nurusheva, A. & Boranbayev, S. 2019. Information security risk assessment in critical infrastructure: a hybrid MCDM approach. Informatica, 30(1):187-211.

Trimintzios, P., & Gavrila, R. (2013). National-level Risk Assessments. https://doi.org/10.2824/2633

Uljarević, M., Jo, B., Frazier, T.W., Scahill, L., Youngstrom, E.A. & Hardan, A.Y. 2021. Using the big data approach to clarify the structure of restricted and repetitive behaviors across the most commonly used autism spectrum disorder measures. Molecular Autism, 12(1):1-14.

UNCTAD. (2021). Digital Economy Report UNCTAD 2021 (Issue September).

Urquhart, C., Lehmann, H. & Myers, M.D. 2009. Putting the "theory" back into grounded theory: guidelines for grounded theory studies in information systems. Information Systems Journal, 20(4):357-381.

Vaccarino, F., Comrie, M., Murray, N. & Sligo, F. 2006. Action research initiatives: The Wanganui adult literacy and employment programme. Palmerston North, New Zealand: Department of Communication and Journalism, Massey University.

Van Der Schyff, K. & Krauss, K. 2014. Higher Education Cloud Computing in South Africa: Towards Understanding Trust and Adoption issues. South African Computer Journal. 53.

Venable, J. & Baskerville, R. 2012. Eating our own cooking: toward a more rigorous design science of research methods. Electronic Journal of Business Research Methods, 10:141-153.

VMware. 2022. What is Bring Your Own Device (BYOD)? | VMware Glossary. VMware Inc. https://www.vmware.com/topics/glossary/content/bring-your-own-device-byod.html Date of access: 16 Jul. 2022.

Walsham, G. 1995. Interpretive case studies in IS research: nature and method. European Journal of Information Systems, 4(2):74-81.

Wang, Q., Wang, C., Ren, K., Lou, W. & Li, J. 2011. Enabling public auditability and data dynamics for storage security in cloud computing. IEEE Transactions on Parallel and Distributed Systems, 22(5):847-859.

Wenhong, T. & Yong, Z. 2014. Optimized cloud resource management and scheduling. USA: Morgan Kaufmann.

Williams, M.G. 2018. A risk assessment on raspberry PI using NIST standards. International Journal of Computer Science and Network Security, https://www.semanticscholar.org/paper/A-Risk-Assessment-on-Raspberry-PI-using-NIST-Williams/f114cfd5b0b2e845c0a96385a42f8b4f1f8e028d Date of access: 19 Jan. 2022.

Williamson, K. & Johanson, G. 2018. Research methods: information, systems, and contexts. Cambridge, Ma: Chandos Publishing.

Woods, M., Paulus, T., Atkins, D.P. and Macklin, R., 2016. Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS. ti and NVivo, 1994–2013. Social Science Computer Review, 34(5), pp.597-617.

Wortley, R. & Mazerolle, L. 2009. Environmental criminology and crime analysis: situating the theory, analytic approach and application. Crime Prevention and Community Safety, 11(2):144-146.

Xiao, B., Guo, J., Qian, W., Hu, H. & Zhou, A. 2017. NIOSIT: efficient data access for log-structured merge-tree style storage systems. Proceedings of the ACM Turing 50th Celebration Conference - China on - ACM TUR-C ’17.

Xie, X. & Zhang, Q. 2021. An edge-cloud-aided incremental tensor-based fuzzy C-Means approach with big data fusion for exploring smart data. Information Fusion, 76:168-174.

Xin, G. & Fan, P. 2021. A lossless compression method for multi-component medical images based on big data mining. Scientific Reports, 11(1):1-11.

Xiong, H., Zhang, H. & Sun, J. 2019. Attribute-Based privacy-preserving data sharing for dynamic groups in cloud computing. IEEE Systems Journal, 13(3):2739-2750.

Xu, H. 2021. Research on mass monitoring data Retrieval Technology based on HBase. Journal of Physics: Conference Series. 1871(1):1–7.

Xu, Z. & Shi, Y. 2015. Exploring big data analysis: fundamental scientific problems. Annals of Data Science, 2(4):363-372.

Yan, F. & Xu, K. 2018. A set pair analysis based layer of protection analysis and its application in quantitative risk assessment. Journal of Loss Prevention in the Process Industries, 55:313-319.

Yang, J., Zhao, Y., Han, C., Liu, Y. & Yang, M. 2021. Big data, big challenges: risk management of financial market in the digital economy. Journal of Enterprise Information Management. ahead-of-print(ahead-of-print).

