



NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Title

**A CENTRALISED HADOOP-BASED FRAMEWORK FOR BIG DATA ANALYTICS IN
PRIME MINISTERIAL OFFICES: A NAMIBIAN CASE STUDY.**

by

Pombili I.P. Haitamba
222143657

Submitted in partial fulfillment of the requirements for the degree of
Masters of Data Science
in the

Faculty of Computing & Informatics
Department of Informatics

at the

NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Supervisor:

Dr. R. Maliwatu

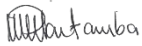
Co-supervisor:

Prof. J. Osakwe

Date of submission:

30/09/2025

METADATA

Study Title	A Centralised Hadoop-Based Framework for Big Data Analytics in Prime Ministerial Offices: A Namibian Case Study
Student Surname, Name	Haitamba Pombili Ipawa Paula
Student Number	222143657
Faculty	Faculty of Computing and Informatics
Department	Department of Informatics
Programme	Master of Data Science (09MADS)
Status	Dissertation
Main Knowledge area	Data Science
Type of research	Applied research
Methodology	Mixed method
Site	Windhoek, Namibia
Documented	30 September 2025
Digital Signature of Student	
Supervisor	Dr Richard Maliwatu
Co-Supervisor	Prof. Jude Osakwe
Digital Signature of Supervisor	
Digital Signature of Programme Coordinator (or equivalent)	
Digital Signature of FCI HDC Representative	

DECLARATION

I, Haitamba Pombili, declare that the work included in this paper presented for the Master of Data Science at the Namibia University of Science and Technology titled “**A Centralised Hadoop-Based Framework for Big Data Analytics in Prime Ministerial Offices: A Namibian Case Study**” is my original work, under the supervision of Dr Richard Maliwatu and I have not previously submitted it for any degree at any other tertiary institutions.

I further declare that I have fully acknowledged sources of information used for the research in accordance with the research institution rules.

Haitamba Pombili

Student Full Name

30 September 2025

Date

RETENTION AND USE OF THESIS

I, Haitamba Pombili, a candidate for the Masters of Data Science, accept the guidelines established by the Namibia University of Science and Technology about the use and preservation of my thesis, which is kept in the Library and Information Services.

I acknowledge that the original copy of my thesis, stored in the Library and Information Services, may be accessed for study and research under these terms. This availability will be subject to the standards for thesis care, lending, or copying set forth by the librarian.

By agreeing to these terms, I acknowledge that my thesis may be used by members of the academic community to further research and understanding in my area.

Signature: 

Date: 30 September 2025

ABSTRACT

Big Data is reshaping the way governments operate, influencing how decisions are made and services delivered. However, most governments, particularly in developing countries, face challenges in managing the growing volume of data generated across public institutions. These challenges can be attributed to a lack of expertise in handling complex and diverse datasets generated in high volume which are becoming difficult for traditional databases to manage.

In Namibia, the Office of the Prime Minister (OPM) plays a significant role in coordinating governance and public service administration. The OPM has a mandate to oversee Information and Communication Technology (ICT) in the Public Service and promote e-governance initiatives, which involve the usage and handling of data. However, despite its central role, data from different government ministries remains fragmented, building up silos. This fragmentation limits the government's ability to extract value from the data, hinders coordination across ministries, and ultimately leads to poor service delivery.

This study addressed the challenges of data fragmentation within the OPM. It explored the feasibility of designing a centralised database system using Hadoop to integrate and analyse big data across ministries, with the goal of improved administrative efficiency, enhanced public service delivery, and promoting e-governance.

Synthetic datasets were generated in Mockaroo, to represent datasets from government ministries. A Hadoop-based setup was undertaken to simulate a centralised database framework integrating all government ministries' data in Hadoop Distributed File System (HDFS) for storage. A MapReduce job was run in Hadoop, using Java code for analyses across ministries, from Ministry of Home Affairs, Immigration, Safety and Security (MHAISS), Ministry of Labour (MoL), and Ministry of Education, Arts and Culture (MEAC). The job compared poverty indicators based on birth year, educational achievement, and employment status across the three ministries and aggregated the results for regional poverty analysis.

The findings of this research show that Hadoop is a cost-effective, open-source framework that has the capabilities to store versatile datasets that currently exists within the Namibian public service into one centralised database that supports big data analysis. The processing layer of Hadoop, MapReduce was able to process a job in minutes that would normally take five to seven working days to complete in the OPMs current administration. Adopting this framework would enable the OPM to make informed decisions backed by evidence, eliminate inefficiencies in public service delivery, and enhance public trust through improved service delivery.

DEDICATION

I dedicate this to my family for their love, motivation, encouragement, and unwavering support along this journey. My most significant source of strength has come from your faith in me.

To my son, for inspiring me every day to set the standard. May this work serve as a testament to the limitless possibilities that lie ahead when we strive for excellence.

ACKNOWLEDGMENTS

First and foremost, I extend my deepest gratitude to God for granting me the wisdom, strength, and guidance throughout this journey. Without His blessings, this achievement would not have been possible.

I am profoundly grateful to my supervisor, Dr Richard Maliwatu, whose expertise, guidance, patience, and encouragement were instrumental in shaping this work. Consulting you about my work always eased the burden of confusion and doubt, your insights were invaluable in giving depth and meaning to this paper. Thank you for always making the time.

And to my co-supervisor, Prof Jude Osakwe, thank you so much for your time and input that has helped piece this work together.

Prof Mamello Thinyane, thank you for lending me a listening ear whenever I needed to bounce off some ideas about my work. Your valuable insights guided me, when I was uncertain about the direction of the research.

The staff of OPM, thank you so much for taking the time to respond to my requests, interviews, and questionnaires. Your responses helped me understand in depth the underlying challenges that played a significant role in grounding the direction of this research.

To my classmates, thank you for your shared experiences and mutual support. Your contributions and encouragement have made this academic journey not only achievable but also enjoyable.

Table of Contents

METADATA	i
DECLARATION	ii
RETENTION AND USE OF THESIS	iii
ABSTRACT	iv
DEDICATION	v
ACKNOWLEDGMENTS	vi
LIST OF FIGURES	x
LIST OF TABLES	x
ABBREVIATIONS/ACRONYMS	xi
1. INTRODUCTION	1
1.1 Background	1
1.2 Research Problem	4
1.3 Research Aim	5
1.4 Objectives and Questions	5
1.5 Scope of the study	6
1.6 Significance of the Study	6
1.7 Structure of the Dissertation	7
2. LITERATURE REVIEW	9
2.1 Introduction	9
2.2 Centralised Database	9
2.2.1 Centralised Database Compliance with Big Data	12
2.3 Big Data in Public Administration	15
2.4 E-governance in Namibia	17
2.4.1 Current State of Data Management in Namibia	17
2.4.2 Benefits of Data Integration	18
2.5 The Case of Hadoop and Big Data Technologies	20
2.5.1 Eliminating Data Redundancy	22
2.5.2 Real-Time Data Processing	23
2.5.3 Standardising Data Entry Across Ministries	24
2.6 Research Gap	25
3. RESEARCH METHODOLOGY	27
3.1 Introduction	27
3.2 Research Design	27
3.3 Population and Sample Size	28
3.3.1 Population	28
3.3.2 Sample Size	28

3.4 Data Collection Methods	28
3.4.1. Interviews	28
3.4.2 Questionnaires.....	29
3.4.3 Synthetic Ministerial Data	29
3.5 Reliability	30
3.6 Validity	31
3.7 Data Analysis Technique	31
3.7.1 Qualitative data analysis	32
3.7.2 Quantitative data analysis.....	32
3.8 Ethical Considerations	33
3.9 Chapter Summary	33
4. HADOOP SETUP.....	35
4.1 Introduction	35
4.2 Implementation Overview	35
4.3 Configuration and Setup	35
4.4 Data Preparation	37
4.5 Ministries Directories in HDFS.....	38
4.6 MapReduce	39
4.6.1 MapReduce Job Structure	40
4.6.2 Compilation and Packaging.....	40
4.6.3 Running the Job	40
4.7 Visualising the Centralised Hadoop Setup	42
4.8 Chapter Summary	46
5. RESULTS AND EVALUATION	47
5.1 Introduction	47
5.2 Data analysis overview	47
5.2.1 Additional benefits of Hadoop	48
5.3 Results of the study	48
5.3.1 Findings on Current Data Management in Namibian Ministries	48
5.3.2 Findings on Inefficiencies within the OPM	51
5.3.3 Benefits of a centralised database with Hadoop for administration	56
5.4 Implementation of a Hadoop-based centralised system	58
5.4.1 Evaluation of the proposed system.....	58
5.5 Chapter Summary	60
6. CONCLUSION AND RECOMMENDATIONS.....	62
6.1 Introduction	62
6.2 Findings of the Study	62
6.3 Recommendations	63

6.3.1 OPM	63
6.3.2 MICT	64
6.3.3 Department of Public Enterprise	64
6.3.4 Government Ministries	65
6.3.4 National Data Governance Authority (To be established)	65
6.3.5 IT Departments across Ministries	66
6.4 Contribution	66
6.4.1 Practical Impact	66
6.4.3 Policy formulation	67
6.5 Limitation of the study	67
6.6 Future Work	67
6.7 Conclusion	68
REFERENCES.....	70
APPENDICES.....	73
Appendix A: Clearance Certificate	73
Appendix B: Letter of Permission from OPM.....	74
Appendix C: Research Questionnaire	75
Appendix D: Raw Data	77
Appendix E: Hadoop 3-node Cluster Configurations.....	79
Appendix F: Java Codes	86
Appendix G: PovertyAnalysis Output File	89

LIST OF FIGURES

Figure 2.1: 5Vs of Big Data	13
Figure 2.2: Centralised Database Systems	16
Figure 2.3: E-governance Roadmap in Namibia	18
Figure 2.4: Benefits of a Centralised Database in the OPM	20
Figure 2.5: Hadoop Components	21
Figure 4.1: Hadoop Cluster Virtual Environment setup	36
Figure 4.2: Ministries Directories in HDFS	38
Figure 4.3: Structured and Unstructured Data Formats in HDFS	39
Figure 4.4: Running MapReduce Job in Hadoop	41
Figure 4.5: MapReduce Analysis output	41
Figure 4.6: Hadoop Setup Visualisation	43
Figure 4.7: MapReduce Job Processing	45
Figure 5.1: Challenges of Current Systems in Ministries	50
Figure 5.2: Benefits of Current Systems in Ministries	50
Figure 5.3: Challenges of Current Systems in the OPM	53
Figure 5.4: Benefits of Current Systems in the OPM	53
Figure 5.5: Channels of Communication	55
Figure 5.6: MapReduce Job Processing Time	56
Figure 5.7: Regional Poverty Analysis Visualisation	58
Figure 5.8: Benefits of Centralising Data	59
Figure 5.9: Ratings on Centralised Database System Implementation	60

LIST OF TABLES

Table 2.1: Comparative Analysis with Estonia and Kenya.....	11
Table 3.1: Summary of Methodology Activities	33
Table 4.1: Summary of Hadoop Multi-node Configurations	36
Table 5.1: Proposed Systems Evaluation	59

ABBREVIATIONS/ACRONYMS

AFIS: Automated Fingerprint Identification Systems

ASYCUDA: Automated System for Custom Data

CPU: Central Processing Unit

DHIS: District Health Information System

DPSITM: Department of Public Service Information Technology Management

HEMIS: Higher Education Management Information System

ICT: Information Communication Technology

IHCMS: Integrated HealthCare Management System

ITAS: Integrated Tax Administration System

JAR: Java Archive

JDK: Java Development Kit

JSON: JavaScript Object Notation

MHAI: Ministry of Home Affairs and Immigration

MIS: Integrated Management System

MOHSS: Ministry of Health and Social Services

NaSIS: Namibia School Feeding Information Management System

NPRS: National Population Registration System

NSFAF: Namibia Student Financial Assistance

NSFP: Namibia School Feeding Program

OMAs: Office of Ministry, Agencies

RAM: Random Access Memory

SDGs: Sustainable Development Goals

SQL: Structured Query Language

TVET: Technical and Vocational Education and Training

UNCTAD: United Nations Conference on Trade and Development

VM: Virtual Machine

XML: Extensible Markup Language

YARN: Yet Another Resource Negotiator

1. INTRODUCTION

1.1 Background

In this digital era, data has become a crucial asset in shaping effective governance and driving socio-economic development. Governments around the world are increasingly turning to data-driven approaches to make informed policy decisions, optimise service delivery, and enhance transparency. However, for many developing countries, the challenge lies not in the absence of data, but in the lack of infrastructure and systems to effectively manage, integrate, and analyse existing and incoming data. Due to the variety in sources, huge volumes, and the complexity of the data, known as Big Data, (El Mohajir et al., 2016).