Yang, M., Zhou, X., Zeng, J., & Xu, J. (2016). Challenges and solutions of information security issues in the age of big data. China Communications, 13(3), 193–202. https://doi.org/10.1109/CC.2016.7445514

Yeasmin, S., & Rahman.K.F. (2012). 'Triangulation' Research Method as the Tool of Social Science Research. Bup Journal, 1(1), 154–163. http://www.bup.edu.bd/journal/154-163.pdf

Yassine, A., Singh, S., Hossain, M.S. & Muhammad, G. 2019. IoT big data analytics for smart homes with fog and cloud computing. Future Generation Computer Systems, 91:563-573.

Yeboah-Boateng, E.O. & Essandoh, K.A. 2014. Factors influencing the adoption of cloud computing by small and medium enterprises in developing economies. International Journal of Emerging Science and Engineering, 2(4):13-20.

Yin, F. & Shi, F. 2021. A comparative survey of big data computing and HPC: from a parallel programming model to a cluster architecture. International Journal of Parallel

Programming, 50(1): 27-64.

Yu, X., Shi, Y., Zhang, L., Nie, G. & Huang, A. 2014. Intelligent knowledge beyond data mining: influences of habitual domains. Communications of the Association for Information Systems, 34(1):3-13.

Zamora-Izquierdo, M.A., Santa, J., Martínez, J.A., Martínez, V. & Skarmeta, A.F. 2019. Smart farming IoT platform based on edge and cloud computing. Biosystems Engineering, 177:4-17.

Zanoon, N., Alhaj, A. & Khwaldeh, S. 2016. Cloud computing and big data is there a relation between the two: a study. International Journal of Applied Engineering Research, 12(17). https://www.researchgate.net/publication/319913597_Cloud_Computing_and_Big_Data_is_there_a_Relation_between_the_Two_A_Study Date of access: 12 May 2019.

Zhang, C., Xiao, C. & Liu, H. 2019. Spatial big data analysis of political risks along the belt and road. Sustainability, 11(8):2-12.

Zhang, S., Yan, H. & Chen, X. 2012. Research on key technologies of cloud computing. Physics Procedia, 33:1791-1797.

Zheng, J., Cai, Y., Wu, Y. & Shen, X. 2019. Dynamic computation offloading for mobile cloud computing: a stochastic game-theoretic approach. IEEE Transactions on Mobile Computing, 18(4):771-786.

Zissis, D. & Lekkas, D. 2012. Addressing cloud computing security issues. Future Generation Computer Systems, 28(3):583-592.

APPENDIX A: RESEARCHER'S CODE OF CONDUCT

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Research and Innovation

CODE OF CONDUCT FOR RESEARCHERS

This code of conduct is applicable to all NWU researchers.

As a researcher of the North-West University (NWU), I subscribe to the rules of the NWU Senate Committee for Research Ethics (SCRE), all applicable policies of the NWU as well as all national and international laws and regulations applicable to my field of study. Furthermore, I commit myself to abide by the ethical principles and responsibilities as set out in the Singapore statement on Research Integrity (22 September 2010), in any and all research endeavours that I undertake as a researcher of the NWU.

The four major principles of research integrity to which I will adhere and that will guide my research are:

- Honesty in all aspects of research;
- Accountability in the conduct of research;
- Professional courtesy and fairness in working with others;
- Good stewardship of research on behalf of others.

Consequently I will also adhere to the following ethical responsibilities:

1. I will take responsibility for the originality and trustworthiness of my research.
2. I will stay abreast of and adhere to all institutional, national, and international laws, regulations, and policies applicable and related to my research.
3. I will at all times employ appropriate research methods, base my conclusions on critical analysis of the evidence and report my findings and interpretations fully and objectively.
4. I will keep clear and accurate records of all research that I have conducted in a manner that will allow verification and replication of my work by others, if applicable.
5. I will, where applicable, share my data and findings openly and promptly, in line with external funding rules. This will be done as soon as possible after I have had an opportunity to establish priority and ownership claims.
6. I will take responsibility for my own contributions to publications, funding applications, reports and other representations of my research. I will also and only include authors who meet valid authorship criteria.
7. I will acknowledge the names and roles of those who made significant contributions to my research in publications, including writers, funders, sponsors, and others, but do not meet authorship criteria.
8. In my peer reviews, I will provide fair, prompt and rigorous evaluations and I will respect confidentiality when I review others' work.
9. I will disclose all conflicts of interest (financial and other) that could compromise the trustworthiness of my work in research proposals, publications, public communications, and in review activities.
10. When I publically address a community in the spirit of academic freedom, I will in all stages base my professional comments on research findings (if applicable) and my expertise. I will distinguish between professional comments and opinions based on personal views.
11. Should any irresponsible research practices and/or research misconduct become known to me or brought under my attention, I will report such irresponsible research activities to the appropriate authorities.
12. I will respond to irresponsible research practices or conduct, by taking prompt actions as set out in the procedures of the university. I will also protect those who report misconduct in good faith, to the best of my abilities.
13. I will endeavour to create and sustain an environment that encourage research integrity through education of students, research teams and peers, as well as abide by policies, and reasonable standards for advancement.
14. I will at all times weigh societal benefits against the risks inherent in my work.