Oyeniran et al., (2019), stated that increasing data is a valuable resource for every organisation, and its expansion and effective management have gained popularity recently, creating research opportunities. However, for data of this capacity to be considered a valuable asset, it needs to be stored efficiently and analysed thoroughly. Modern technologies have replaced traditional storage systems, enabling institutions to analyse large datasets for evidence-based coordination and policy making.

The Office of the Prime Minister (OPM) stands at the pinnacle of the government structure in the Republic of Namibia. According to the Customer Service Charter by the Office of The Prime Minister, (2023), they are responsible for administration and coordination on behalf of the Namibian Cabinet and the Public Service, ensuring result-driven service delivery to citizens. This means their role requires them to interact with 21 government ministries to fulfil their mandate of public service delivery. Each of these ministries deals with data daily and store it in isolated systems, managed independently, and rarely shared across departments. This fragmentation results in duplicated efforts, delays in service delivery, inefficiencies in public administration, and poor coordination in national planning. The Namibia Statistics Agency, (2023), reported a population increase of 30.08% between 2011 and 2023, reaching a total of 3.02 million people, reflecting an annual growth rate of 3.0%. This increase means that there is an influx of data that each ministry captures and handles daily, as a result of more people requiring public services.

However, the ministry's current data management practices are becoming ineffective in handling the kind of data that is growing at such an increasing rate, due to scalability issues (Nengomasha et al., 2018). And it is becoming even more challenging to extract any value, to

get a holistic overview on the nation's state of affairs, from historic and present data with the current traditional database systems. There is a growing need to serve better and tailor service delivery to suit the growing needs of every Namibian citizen. Yet, without a centralised and scalable system to manage and analyse cross-ministerial data, the OPM faces significant limitations in its ability to achieve efficiency, and enhance decision-making backed by data.

Information gathered by the researcher revealed that, the OPM currently has physical servers where it hosts some of the government's ministries data backups, emails, websites, and ensures security measures are in place to protect against malicious attacks and unauthorised access. There are also separate departments that are responsible for handling different operations under the same Ministry, with no central point of reference and no collective data for informed decision making either. For instance, for the OPM to administer a more effective and targeted approach to distributing drought relief resources, they would need to analyse from various different ministries, which should be readily accessible for emergency cases. To effectively carry out administration operations and improve service delivery, the OPM needs to employ advanced technologies that can handle large volumes of unified data and adopt a framework that can extract value for better coordination and planning and decision making.

Chen et al., (2019), describes the term "big data" as the collection, examination, and storage of large data sets that exceed the scale and complexity of typical data processing and traditional data management systems. He further added that governments face increasing pressure to deliver personalised, efficient services like the private sector but struggle to manage the rapid growth of Big Data due to resource constraints, outdated systems, security concerns, and a lack of expertise. To analyse governance growth over time and make decisions based on historical data, it is necessary to store this data, which might be challenging, within the traditional systems. However newer technologies such as Hadoop, provide a scalable, fault-tolerant, and affordable way to store and deal with huge amounts of historical data, both structured and unstructured.

According to Gomathy, (2022), the Hadoop ecosystem is a set of tools or a platform that provides a variety of services to deal with big data challenges providing a large-scale data processing framework. Gomathy (2022), further emphasised that Hadoop supports all formats of datasets, and it is scalable which means it can handle growing datasets, as well as allows for simultaneous data analysis. This framework has the potential and full capacity to handle data

from all 21 ministries in one centralised location, which is what this paper proposes for the OPM, to enable them to carry out administrative operations efficiently. According to Mohammed and Saleh (2017), a centralised database typically displays data or information in a specific location within a network.

While the need to centralise and integrate big data is widely recognised, the technology chosen to implement such a system must be scalable, cost-effective, and suitable for handling diverse data sources across government institutions.

Traditional Relational Database Management Systems (RDBMS), such as MySQL, PostgreSQL, and Oracle, are highly effective for structured data but often struggle to scale or efficiently handle unstructured, semi-structured, and high-volume data, particularly as generated across multiple ministries and agencies, (Silva et al., 2016). Garzon, (2024), highlighted that platform used for modern data warehousing, such Google BigQuery, Amazon Redshift, and Snowflake, have expanded their support for semi-structured and unstructured data; however, they still require schema definitions and often involve additional Extract, Transform, Load (ETL) steps for handling accurate unstructured content. He further emphasised that these solutions are usually proprietary, incur substantial licensing and infrastructure costs, and require significant technical support.

Cloud-based technologies, such as Microsoft Azure, Amazon Web Services (AWS), and Google Cloud Platform, offer advanced data storage, analytics, and integrated AI/ML capabilities, (Somasundaram, 2020). However, governments must consider data sovereignty, since cross-border cloud storage can expose sensitive data to foreign jurisdiction and surveillance, complicate compliance with local data protection laws, and create long-term dependencies on third-party vendors, as highlighted by Scherenberg et al., (2024). Additional challenges include internet reliability and budget constraints, all of which can impact sustainability and control over public sector data management.

Hadoop is an open-source, cost-effective framework that offers centralised databases with improved scalability, enhanced data consistency, reduced redundancy, and improved data security, (Güvenoğlu & Razbonyalı, 2016). By integrating various data sources from different ministries into a single, secure system, the OPM would be able to make data-driven decisions that enhance public policy, resource allocation, and deliver customer-centric services. This research explores how a centralised database using Hadoop can transform the OPM's current data management practices and leverage big data analytics, making it more responsive to the needs of citizens, businesses, and other stakeholders.

By simulating Hadoop's capabilities on synthetic datasets that replicate real-life citizen records, this study demonstrated how big data integration can support Namibia's broader development goals and enhance the lives of its citizens through smarter, more responsive governance. The OPM will be able to carry out analysis on budget allocation, drought resource allocation, poverty distribution, unemployment rates, and disease trends outbreaks at district to national level from a central point of reference. This results in effective administrative processes, better decision-making, and improved public service delivery for all citizens. It also helps the OPM achieve its mandate and work towards achieving Namibia's Vision 2030, of becoming a prosperous and industrialised nation.

1.2 Research Problem

The OPM in Namibia is responsible for overseeing public administration and coordinating ICT and e-governance initiatives across government ministries. In fulfilling this role, the OPM collects vast amounts of data from various ministries and departments. However, this data is often fragmented, siloed, inconsistently formatted, and stored on disparate databases, which makes it difficult to aggregate, analyse, and utilise effectively for strategic decision-making.

According to Oyeniran et al., (2019), the absence of a centralised data infrastructure presents significant challenges in data security, accessibility, and privacy, especially in environments where no unified system exists to uphold compliance and governance standards. In the case of Namibia, where a national data protection bill is still pending implementation, these concerns are even more certain.

This lack of integration contributes to a range of administrative inefficiencies at the OPM, including duplicate records, delayed service delivery, poor inter-ministerial coordination, and slow policy response to pressing socio-economic issues such as unemployment, poverty, and regional inequality. Furthermore, the traditional systems currently employed by government entities are often not equipped to handle the volume, variety, and velocity of modern data.

As highlighted by Abuljadail et al., (2023) big data requires a suite of specialised tools and frameworks for meaningful analysis, which traditional relational databases are unable to support efficiently. There is, therefore, a pressing need for a centralised, scalable, and open-source data architecture that supports not only integrated storage but also real-time data analytics, secure access, and inter-ministerial collaboration.

This study investigated the feasibility of designing a centralised database using Hadoop, an open-source framework known for its capacity to store and process large datasets in a distributed and fault-tolerant manner, (Poladi, 2023).

By simulating and proposing a Hadoop-based system, the study aimed to demonstrate how such an architecture could help the OPM overcome current inefficiencies, enable data-driven governance, and ultimately enhance public administration and service delivery in Namibia.

1.3 Research Aim

This research investigates the design of a centralised big data framework using Hadoop to support integrated data analysis across government ministries, to transform the administrative functions of the OPM. The study combines ministries' data management practices with big data technology adoption frameworks, offering a novel approach for analysing digital transformation in resource-constrained environments.

1.4 Objectives and Questions

	Research Objectives	Research Questions
Main	To design and evaluate a centralised database framework using Hadoop for integrating fragmented big data across Namibian government ministries, supporting data-driven decision-making and improving administrative efficiency within the OPM.	How can a Hadoop-based centralised database system be designed and applied to integrate big data across Namibian government ministries, improving administrative efficiency and supporting data-driven governance in the OPM?
Sub		
1	Examine the current state of data management across the 21 Namibian government ministries.	What are the current data storage and management practices across Namibian government ministries?
2	Identify challenges and inefficiencies caused by decentralised and inconsistent data storage.	What inefficiencies arise from the lack of centralised data integration at the OPM?

3	Explore the potential benefits of a centralised database system for administrative optimisation.	How can a Hadoop-based system enhance data integration and decision-making?
4	To design a prototype Hadoop-based centralised database system that supports cross-ministerial data processing and analytics	How effective is a Hadoop-based centralised database system in processing and analysing cross-ministerial data for public service delivery?

1.5 Scope of the study

The study was limited to analysing the potential benefits and challenges associated with integrating data for a centralised database framework for the OPM’s administration only.

1.6 Significance of the Study

This study will help inform the current data management challenges within government institutions, primarily due to the lack of data integration across ministries, which results in inefficiency in OPM’s administrative operations. A centralised database is more citizen-centric and will promote coordination amongst government ministries to enhance citizen public service delivery. It will also serve as a stepping stone towards achieving the e-governance strategic action plan.

The study addressed a critical gap in the literature by focusing on Namibia, where less than 5% of big data governance studies target African governments, (Martinez-Mosquera & Luján-Mora, 2019).

It lays the foundation for a replicable blueprint for African Union member states struggling with similar data fragmentation, aligning with SDG 16 (Peace, Justice, and Strong Institutions) through corruption-resistant, evidence-based policymaking.

It also proposes a transition from traditional RDBMS models to newer technology, an open-source Hadoop integration for public governance, combining distributed storage (HDFS), schema-on-read standardisation, and big data analytics.

1.7 Structure of the Dissertation

Chapter 1: Introduction

This chapter introduces the study by outlining the background and context of data fragmentation challenges within the Namibian OPM. The problem statement highlights inefficiencies arising from siloed data storage and the lack of a unified system to support real-time decision making. The research aim, objectives, and questions are introduced, followed by the significance of the centralised system to support data-driven governance.

Chapter 2: Literature Review

The literature review adopts a narrative approach to explore theories and practices in public sector data management. It highlights the challenges of fragmented data, the rise of big data technologies, and their role in open-source frameworks, such as Hadoop. The review identifies gaps in the Namibian context, particularly the absence of scalable, interoperable systems and limited policy frameworks supporting integrated data analytics, thereby justifying the need for this research.

Chapter 3: Research Methodology

This chapter presents a mixed-methods research design used in this project. Data was collected via interviews and questionnaires to assess current practices. Synthetic CSV datasets were generated for a MapReduce job to demonstrate cross-ministerial data analysis using Hadoop. Measures to ensure reliability and validity, such as piloting, are discussed, alongside ethical considerations related to data privacy and informed consent.

Chapter 4: Hadoop Setup

This chapter provides an overview of Hadoop 3-node configurations on a virtual machine. Government directories were created, and different file formats were loaded and stored in HDFS to demonstrate the setup of the different government ministries' databases in a centralised framework. A MapReduce job was successfully run to demonstrate Hadoop's capabilities in supporting combined data analysis for improved service delivery.

Chapter 5: Results and Evaluation

The chapter reports on the findings, which reveal that data across Namibian ministries is highly fragmented, inconsistently formatted, and stored in non-integrated systems, leading to inefficiencies such as duplicated efforts, delayed policy responses, and poor service delivery.

The chapter also presents Hadoop as a system that can solve all these challenges by integrating data into a single framework and eliminating the need for intermediaries in data collection. It also showcases Hadoop's capabilities to perform analysis faster and more efficiently.

Chapter 6: Conclusion and Recommendations

This chapter concludes the dissertation and provides recommendations for the relevant stakeholders. These include adopting Hadoop for centralised data, creating interoperability standards, implementing a data protection framework, and building technical capacity for big data analytics within government systems.

2. LITERATURE REVIEW

2.1 Introduction

This chapter presents a narrative literature review for the study by exploring relevant research related to the application of centralised databases in governance. It demonstrates the relationship between existing knowledge and the areas of inefficiencies and how leveraging big data through a central database will improve service delivery in Namibia's OPM. The chapter explores the conceptual understanding of interdisciplinary research areas, including public sector data management and integration, the emerging role of big data technologies, and e-governance, thereby setting the stage for examining Hadoop-based solutions in the Namibian context.

2.2 Centralised Database

A centralised database keeps information or data in one specific place on a network. It enables the collection and storage of data from pre-existing databases in a single database for internal sharing, analysis, and upgrading, (Mohammed & Saleh, 2017).

Governments act as custodians of administrative data, including birth and death certificates, tax returns, and educational records (Dolan et al., n.d.). This places them in a position to make decisions on the responsible collection, handling, and dissemination of such data. Governments, businesses, and citizens can all benefit from national data sharing systems in various ways, including enhanced public service and socioeconomic growth.

A centralised database implementation within the OPM will enable the integration, processing, distribution, and storage of data from different government ministries. Although extensive prior research has been done on this subject, it has not kept pace with the latest technological advances, such as the use of Hadoop technologies to leverage data from various sources for storage and analysis, which in turn becomes big data.

According to the eGovernment Strategic Action Plan for the Public Service of Namibia by the Office of the Prime Minister, (2014), different ministries employ varied data formats, storage methods, and classification systems, making integration extremely challenging and leading to compliance risks. There are no standard metadata definitions, resulting in conflicting reports on key metrics, such as poverty rates, unemployment figures, and regional population statistics.

For instance, population figures from the Ministry of Home Affairs, Immigration, Safety, and Security (MHAISS) may differ from those provided by the Ministry of Health and Social Services (MHSS), making it difficult to produce reliable national reports. A study by Nengomasha et al., (2018), identifies a major issue in the reliance on birth certificates as proof of citizenship. Because of this, the government is sometimes reluctant to issue birth certificates, such as when the parents don't have the necessary paperwork. Additionally, clerks are sometimes given permission to issue "Namibian" and "non-Namibian" birth certificates without the necessary supervision. The World Bank Group, (2016), Identity Management Systems Report for Namibia revealed that, the MAIHSS's use of paper-based, registers for recording patients' illnesses without standardised identity numbers leads to fragmented patient data, making it hard to track individuals and causing more duplication than under-registration.

The impact of fragmented data systems was highlighted during the COVID-19 pandemic, where timely and accurate data became critical for public health responses around the globe. According to the World Health Organization, (2020), report on COVID-19 updates, data was the key for providing broader insights to the pandemic's effects in numerous nations, with statistics dashboards displaying numbers across various platforms to inform on the status and spread of the pandemic in each country across the world from one system. The availability of data in a unified system made it easy to provide constant and precise live data.

Similar issues of fragmentation are not unique to Namibia. Mohammed & Saleh, (2017), observed that despite Nigeria's government and agencies' efforts to establish databases at various organizations gathering citizen data, including biometrics, a lack of integration between them has resulted in redundancy and inconsistencies, with people giving contradicting information on various platforms. Integrated government data reduces risk by minimising manual input errors, ensuring consistent records across registers, and enabling fraud detection through advanced analytics tools. Estonia is at the forefront of this field and uses satellite imagery and data from agricultural registrations to determine if land that receives government subsidies is being used for farming (Axel Domeyer et al., 2021).

According to Rutto & Yudah, (2015), the Nairobi-based Huduma Centres offer a one-stop shop for public services, streamlining processes by enabling citizens to access multiple services through a single, coordinated entity, rather than visiting various ministry offices. These are

some of the examples that demonstrate how centralised systems can enhance data integrity, reduce redundancy, and improve efficiency in public service delivery.

Rwanda’s centralised e-governance system, Irembo, is another platform that boasts a good turnaround time, offering fast and efficient service delivery, Rwanda, (n.d.). Another study on Germany found that completely integrated and interoperable government data might reduce case-processing times for important public services by 60% (Axel Domeyer et al., 2021).

According to comparative analysis, nations such as Kenya, which established the Huduma Center to offer residents of Kenya prompt and effective government services, have had great success in providing their population with satisfactory services. The former Youth and Gender Affairs Cabinet Secretary, Sicily Kariuki, indicated that the centres serve at least 30,000 customers daily, and have a turnaround time of 10 minutes (Rutto & Yudah, 2015). The report goes on to say that the platform offers a variety of services, which makes it easy for individuals to access them instantly without having to go to different locations to acquire public services. The table below gives an overview on Estonia and Kenya, before and after implementing centralised systems.

Table 2.1: Comparative Analysis with Estonia and Kenya

	Estonia	Kenya
Response time before	<ul style="list-style-type: none"> • Tax processing 3-6 months • Benefit approval, 8-12 weeks • Crisis response time 48hrs 	<ul style="list-style-type: none"> • Document processing 2-4 weeks • Queue waits 4-6 hrs
Implementation	X-road in 2003	Huduma Centres in 2013
Features	<ul style="list-style-type: none"> • 99% services are digital • Blockchain-secured data exchange • Real-time dashboards 	<ul style="list-style-type: none"> • Physical one-stop-centres • Integrated payment systems • Digital queue management

Services integrated	900+ public services across 13 ministries	100+ services from 26 agencies
Response time after	Tax processing is less than a month Crisis response time 2 hrs	Same day service completion Average wait time: 15 minutes
Corruption impact	87% reduction in data tampering incidents	62% decrease in petty bribery reports

The analysis above demonstrates that centralised systems reduce administrative response time delays by 60-85%, with Estonia's digital-first model showing 3x greater efficiency gains than Kenya's hybrid approach.

2.2.1 Centralised Database Compliance with Big Data

The potential of centralised databases to improve data consistency across ministries, reduce redundancy, and streamline data storage and retrieval has been the subject of numerous studies. The Central Database serves as a means of facilitating data integration and exchange among different ministries, (Saleh et al., 2013). This aligns with the aim of removing irregularities and redundant data while optimising the use of resources.

Local government ministries capture and utilise a wealth of data, such as resident information, budget and finances, in their day-to-day operations, which becomes difficult to store and process using traditional databases. As such modern advanced solutions are required to extract the value needed from these data. In public organisations, "data culture" refers to viewing big data as something that needs support from organisational-wide structures and skills rather than just being an IT problem, (Giest, 2017).

According to Omar and Jumaa, (2019), big data's main characteristics are identified by the 5Vs: Volume, Variety, Velocity, and Veracity, and Value.

- a) **Volume:** unpredictable scale of data being produced, which demands a large storage volume (Jadhav, 2018). The advantage of handling and analysing such large volumes of data is that insightful information about businesses and citizens can be derived.

- b) **Variety**: describes gathering various types of data via various technologies, including documents and multimedia data. Moreover, these data types may be semi-structured, structured, or unstructured (Omar & Jumaa, 2019).
- c) **Velocity**: describes how quickly data is generated (per second/minute). Throughout the data collection process, the contents of the data frequently change, resulting in various forms that originate from multiple sources. From this point of view, new technologies and methods are required to explore the flowing data adequately (Omar & Jumaa, 2019).
- d) **Veracity**: relates to the precision or accuracy of the knowledge (e.g., helpful and instructive) (Omar & Jumaa, 2019). The data generated may contain errors, primarily due to the use of unreliable sources.
- e) **Value**: includes the process of extracting valuable information from massive datasets. Value is the most important component of any big data solution since it allows for the production of helpful information (Omar & Jumaa, 2019).

The 5Vs of Big Data are summarised in the figure below.

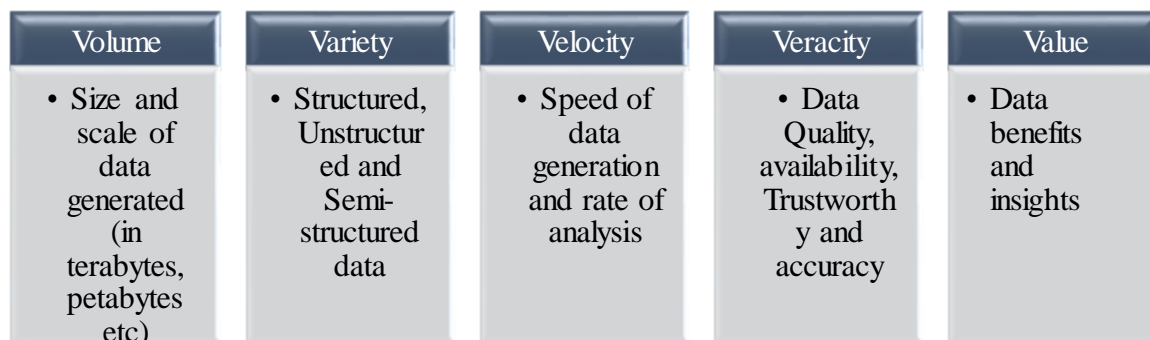


Figure 2.1: 5Vs of Big Data

Jadhav (2018) highlights that big data is too large to store, generated at a fast pace, and too difficult to handle, with traditional databases unable to effectively collect, process, manage, and analyse. Therefore, the solution to processing, organising, and evaluating large volume of data over the network or internet is big data analytics. However, implementing centralised databases for big data compliance presents unique challenges due to the scale, diversity, and regulatory complexity of large datasets. Concerns have also been raised over the best ways to control big data, use it efficiently in digital public services, and integrate it into policymaking at various levels of government. Therefore, strict adherence to regulatory, security, and ethical compliance standards is necessary to ensure that lawful and efficient data management practices are followed.