Name: Fadi Fataftah Signature: fataftah Date: June 22, 2021

Original details: (11884754) P:\9. Research and Post-graduate Education\9.1 Implementation of the research strategy\9.1.5 Ethics\9.1.5.1.3_Code_Conduct_2017.docm
18 July 2017

File reference: 9.1.5.1.3

APPENDIX B: ETHICAL CLEARANCE

NWU®
NORTH-WEST UNIVERSITY
NOORDWES-UNIVERSITEIT
YUNIBESITI YA BOKONE-BOPHIRIMA

Private Bag X1290, Potchefstroom
South Africa 2520
Tel: 018 299-1111/2222
Fax: 018 299-4910
Web: http://www.nwu.ac.za

Senate Commitee for Research Ethics
Tel: 018 299-4849
Email: nkosinathi.machine@nwu.ac.za

ETHICS APPROVAL LETTER OF STUDY

Based on the review by the **Faculty of Natural and Agricultural Sciences Ethics Committee (FNASREC)**, the Committee hereby clears your study as no ethical risk. This implies that the FNASREC grants permission that, provided the general conditions specified below are met, the study may be initiated, using the ethics number below.

Study title: Risk assessment processes for big data based on cloud computing technologies: a comparative study
Study Leader/Supervisor: Dr B Isong
Student: FMM Fataftah

Ethics number: N|W|U|-|0|1|5|5|8|-|2|0|-|A|9
Institution | Study Number | Year | Status

Status: S = Submission; R = Re-Submission; P = Provisional Authorisation; A = Authorisation

Application type: Single **Risk Category:** **No Risk**

Commencement date: 01/02/2020
Expiry date: 01/10/2022

General conditions:
The following general terms and conditions apply:

- *The commencement date indicates the date when the study may be started.*
- *In the interest of ethical responsibility, the NWU-SCRE and FNASREC reserves the right to:*
  - *request access to any information or data at any time during the course or after completion of the study;*
  - *to ask further questions, seek additional information, require further modification or monitor the conduct of your research or the informed consent process;*
  - *withdraw or postpone approval if:*
    - *any unethical principles or practices of the study are revealed or suspected;*
    - *it becomes apparent that any relevant information was withheld from the FNASREC or that information has been false or misrepresented;*
    - *submission of the annual (or otherwise stipulated) monitoring report, the required amendments, or reporting of adverse events or incidents was not done in a timely manner and accurately; and / or*
    - *new institutional rules, national legislation or international conventions deem it necessary.*
- *FNASREC can be contacted for further information or any report templates via Roelof.Burger@nwu.ac.za 018 299 4269*

The FNASREC would like to remain at your service as scientist and researcher, and wishes you well with your study. Please do not hesitate to contact the FNASREC or the NWU-SCRE for any further enquiries or requests for assistance.

Yours sincerely,

Prof Roelof Burger
Chairperson Faculty of Natural and Agricultural Sciences Ethics Committee (FNASREC)

File reference: 9.1.5.4.2

Page 1 of 1

APPENDIX C: PROOF OF PAPER ACCEPTANCE

ICECET 2022
International Conference on
Electrical, Computer and Energy Technologies
20-22 July 2022, Prague

05/05/2022

ACCEPTANCE LETTER

Dear ***Fadi Fataftah, Bassey Isong,***

Thank you for your submission to the ICECET 2022 conference. We are pleased to inform you that your paper entitled **"ID-110: Security Issues and Possible Solutions in Cloud Computing and Big Data: A Review"** has been accepted as a full paper for **oral presentation** by the conference committee of *International Conference on Electrical, Computer, and Energy Technologies (ICECET).* The event will take place in Prague, Czech Republic on 20-22 July 2022 **online** and **physically.**

According to the conference regulations, only those papers which have been duly registered and presented on the conference day are considered for submission to IEEE Explore. The conference program will be communicated in due course.