The risk of document sourcing or doxing, which is the process of assembling disparate pieces of information to discover more about an individual, increases privacy risks because big data is simply huge metadata that may contain personally identifying information, (Abuljadail et al., 2023). According to DLA Piper (2025), comprehensive data privacy law has not been passed in Namibia. Nonetheless, a number of laws tailored to particular industries, such as banking and law, are in place to safeguard customer data. The absence of a fully enacted data protection law in Namibia, calls for a system that aligns with international frameworks such as the General Data Protection Regulation (GDPR) and the Protection of Personal Information Act (POPIA) to safeguard sensitive data. Compliance measures should include robust encryption protocols, multi-factor authentication, and access controls to prevent unauthorised access.

Government ministries use systems that span departments or sectors, each of which has its own data collection systems. This leads to siloed databases that hinder effective data exchange and decision-making. When trying to combine datasets or extract insights from several information sources, the incompatibility of these technologies makes things difficult (Giest 2017). It has been noted that the digitization process has resulted in data format inconsistencies, with data being kept in a variety of formats, and data being recorded in disparate ways. It is difficult to communicate the pertinent information or draw insightful conclusions as a result. Therefore, the proposed centralised database will adopt ETL/ELT pipelines (Extract, Transform, Load/Extract, Load, Transform) to streamline data ingestion and transformation, ensuring scalability.

Abuljadail et al., (2023) stated that open data ethics should be formally defined because the urban informatics system's data collection, administration, distribution, and usage processes affect fairness in democratic decisions about cities. Therefore, the proposed system should employ role-based access controls (RBAC) and attribute-based masking to limit the exposure of sensitive data. This is to ensure that the data is kept safe from unauthorised access and to impose a hierarchical level of control, as it will be a combination of data from all government ministries.

To make well-informed decisions, it is crucial to guarantee that government data is reliable, comprehensive, and consistent. Maintaining high-quality data, however, can be difficult because of factors such as input errors or antiquated legacy systems without automated validation procedures. Data governance frameworks should be established by governments that

include data quality assessments, better training for those who handle the data, and better integration with outside sources for verification. Data use is directly related to the digitization of public services: Numerous opportunities exist for data aggregation and analysis through digital applications, and data analysis can subsequently facilitate the deployment of digital services, (Giest, 2017).

The entire big data processing pipeline revolves around data analysis since it is the analysis that gives big data its worth. Data analysis requires raw data, which is information that has been retrieved and combined from various data sources. It is possible to choose all or some of this data for analysis based on the requirements of various applications, (Hong et al., 2018).

2.3 Big Data in Public Administration

To support the growing needs of Namibian citizens and for an efficient and successful data-driven strategy, this study proposes an interoperable and connected data warehouse utilising Hadoop. Apache Hadoop is an open-source software platform that allows huge datasets to be processed equally over clusters of commodity systems. It is built with a high level of fault tolerance and can expand from a single server to hundreds of computers, (Bhushan Jadhav et al., 2018) .

The relationship between big data and Hadoop was discussed by Al-Sai & Albualigah, (2017), stating that Hadoop is a commonly used technology in the field of big data analysis, operating on a master-slave design and storing data in the Hadoop Distributed File System (HDFS). With Hadoop, each ministry will have a certain level of access control, with an interface to query the system. Each user will be assigned a unique identification number that can be used to retrieve a complete record of information required about citizens from the database.

A typical example of inefficiency in current government data practices can be observed in the Directorate of National Examinations and Assessment (DNEA) under the Ministry of Education. When a citizen requests a duplicate certificate from as far back as 2009, they must manually complete a form with personal details such as name, national ID number, and examination center code, information that should already exist in a centralised database. Despite this, citizens are instructed to wait 5–7 working days while officials manually verify their records. If the data is unavailable at the office, staff must contact individual examination centers to trace the certificate. This process reflects the fragmented nature of government data

systems, which not only delays service delivery but also illustrates the pressing need for an integrated data infrastructure to support real-time access and improve administrative responsiveness.

A centralised database would resolve such scenarios within minutes by maintaining a unified trail of all examination records from the district to the national level. With a single query, such as a national ID, officials could instantly access a candidate’s certificate, eliminating delays, paper-based redundancies, and the need for repeated office visits. This streamlines service delivery, reduces administrative overhead, and minimises reliance on intermediaries. Moreover, it fosters inter-ministerial collaboration, improves transparency, and lays the groundwork for scalable e-government services. Ultimately, such infrastructure strengthens citizen trust, remodels public administration, and positions the country to align with innovative governance initiatives.

Centralising data minimises information silos, enhances data quality, and facilitates the identification and prediction of potential impacts of emerging trends or proposed changes across various ministries. The diagram below highlights the actors, opportunities, tensions and applications of such systems.

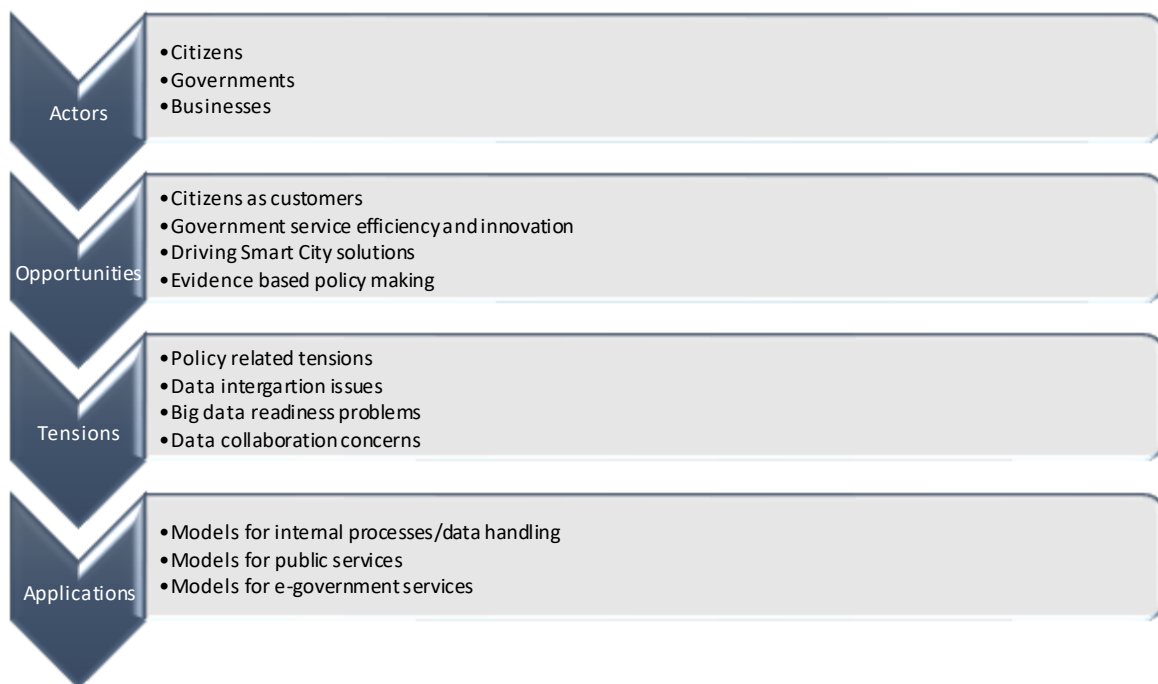


Figure 2.2: Centralised Database Systems

2.4 E-governance in Namibia

2.4.1 Current State of Data Management in Namibia

According to the Center for African Studies, (2023), there is currently no national data center in Namibia, however, the Fourth Industrial Revolution (4IR) task force (consisting of 8 people), appointed by the president in 2021 to evaluate Namibia's preparedness for the Fourth Industrial Revolution, advocated for the establishment of one. The Centre for African Studies (2023) further states that, "Anicia Peters, the chairperson of the task force, noted that the government does not have full control over citizens' data as it is stored abroad". Today, public institutions have isolated data servers that are not integrated, making data sharing between them more difficult.

The COVID-19 outbreak has brought attention to the difficulties in using public data. For instance, email, phone calls, and faxes have frequently been used to aggregate case numbers from hospitals and labs. Accessing and combining data gathered by various government agencies has proven to be very difficult, (Axel Domeyer et al., 2021). Making decisions based on data and gaining full insights are made difficult by these fragmented storage techniques. However, there are a number of barriers that hinder the integration and processing of data from other ministries, some of which may have associations with politics.

Namibia's move toward data governance is gradually taking shape, with the Data Protection Bill of 2022 marking a significant legislative step. The bill asserts that "cross-border transfers of personal data to other countries and international organisations may not take place, unless an appropriate level of data protection is guaranteed" (MICT, 2021). Despite this, many government institutions still rely on manual systems for capturing and storing citizen information. As records accumulate, this fragmented and paper-based approach becomes increasingly complex to manage, resulting in inefficiencies, rising administrative costs, and delays in service delivery.

Digitising these records through ICT infrastructure creates a comprehensive "big data trail" that tracks interactions, enables efficient data retrieval, and enhances transparency. According to (Roger & Schuster, 2023) such digital footprints provide vital opportunities to strengthen public administration by enabling data-driven decision-making.

Although e-governance is not a new concept in Namibia, its implementation has been inconsistent. However, a foundational roadmap has been established since the 1990s, laying

the groundwork for various ICT and digital governance initiatives. Building upon this progress, a centralised big data system powered by open-source technologies, such as Hadoop, could significantly enhance administrative capacity, improve inter-ministerial coordination, and accelerate the country's e-governance initiatives. The figure below displays a roadmap of what has been done thus far, dating back to 1990.

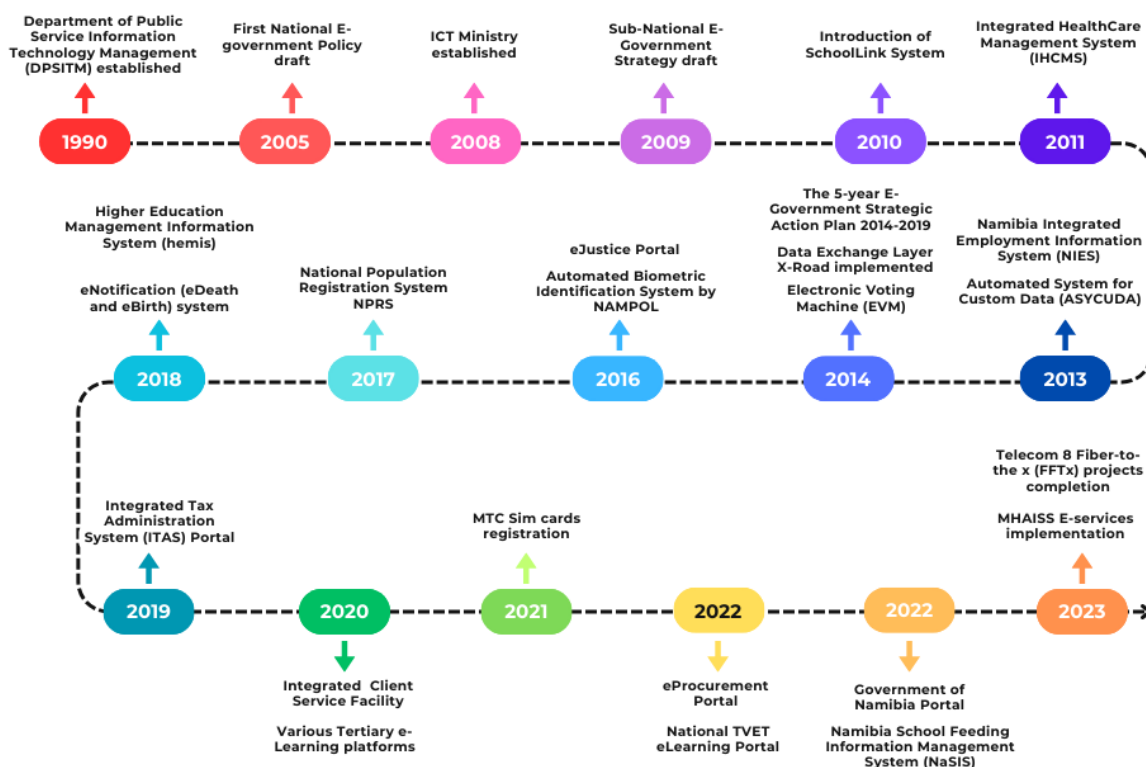


Figure 2.3: E-governance Roadmap in Namibia

2.4.2 Benefits of Data Integration

Integrating data from different ministries to a centralised database using Hadoop technology can address many challenges faced by governments today. Benefits include enhanced decision-making, improved service delivery, targeted interventions for poverty and hunger eradication, and more effective monitoring of health and education outcomes. According to the UN, the significant data era has arrived, and with the abundance of data resources available to people, both new and old, it is possible to conduct novel real-time population analysis.

Based on the idea of large-scale systems, streamlining and optimizing government work processes is essential to creating an innovative, service-oriented government. Cross-departmental collaboration and knowledge sharing are the objectives of the information service. Using new technology as the foundation, resource integration and application deepening, as well as decision-making science, accurate management, and efficient service delivery, should be encouraged. The new era of "innovation" service implies that e-government should prioritise enhancing active, public-centred online services, successfully promoting a comprehensive one-stop shop for underrepresented communities and effectively promoting online interactions (Hong et al., 2018). It was acknowledged in Namibia's 2005 e-Government Policy and the e-Government Strategic Action Plan 2014-2018, that e-government information systems must be integrated and interoperable, and that standards and guidelines must be established.

Prior research has noted the rise in social media usage in the US and Ireland as an illustration of the early warning indicators of growing unemployment, suggesting that the government should "come in" and act swiftly if it can rationally evaluate the data resources at its disposal. Additionally, the research describes how governments may make greater use of big data to adapt to shifts in social and economic indices, including food prices, unemployment, and income, (Hong et al., 2018).

Big data can transform areas of inefficiency in the OPM, significantly enhancing its administrative capabilities and strategic decision-making. It can reduce costs, improve the time required to resolve inquiries, and deliver good service tailored to meet the growing needs of citizens. The insights gained from big data analytics can directly support the achievement of various Sustainable Development Goals (SDGs), which the government aims to meet as stated in the E-Governance policy for the Public Service in Namibia.

Ultimately, the benefits of integrating data analytics include better decision-making, improved service delivery, targeted resource allocation for poverty and hunger eradication, and more effective monitoring of health and education outcomes. There are many benefits that can be derived from centralised databases, a few are summarised below.



Figure 2.4: Benefits of a Centralised Database in the OPM

2.5 The Case of Hadoop and Big Data Technologies

A Hadoop-based system offers a scalable, fault-tolerant, and efficient solution for integrating data across Namibia’s government ministries, from the district level to the national level. The OPM’s e-governance centre would manage this. By leveraging distributed storage and processing, Hadoop enables the OPM to consolidate fragmented datasets, addressing key challenges in the ministry’s administrative operations.

Hadoop is a popular and widely used platform for large data analysis. It has a master-slave architecture and stores data in the HDFS. HDFS supports read, write, rename, and add processing in parallel (Greeshma & Pradeepini, 2016). It is a distributed file system capable of analysing enormous volumes of data stored in disparate government institutions. As a result, it will function using a master-slave architecture, in which the agent node processes input locally before sending the results to the controller node.

Each data node, which houses information from every government ministry, can function in a tree-like master-slave architecture. The root node, which is linked to the servers of its lower levels, represents the highest level of government ministry. The child nodes will act as slaves to the root node, which will act as the master node. To store data, the master node will have its own HDFS. The Hadoop framework comprises several key components, including the HDFS for storage, MapReduce for processing, and various other tools such as Hive, Pig, and Spark for data manipulation and analysis.

Essentially, the Hadoop framework is divided into three major components: MapReduce, YARN, and HDFS as shown below.

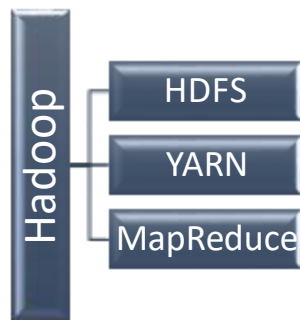


Figure 2.5: Hadoop Components

a) **MapReduce**

This is the model of programming that is utilized for parallel data processing. Large volumes of data can be processed independently on different data clients by breaking them up into smaller pieces. The findings from the different users are then automatically compiled to produce a single result (Omar & Jumaa, 2019).

For instance, data from the Ministries of Health, Finance, and Social Welfare must be used by the OPM to distribute money for healthcare. MapReduce can process healthcare data from these ministries in parallel, combine the results, and present a unified view that includes patient records, financial reports, and social welfare statistics. The OPM will be able to allocate resources more easily with the support of this extensive dataset.

As a result, this will ultimately improve the OPM's decision-making and administrative processes.

b) **YARN**

YARN manages the memory, CPU, and other resources of the Hadoop cluster, (Omar & Jumaa, 2019).

Having various datasets sitting in Hadoop's HDFS that needs to be analysed, when the OPM puts in a job request, YARN would dynamically assign resources to handle these several datasets simultaneously. This ensures effective data collection and processing, enabling the OPM to make informed decisions based on comprehensive and readily available data.

This improves administrative efficiency and decision-making capacities by strengthening the OPM's capacity to centralise and evaluate data from multiple ministries.

c) **Hadoop Distributed File System (HDFS)**

This is Hadoop's main storage system, made to hold huge amounts of files on several computers, called nodes. The file systems are often separated into data and metadata using the HDFS. When compared to the conventional distributed file system, HDFS offers two significant advantages. First, excellent fault tolerance allows for the recovery of data from other data clients by saving copies of the data across multiple data clients, hence allowing for distinct errors to be made. Because Hadoop clusters can store datasets in petabytes, the second benefit is that they enable the use of large data sizes (Omar & Jumaa, 2019).

For the OPM to centralise and analyse data from multiple ministries, HDFS's fault tolerance and capacity to handle enormous data quantities are important. In addition to enhancing the OPM's administrative effectiveness and decision-making capabilities, this ensures reliable data storage and efficient processing.

2.5.1 Eliminating Data Redundancy

Hadoop's Distributed File System (HDFS) enables a centralised repository where all ministries can store structured, semi-structured and unstructured data; this is known as the data storage layer. Structured data follows a fixed schema and is stored in relational databases; semi-structured data lacks a rigid schema but uses metadata for organisation (e.g., JSON, XML), while unstructured data has no predefined format and includes media files and documents, (Bhushan Jadhav et al., 2018).

The duplication of data collection employed across all government ministries is a significant issue in defragmented data systems. Each ministry maintains its records from the district to national level on key national issues, such as population demographics, basic education, employment, health, poverty levels, and tourism, without a centralised platform to reference existing information. Ministries often end up collecting the same data for different purposes, resulting in administrative inefficiencies. For example, tertiary institutions will require national certificates for admission purposes that already exist in the Ministry of Education database, or ID documents that already exist at Home Affairs. Similarly, Namibia Student Financial Assistance (NSFAF) would require the same documents for loan applicants that already exist in another government ministry's database for funding.

Namibia's multiple registration systems, including the legacy South West Africa (SWA) population register implemented by South Africa before independence (1990), create risks of duplicate records despite efforts to ensure unique identities. The Ministry of Home Affairs

Immigration, Safety and Security (MHAISS) continues to consult the SWA register for verification, which compounds procedural inconsistencies between the current and historical systems (Namibia, 2016).

This results in the repetition of existing data, leading to redundancy and increased storage costs, wasted computational resources, and a higher probability of data inconsistencies. According to Nengomasha et al., (2012), The Namibian National Archives, which is designed to preserve records of lasting significance, lacks the space and personnel necessary to preserve large volumes of semi-current records.

Estonia is one of the countries that got it right in implementing an advanced e-governance ecosystem centred around its X-Road platform. The Public Information Act 3 of Estonia forbids the creation of distinct databases for the purpose of gathering identical data. As a result, state institutions are unable to continually request the same personal data if it is already in one of the data repositories linked to X-Road, a platform for data interchange that links all apps related to e-governance, (Vassil, 2015). This highlights the connection between enabling technologies and laws designed to enhance citizen interaction.

HDFS combines data from disparate sources, such as IoT devices, ministry databases, documents, images, and satellite imagery, into a centralised repository, while ensuring seamless integration when needed. Ministries can share information through a common access layer, reducing redundant data collection efforts. HDFS is a scalable, reliable, and manageable solution for handling large datasets. It supports clusters ranging from 10 to 4,000 nodes and is used by major companies like Yahoo, Facebook, Twitter, and eBay for production-scale data processing, (Greeshma & Pradeepini, 2016). This will also address fraudulent activities, as once a citizen's record is updated, it will automatically be reflected in all other databases across all ministries.

2.5.2 Real-Time Data Processing

MapReduce makes it possible to handle big datasets in parallel by dividing them into smaller pieces that may be processed separately by several nodes; which is done in the processing layer. It automatically aggregates results and supports structured, semi-structured, and unstructured data (Omar & Jumaa, 2019). For instance, to deduce analytics for drought relief resource allocation, data should be collected from multiple ministries and may come in different formats. MapReduce will enable the OPM to process data from various ministries' repositories stored in HDFS, identifying high-risk regions, ensuring budget efficiency, monitoring impact, and improving response time, all from a central location.

This will enable the automated combination of analytics pipelines for key indicators, such as poverty rates and drought-affected regions, without requiring manual data requests that may delay the process. Accurate and readily available insights into population demographics and regional resource needs would enable OPM to allocate resources based on evidence.