We look forward to seeing you for a fruitful research and innovation event and for a great time in the wonderful environment of Prague

Yours sincerely,

Dr. Simon Winberg
Chair

Web: http://www.icecet.com E-mail: icecet.conference@gmail.com

APPENDIX D: PROOF OF PAPER SUBMISSION

Gmail

Fadi Fataftah <fadi.fataftah@gmail.com>

Submission Confirmation for PONE-D-22-16714 - [EMID:6cab57c135daafca]

PLOS ONE <em@editorialmanager.com> Thu, Jun 9, 2022 at 7:25 PM
Reply-To: PLOS ONE <plosone@plos.org>
To: Fadi Fataftah <fadi.fataftah@gmail.com>

PONE-D-22-16714
A Comparative Risk Assessment Processes for Big Data Based on Cloud Computing: a case study of four countries
PLOS ONE

Dear Mr. Fataftah,

Thank you for submitting your manuscript entitled 'A Comparative Risk Assessment Processes for Big Data Based on Cloud Computing: a case study of four countries' to PLOS ONE. Your assigned manuscript number is PONE-D-22-16714.

We will now begin processing your manuscript and may contact you if we require any further information. You will receive an update once your manuscript passes our in-house technical check; you can also check the status of your manuscript by logging into your account at https://www.editorialmanager.com/pone/.

If during submission you selected the option for your manuscript to be posted on the bioRxiv preprint server (http://biorxiv.org), we will be assessing the manuscript for suitability shortly. If suitable, your preprint will be made publicly available on bioRxiv and you will receive an email confirmation from them when it has posted. Please check your response to this question and email us as soon as possible at plosone@plos.org if it has been answered incorrectly. Further information about our partnership with bioRxiv to facilitate the rapid availability of life sciences research is available at http://journals.plos.org/plosone/s/preprints.

If you have any inquiries or other comments regarding this manuscript please contact plosone@plos.org.

Thank you for your support of PLOS ONE.

Kind regards,
PLOS ONE

APPENDIX E: CONSENT FORM

Consent form

TITLE:

Risk assessment processes for big data based on cloud computing technologies: a comparative study.

	Initial
1. I confirm that I have understood the information provided for the above study and have had the opportunity to ask questions.	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reasons or to decline to answer any particular question in the study.	☐
3. I agree to take part in the study titled above.	☐
4. I agree to the interview being audio recorded.	☐
5. I agree to the use of quotes in publications.	☐
6. I agree to provide information to the researchers under the conditions of confidentiality set out by NWU ethics comity.	☐

First Name	Date	Signature (Optional)
Researcher	Date	Signature/ initials

APPENDIX F: RESEARCH QUESTIONS

Position:
Years of experience:
Place of Work, Country:
Email:
Skype Id:

Title: Risk assessment processes for big data based on cloud computing technologies: a comparative study.

1. Interview introduction

Background information (research overview and consent form)
Length: 60 -90 minutes

2. Written consent

Would you like to participate in this interview?

- Consent was obtained from the study participant
- Consent was NOT obtained from the study participant

Questions:

1) Would you please explain your level of knowledge on:
   a) Big data
   b) Cloud computing
   c) Risk assessment/ Management

2) How do you manage your network?

3) Are big data technologies implemented with your Organisation?
   (If answer for Q3 is yes then skip to 3.3)

   3.1) If no, why has it not been implemented yet?
   3.2) To what extent do you support implementing the technology and Why?
   3.3) What are the advantages/ disadvantages in your opinion?

4) Do you run or plan to run your big data applications on a traditional storage medium or in the cloud? (If the answer is on the cloud follow with) What challenges have you experienced with big data and how have you dealt with them?

5) Are cloud computing technologies implemented within your organisation?
   (If answer for Q5 is yes, skip to 5.3)
   5.1) If No, why has it not been implemented yet?
   5.2) To what extent do you support implementing the technology and Why?
   5.3) What are the advantages/ disadvantages in your opinion?

6) Are specific arrangements (policies) in place concerning your information in case you want to terminate or transfer to the cloud service? Yes / No
(If No ask to explain.) (Based on time and previous answers I may consider discussing why there were no arrangements)

7) Do you have a clear risk assessment plan? If yes, explain more please on the part related to cloud computing and big data.

8) From a technical point of view, can you point out what is covered in your current risk plan based on the following in detail;
a) Interpreting data in transit
b) Data leakage

9) From a legal perspective, what are the risks related to data protection?

10) Is there anything else you would like to add?

Thank you.