2.5.3 Standardising Data Entry Across Ministries

Government ministries currently use diverse data formats (e.g., spreadsheets, CSV, PDFs, and various databases) with no standardised classification, making integration difficult. Government databases and portals must import or export big data in a variety of forms to Hadoop and vice versa using the data acquisition layer, which includes components like Flume. Namibia currently lacks comprehensive data protection legislation, which has resulted in a lack of standardised protocols for data sharing and protection, (DLA Piper, 2025).

There are no official limitations on government data gathering or sharing due to the lack of a data protection law, which facilitates possible data breaches and the exploitation of personal data. According to DLA Piper, (2025), the government published a Draft Data Protection Bill in 2021, which aims to establish a Data Protection Supervisory Authority and regulate the processing of personal information. However, this bill has not yet been passed into law.

The lack of standardised protocols has led to inconsistent practices across different ministries, with each ministry operating at different levels. For instance, during the COVID-19 pandemic, businesses and public offices began collecting personal data for contact tracing without a clear legal framework for data retention, sharing, or protection. In some government institutions, personal information is not protected. Because of this, it is now simple for anyone to look up the contact information of clients at different public and private offices, (Nashilongo Gervasius, 2020).

This makes citizens' data easily accessible from different locations, with no proper trace of how this information was leaked. It also makes it hard to protect data that is collected and stored at different government ministries. These ministries share information requests via email and physical files, without following any standard procedures, which also delays the timely delivery of required services to citizens. Data becomes vulnerable at each point of access as it is passed from one location to another across all 21 government ministries.

Flume is used to transform data from various datasets to standardise data ingestion for processing. According to Bhushan Jadhav et al., (2018), Flume offers a dependable, distributed solution with an integrated fault tolerance mechanism for effectively gathering, combining, and transferring massive volumes of log data.

A typical data governance scheme can be established with predefined categories for administrative and decision-making processes (e.g., unemployment stats, healthcare metrics, financial records). This ensures consistency in reporting, allowing OPM to generate accurate, comparable, and evident-based reports.

2.6 Research Gap

Numerous studies demonstrate that this is not an entirely novel concept; rather, it has already been considered and tested. However, little experimental work exists on how these technologies can be practically implemented in developing countries, particularly within African governance contexts. Most existing literature focuses on developed economies with robust digital infrastructures, regulatory frameworks, and established data protection laws. As a result, there is a noticeable lack of context-specific solutions tailored to the administrative, infrastructural, and legislative challenges faced by countries like Namibia.

In the Namibian public sector, data remains largely fragmented across ministries, with little integration or interoperability between systems. Although efforts have been made to digitise some services, a comprehensive, scalable, and centralised data management infrastructure is yet to be implemented. Furthermore, there is limited scholarly exploration on how open-source tools like Hadoop can be leveraged not only to consolidate government data but also to enable data analytics to support evidence-based decision-making in government.

This study addresses this gap by designing and evaluating a Hadoop-based centralised database system for the OPM. It contributes both theoretically, by extending the discourse on big data applications in African governance and service delivery, and practically, by demonstrating a working prototype through simulated multi-ministerial datasets and MapReduce analysis to support administrative decisions. The study also considers Namibia's unique legislative context, where data protection laws remain in draft form, adding a critical layer of inquiry into how such technologies can be adopted responsibly and securely.

2.7 Chapter Summary

This literature study presented a discussion of the advantages and disadvantages of using centralised databases for efficient governance. While challenges in public sector data management are widespread, global trends and case studies affirm the potential of big data technologies to enhance governance. As indicated by the research aims and questions, it paves the way for additional research into the possibility of streaming data into the Hadoop framework from various sources to understand further citizens needs for public services. Subsequent studies will delve further into e-service portals, data security, data governance policies, and the transformation of a fully-fledged e-governance nation in Namibia.

A Hadoop-based system eliminates data fragmentation, enhances evidence-based decision-making, standardises formats, and strengthens security, positioning OPM for data-driven governance. This framework enables the OPM to optimise and deliver on its ongoing projects within the departments of E-governance, National Food and Nutrition Security, the Directorate of Disaster Risk Management, and Drought Relief through a central point of reference.

This study also sets the stage for further research on e-government services and various government portals, resulting from the proposed centralised database in the OPM. Hong et al., (2018) states that in this era of big data, it is crucial to leverage these resources to enhance the government's ability to provide services and make predictions.

3. RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes how the study's research design and methodology were used, and how the findings aligned with the study's goals and questions. Asenahabi, (2019), defines research design as a strategy, methodology, and method of inquiry used to answer research questions while preserving the highest level of control over variables.

To investigate the feasibility and impact of implementing a Hadoop-based centralised database system at the OPM to address administrative inefficiencies, both qualitative and quantitative research methods were used. A mixed-methods research design was adopted to combine qualitative insights from the ministries' staff, with practical simulations and demonstrations of data integration in Hadoop. This approach enabled the researcher to collect information from a variety of sources, aiming to gather more insights on the research questions and address them effectively.

The research was designed to address both the context on administrative practices of data fragmentation across Namibian ministries, how they negatively impact administration at the OPM and the technical setup of Hadoop towards solving these challenges.

3.2 Research Design

The purpose of the research design is to provide a structured foundation for the study which involved both qualitative and quantitative mixed methods approach.

The primary goal of qualitative research is to investigate and comprehend how an individual's or a group's characteristics relate to a social or human issue (Asenahabi, 2019). This part of the study, involved collecting data through semi-structured interviews from OPM as the primary focus, Ministry of Education Arts and Culture (MEAC), Ministry of Justice (MoJ), Ministry of Agriculture, Water and Land Reform (MAWLR), and Ministry of Labour (MoL), to understand current data storage and management practices. A comparative analysis was also done to understand the relevance of the proposed framework, in comparison to Estonia and Kenya that have successfully implemented similar solutions in governance.

The methods and measurements that yield measurable or discrete results are known as quantitative research design (Asenahabi, 2019). In this part of the study, well-structured questionnaires were used to collect data from government officials with the following roles,

Systems Analysts, Systems Administrators, ICT Personnel, Systems Administrators, Analyst Programmers, and Senior Education Officers to quantify the challenges and efficiencies that result from data fragmentation across Namibian ministries.

Further approaches employed in this research involved simulating the integration of centralised database repositories for all government ministries in Hadoop's HDFS, which allowed for a practical demonstration of how it would support integrated data analytics for governance.

This approach was chosen to ensure that the solution is not only technically feasible but also contextually relevant and responsive to the actual administrative needs within the Namibian public sector, thereby facilitating better service delivery.

3.3 Population and Sample Size

3.3.1 Population

A population is the group of people, things, or events that are the subject of research and have similar traits. At the time of the study, Namibia has 21 ministries, including the OPM, which made up the population of the study. Most ministries operate at the district, regional, and national levels, with different departments that utilise various systems, databases, and data capture methods, which are not integrated.

3.3.2 Sample Size

A sample is a collection of cases that forms a subset of the population. The sample of this study consisted of ministries at the national level based in Windhoek, totalling 5, including the OPM. Within these ministries, only the following departments were selected where insights were captured: the Directorate of National Examinations Assessment (DNEA), Planning and Administration, ICT, Finance and Administration, Solutions Architecture, Public Service Information Systems Management, Disaster Risk Management, and the Department of Public Service and E-government Management.

3.4 Data Collection Methods

3.4.1. Interviews

Semi-structured interviews are typically conducted in person, providing the researcher with the opportunity to ask questions for a thorough understanding of the factors and consequences of current work practices. One benefit of doing an interview is that it gives participants the opportunity to discuss subjects the interviewer may not have thought about.

The interviews were conducted with key personnel from the following Namibian government ministries, OPM, MEAC, MoJ, MAWLR, and MoL. The key informants from these interviews were administrative officers, education officers, transport officers, and systems analysts. The goal was to understand current data management strategies, the systems used, and inefficiencies in their roles resulting from the current state of data fragmentation. These interviews provided insights into the human and institutional factors that influence data fragmentation and integration.

3.4.2 Questionnaires

Questionnaires are the primary means of gathering data in practical research since the researcher has complete control over the sample and the questions to be answered. For this study, structured questionnaires were shared via a JotForm to Systems Analysts, Systems Administrators, ICT Personnel, Systems Administrators, Analyst Programmers, and Senior Education Officers from the same ministries where the interviews were conducted. Approximately 60 questionnaires were distributed among the staff of Namibian ministries, and only 25 responses were received. The questionnaires sought to capture the storage formats of data in different ministries departments, challenges and benefits of the computer systems currently being used, number of databases they have and whether they are integrated, how information is shared across departments, how long it takes to provide services internally and externally, data security measures in place, and data backups, as well as how easy it is to access information. The questionnaire also probed into some of the challenges and benefits that the respondents thought would result from a centralised database system and asked for their rating in favour of such a system.

3.4.3 Synthetic Ministerial Data

The interviews and questionnaires provided valuable insights into how the OPM operates in coordination with other ministries. From these findings, it became clear that while each ministry captures data specific to its mandate, much of the information relates to the same citizens tied to a common identifier, such as a *national_id*. However, because this data is currently stored in isolation, linkages across ministries are often missing. The synthetic datasets were therefore designed to reflect this reality-anticipating the types of data individual ministries are likely to capture, with a shared identifier enabling cross-ministerial integration.

The synthetic datasets generated using Mockaroo, each contained 1000 records. MHETI dataset containing the columns, *national_id*, *education_level*, *institution*, *current_enrollment_status*, *region*. MoL contains, *national_id*, *employment_status*,

year_of_employment, job_sector, income_bracket, field_of_study, region. MHAISS dataset contains, *national_id, citizenship, region, gender, ethnicity*. This data was used to perform cross ministerial analysis, on the poverty levels per region, by checking their demographic background and regional tagging from MHAISS, then checks if they have a tertiary education or dropped out in MEAC and if unemployed, self-employed (low income), or income marked as low. If they meet all these criteria, then they flagged higher poverty risk, then Hadoop aggregates all the identified using *national_id* as the primary key their regions, to demonstrate how a centralised database will enable for ministerial collaboration for better insights.

3.5 Reliability

When a measurement device is said to be reliable, it means that it is free from bias or random error and consistently yields the same results under the same circumstances (Sileyew, 2019). In this study, several measures were taken to ensure reliability:

- **Pilot Testing:** The questionnaires and interview guides were piloted with 3 participants from the OPM and 2 from the MOEAC. Feedback from this pilot was used to refine ambiguous questions and enhance clarity, ensuring that respondents consistently interpreted all items.
- **Standardised Administration:** The JotForm survey was distributed across the following government ministries departments: **Ministry of Education, Arts and Culture** (Directorate National Examinations and Assessment (DNEA), National Examinations), **Ministry of Justice** (Planning and Administration), **Ministry of Agriculture, Water and Land Reform** (ICT), **Ministry of Labour** (Finance and administration). The same form was distributed across the following departments in the **OPM**: Solution Architecture, ICT, Public Service Information Systems Management, Disaster Risk Management, Public Service E-government Management, Administration and Information Technology Management. This standardisation minimised the risk of variation in how questions were presented or understood by different respondents.
- **Transparent Coding:** The Java code developed for the MapReduce poverty analysis was thoroughly commented, providing a clear, step-by-step explanation of the logic used. This transparency allows others to review, replicate, and verify the processing steps, enhancing the dependability of the technical component.

- **Consistent Findings:** Feedback from government officials consistently highlighted inefficiencies in the current decentralised system and recognised the potential benefits of centralised data integration, supporting the reliability of the qualitative findings.

3.6 Validity

Validity is the degree to which a tool or procedure measures what it is supposed to measure, guaranteeing the accuracy and relevance of the findings, (Sileyew, 2019). In this research, validity was addressed through the following strategies:

- **Construct Validity:** The MapReduce job was designed to reflect real-world poverty profiling logic by integrating income, education, and employment attributes to classify poverty status accurately. This ensured that the computational analysis was aligned with established social science frameworks for poverty measurement.
- **Content Validity:** The synthetic data generated for the simulations included attributes that closely mimic those of real citizens, making the results relevant and applicable to actual government data scenarios.
- **Triangulation:** The use of both qualitative (interviews, questionnaires) and quantitative (Hadoop simulation) methods allowed for cross-verification of findings, further enhancing the validity of the conclusions drawn.

3.7 Data Analysis Technique

Data analysis techniques were employed to analyse the data collected through a combination of both qualitative and quantitative methods, using statistical tools and Hadoop for data processing. The analysis was conducted in two primary phases:

1. Understanding current data management practices in government ministries.
2. Understanding current inefficiencies in the OPM's administration.
3. Implementing a technical simulation to assess the feasibility of a centralised big data framework for data storage and consolidated analysis at the OPM for administration efficiency.

3.7.1 Qualitative data analysis

To gain insight into the current state of data storage and management practices across government ministries, qualitative data was collected through semi-structured interviews and questionnaires administered to key personnel in various ministries. The qualitative data from interviews was analysed using thematic analysis, which involved identifying recurring patterns, key challenges, and shared understandings related to inefficiencies in data handling and access. This method was appropriate because it allowed for a flexible and in-depth exploration of participants' experiences and perspectives, highlighting common themes such as fragmented systems, delayed data access, and a lack of interoperability. Additionally, content analysis was used to systematically categorise open-ended questionnaire responses and quantify the frequency of specific issues raised by respondents. Together, these techniques provided relevant context and insights that informed the rationale for implementing a centralised data framework, such as Hadoop, to address the identified challenges.

3.7.2 Quantitative data analysis

Quantitative data from questionnaires was analysed using descriptive statistical methods, with tools such as Microsoft Excel used to tabulate, group, and visualise the results. Responses were grouped separately for the OPM and other government ministries to highlight unique challenges and benefits experienced by each in their current data management practices. This comparative approach allowed for more precise identification of disparities in administrative efficiency and access to data resources across ministries.

The application of frequency analysis helped identify the most cited challenges, including fragmented systems, data retrieval delays, and reliance on manual processes. These quantitative techniques were suitable for the study, as they provided a clear and measurable understanding of trends across a diverse set of ministries and supported evidence-based conclusions about inefficiencies in the current system.

In addition, simulated quantitative analysis was conducted using a MapReduce job on synthetic datasets representing ministry data. While primarily used to demonstrate system functionality, the results, such as counts of impoverished individuals by region, were also used in a quantitative context to support the study's objectives on integrated analytics and informed policymaking. These results were further visualised using Power BI to illustrate how integrated, real-time data analysis can enhance decision-making and planning at the national level.

3.8 Ethical Considerations

The research complied with the Namibia University of Science and Technology's (NUST) ethical standards, and an ethical clearance certificate was obtained (see Appendix A). A letter of permission was also obtained from the OPM, allowing for the collection of data (see Appendix B).

All participants were informed of the study's purpose and provided consent for their responses to be used. Confidentiality and anonymity were strictly maintained throughout the data collection and analysis process. The simulation used synthetic data; no real citizen information was accessed or compromised.

3.9 Chapter Summary

This chapter outlined the methodological approach used to investigate the feasibility and effectiveness of implementing a centralised Hadoop-based database system for the Namibian Office of the Prime Minister (OPM). A mixed-methods research design was adopted, combining qualitative and quantitative techniques to gain a holistic understanding of the existing data management challenges and the potential benefits of a big data solution.

The chapter also detailed the research activities carried out to fulfil the study's objectives. Below is a summary detailing how each activity aligns with a specific research question and its outcome.

Table 3.1: Summary of Methodology Activities

Research Question	Research Activity	Outputs and outcomes
1. What are the current data storage and management practices across Namibian government Ministries?	Conduct semi-structured interviews Distribute questionnaires to respondents Literature review	Identified challenges and benefits of current data management systems. Baseline understanding of existing data silos.

<p>2. What inefficiencies arise from the lack of centralised data integration at the OPM</p>	<p>Analyse interview and questionnaire data</p> <p>Review OPM's existing infrastructure and workflows</p> <p>Literature Review</p>	<p>Documentation of redundancies, delays, and inconsistencies</p> <p>Identified inefficiencies such as duplication, delays, and lack of coordination.</p>
<p>3. How can a Hadoop-based system enhance data integration and decision-making?</p>	<p>Comparative analysis of similar systems in governance</p> <p>Design a Hadoop-based prototype and store data in HDFS directories to simulate centralised database</p>	<p>Centralised data framework</p> <p>Better accessibility</p> <p>Coordination</p>
<p>4. How effective is a Hadoop-based centralised database system in processing and analysing cross-ministerial data?</p>	<p>Develop and run MapReduce job to analyse poverty across regions</p> <p>Visualise results using Power BI</p>	<p>Production of regional poverty trends</p> <p>Time efficient processing</p> <p>Better planning, coordination and enhanced service delivery</p>

The combination of respondents' perspectives and cross-ministerial data processing offered a holistic view of the problem and the feasibility of the proposed solution. In the next chapter, the results obtained from applying this methodology are presented and discussed in light of the study questions and objectives.

4. HADOOP SETUP

4.1 Introduction

This chapter presents the implementation of a Hadoop-based framework to simulate a centralised database system for the Namibian government. The objective is to demonstrate how Hadoop can support inter-ministerial data integration and analysis by setting up a multi-node Hadoop cluster, uploading a variety of data from simulated ministry datasets into the HDFS, and performing a cross-ministerial analytical job using MapReduce. This implementation showcases the feasibility, scalability, and performance of a big data architecture designed to resolve current inefficiencies in public data management and service delivery.

4.2 Implementation Overview

The open-source framework Hadoop was chosen because it can handle organised, semi-structured, and unstructured data, all of which are prevalent in Namibian government today. The framework is also scalable, fault-tolerant, and can process and analyse large amounts of data. The HDFS was set up and used to create ministry-specific directories, simulating real-world data segregation by government institution. A MapReduce job was executed to demonstrate Hadoop's capabilities in performing analyses that would support cross-ministerial collaboration for administrative efficiency.

4.3 Configuration and Setup

A multi-node Hadoop cluster was set up and configured on an Ubuntu Linux virtual machine (VM) as shown below, to provide an environment where the services can run and jobs be executed.

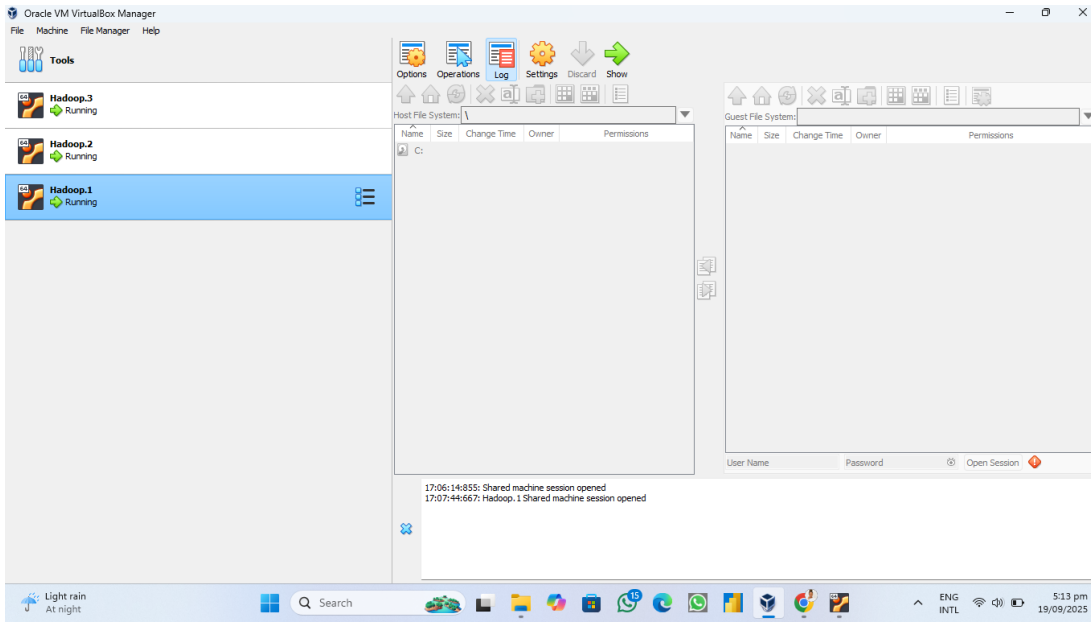


Figure 4.1: Hadoop Cluster Virtual Environment setup

Table 4.1 below provides a summary of the cluster components' configurations. For a detailed, step-by-step guide, refer to Appendix E.

Table 4.1: Summary of Hadoop Multi-node Configurations

Component	Configuration Details
Hadoop version	Hadoop 3.x
Cluster type	Multi-node
Number of nodes	1 master node, 2 datanodes
Operating system	Ubuntu Linux
Java version	Java 8 or later
HDFS setup	Enabled with dedicated directories for Namenode and Datanodes
YARN setup	Enabled for cluster resource management
Replication factor	2 to demonstrate fault tolerance and data availability
Block size	128 MB for efficient large file processing

Data storage paths	NameNode: /usr/local/hadoop/hdfs/namenode DataNodes: /usr/local/hadoop/hdfs/datanode
Core configuration files	core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml
Cluster communication	SSH passwordless communication configured between all nodes
Hadoop services	NameNode, DataNode, ResourceManager, NodeManager
Dataset formats	CSV, PDF, JPG
Processing framework	MapReduce is used for distributed data analysis tasks.

The cluster was configured to mimic a real-world setup where the OPM hosts a central database and individual ministries act as repositories for storing data that they currently have fragmented in various formats. The datasets that were used for this study, were generated with Mockaroo, due to the sensitivity of citizen data, and also the absence of existing standardised datasets at the OPM for experimental purposes.

4.4 Data Preparation

To simulate real-world data scenarios reflective of government operations, CSV files were generated using Mockaroo. These datasets were carefully customised to represent typical records collected by various ministries, such as citizen demographics, education statistics, employment data, and health-related information. Each dataset comprised 1,000 records to ensure consistency and sufficient volume for testing. In addition to the CSV files, a PDF document and a JPG image were downloaded from the MHAISS website. These files were included to demonstrate Hadoop's capability to handle diverse file formats, structured (CSV), semi-structured (PDF), and unstructured (JPG). This approach mirrors the current data environment in Namibia, where government information is collected and stored in various formats, highlighting Hadoop's suitability for building a flexible and inclusive centralised data infrastructure.

4.5 Ministries Directories in HDFS

Once the Hadoop cluster was fully operational and all core services were running, individual directories were created in HDFS for each of the 21 government ministries. HDFS, the primary storage layer in the Hadoop ecosystem, served as the backbone for this setup. Each directory functioned as a standalone repository, storing datasets relevant to its respective ministry. By doing so, Hadoop was effectively used to simulate a centralised database infrastructure, integrating disparate data sources under a unified framework. A few of the Ministries' directories are shown below as they would appear on the Hadoop cluster.

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
drwxr-xr-x	pombili	supergroup	0 B	Apr 23 07:33	0	0 B	M_A_W_L_R	🗑️
drwxr-xr-x	pombili	supergroup	0 B	Apr 23 07:34	0	0 B	M_E_F_T	🗑️
drwxr-xr-x	pombili	supergroup	0 B	Apr 23 07:34	0	0 B	M_F_M_R	🗑️
drwxr-xr-x	pombili	supergroup	0 B	Apr 23 07:36	0	0 B	M_H_A_I_S_S	🗑️
drwxr-xr-x	pombili	supergroup	0 B	Apr 23 07:36	0	0 B	M_H_E_T_I	🗑️

Figure 4.2: Ministries Directories in HDFS

This setup directly addresses the issue of fragmented data silos currently prevalent across Namibian government ministries. It enables cross-ministerial access and analysis of information while maintaining the sovereignty of each ministry's data. Furthermore, Hadoop's support for various data formats as shown in Figure 4.2, includes but is not limited to, structured CSVs, semi-structured PDFs, and unstructured JPGs. This is to ensure that historic data, which already exists in multiple formats, is preserved and remains relevant. This compatibility with heterogeneous data types not only safeguards historical records but also facilitates their integration with newer datasets. As a result, it ultimately enhances administrative efficiency and decision-making capacity at the OPM that is inclusive of every Namibian citizen's record, from older records to the latest.

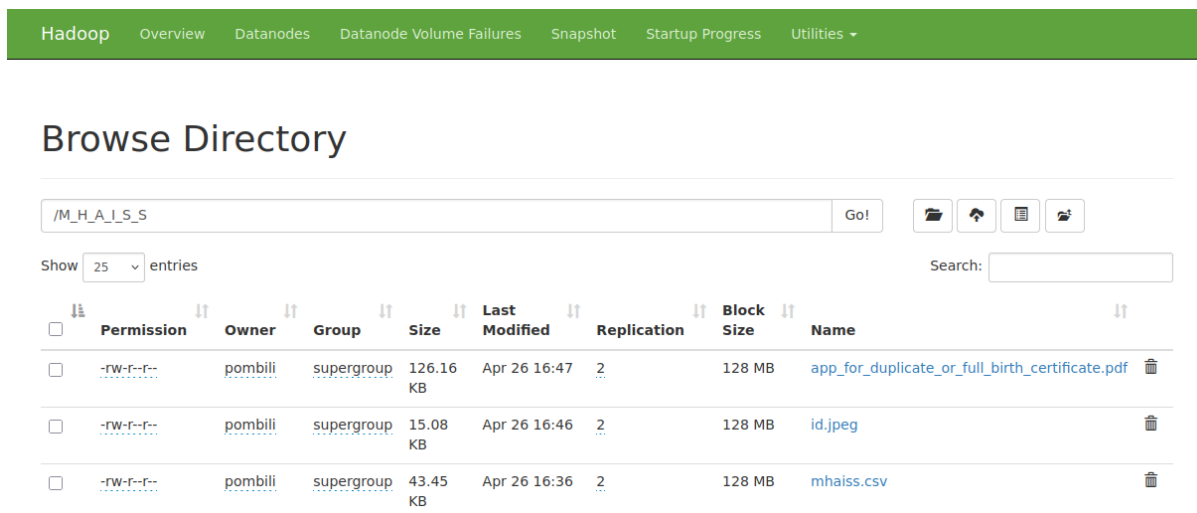


Figure 4.3: Structured and Unstructured Data Formats in HDFS

4.6 MapReduce

MapReduce, the processing engine within the Hadoop framework, was employed to run a cross-ministerial data analysis, simulating how a centralised framework could streamline data-driven decision-making. The analysis was completed in under three minutes, demonstrating the efficiency and scalability of Hadoop in handling large, distributed datasets. This is not the case with the current processes at the OPM, where collecting cross-ministerial data involves manual communication between departments, often taking several days to compile and verify information.

Such delays can severely impact government responsiveness, particularly in situations that require urgent intervention, such as responding to public health crises, allocating emergency resources during natural disasters, or tracking sudden spikes in unemployment. Additionally, manual data handling increases the likelihood of human error, duplication, and inconsistencies, all of which can compromise the accuracy and reliability of national reporting and planning.

MapReduce, by contrast, automates the extraction, transformation, and aggregation of data across ministries in a standardised and error-minimised process (Apache Hadoop, 2025c). It will enable the OPM to gain timely insights, reduce administrative overhead, and support real-time policymaking. This not only enhances internal efficiency but also significantly improves the quality and speed of public service delivery, supporting Namibia's broader goals for digital governance and improved service delivery.

4.6.1 MapReduce Job Structure

A MapReduce job was developed to perform a poverty-level analysis by joining datasets on the *national_id* field. The MapReduce job was compiled using Java code in Eclipse. The key MapReduce classes developed were: Mapper Class, Reducer Class, and Driver Class.

- a) **Mapper Class:** The PovertyMapper.java class is responsible for parsing input records from different CSV files representing ministry-specific datasets, (Apache Hadoop, 2025c). Each line of data was tagged based on its source (e.g., “mheti” for education, “mol” for labour, and “mhaiss” for home affairs) and transformed into a key-value pair using *national_id* as the key.
- b) **Reducer Class:** The PovertyReducer.java class received all key-value pairs grouped by *national_id*, (Apache Hadoop, 2025). It merged data from different ministries to form a unified profile per individual and applied a decision rule to classify poverty status. An individual was classified as “Below Poverty Line” if they met any of the following combined criteria: attained primary education or lower, or dropped out of secondary/tertiary education; were unemployed, temporarily employed, or self-employed with low income. A valid citizen profile was recorded and aggregated by region value.
- c) **Driver Class:** The PovertyDriver.java class acted as the driver of the MapReduce job. It configured the input and output paths for each ministry dataset, registered the Mapper and Reducer classes, and defined key output parameters, including the input format and data types, (Apache Hadoop, 2025c). It ensured that all datasets were processed in a single job, producing an output file stored in the HDFS for further analysis.

4.6.2 Compilation and Packaging

The code was compiled and packaged into a JAR file, cleaned and built in Eclipse, exported as a runnable file named *PovertyAnalysis.jar* with main class *org.poverty.analysis.PovertyAnalysisDriver*, and then saved in home/pombili virtual environment.

4.6.3 Running the Job

The MapReduce Job was run using the command:

```
hadoop jar /home/pombili/PovertyAnalysis.jar org.poverty.analysis.PovertyAnalysisDriver /input/mhaiss.csv /input/mheti.csv /input/mol.csv /output/poverty_analysis,
```

as displayed in the figure below.

```
pombili@master: $ hadoop jar /home/pombili/PovertyAnalysis.jar org.poverty.analysis.PovertyAnalysisDriver /input/mhaiss.csv /input/m
heti.csv /input/mol.csv /output/poverty_analysis
2025-05-03 11:57:27,946 INFO client.DefaultNoHARMFaloverProxyProvider: Connecting to ResourceManager at master/192.168.0.11:8050
2025-05-03 11:57:31,296 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool int
erface and execute your application with ToolRunner to remedy this.
2025-05-03 11:57:31,557 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/pombili/.sta
ging/job_1746252820347_0002
2025-05-03 11:57:32,894 INFO input.FileInputFormat: Total input files to process : 1
2025-05-03 11:57:33,151 INFO input.FileInputFormat: Total input files to process : 1
2025-05-03 11:57:33,198 INFO input.FileInputFormat: Total input files to process : 1
2025-05-03 11:57:33,541 INFO mapreduce.JobSubmitter: number of splits:3
2025-05-03 11:57:34,180 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1746252820347_0002
2025-05-03 11:57:34,180 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-03 11:57:35,449 INFO conf.Configuration: resource-types.xml not found
2025-05-03 11:57:35,449 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-05-03 11:57:37,879 INFO impl.YarnClientImpl: Submitted application application_1746252820347_0002
2025-05-03 11:57:38,478 INFO mapreduce.Job: The url to track the job: http://master:8088/proxy/application_1746252820347_0002/
2025-05-03 11:57:38,481 INFO mapreduce.Job: Running job: job_1746252820347_0002
2025-05-03 11:58:08,194 INFO mapreduce.Job: Job job_1746252820347_0002 running in uber mode : false
2025-05-03 11:58:08,591 INFO mapreduce.Job: map 0% reduce 0%
2025-05-03 11:59:12,118 INFO mapreduce.Job: map 33% reduce 0%
2025-05-03 11:59:13,160 INFO mapreduce.Job: map 67% reduce 0%
2025-05-03 11:59:43,787 INFO mapreduce.Job: map 100% reduce 0%
2025-05-03 11:59:45,847 INFO mapreduce.Job: map 100% reduce 100%
2025-05-03 11:59:46,884 INFO mapreduce.Job: Job job_1746252820347_0002 completed successfully
2025-05-03 11:59:48,621 INFO mapreduce.Job: Counters: 55
```

Figure 4.4: Running MapReduce Job in Hadoop

The job’s status was tracked on the Hadoop application page, available at <http://localhost:8088>. After the successful completion of the job, the results were automatically stored in `/output/poverty_analysis/part-r-00000`, listing: each `national_id` of individuals matching the poverty criteria, shown below.

```
pombili@master: $ hdfs dfs -cat /output/poverty_analysis_per_region/part-r-00000
Omusati Oshana At Risk of Poverty
At Risk of Poverty
Erongo At Risk of Poverty
Omusati Oshana At Risk of Poverty
Caprivi At Risk of Poverty
Otjikoto At Risk of Poverty
Kavango West At Risk of Poverty
At Risk of Poverty
Khomas At Risk of Poverty
Omaheke At Risk of Poverty
Otjikoto At Risk of Poverty
Kunene At Risk of Poverty
Kavango West At Risk of Poverty
Omaheke At Risk of Poverty
Otjikoto At Risk of Poverty
At Risk of Poverty
Erongo At Risk of Poverty
Omaheke At Risk of Poverty
Hardap At Risk of Poverty
Otjikoto At Risk of Poverty
Otjikoto At Risk of Poverty
Otjozondjupa At Risk of Poverty
Kavango West At Risk of Poverty
Otjozondjupa At Risk of Poverty
Ohangwena At Risk of Poverty
Kavango West At Risk of Poverty
Erongo At Risk of Poverty
Hardap At Risk of Poverty
Kavango West At Risk of Poverty
Omusati Oshana At Risk of Poverty
Hardap At Risk of Poverty
```

Figure 4.5: MapReduce Analysis output

This setup confirms that Hadoop is well-suited to support a centralised database framework for all government ministries, enabling consolidated data analysis and providing a comprehensive overview for administrative decision-making. Its scalability, through parallel processing across datanodes, ensures that as data volumes or analytical demands grow, the system can easily

expand by adding more datanodes to support the processing demand. With its robust architecture and analytical capabilities, Hadoop offers the OPM a practical and powerful tool to enhance administrative efficiency and leverage big data for improved, citizen-centric service delivery.

4.7 Visualising the Centralised Hadoop Setup

A Hadoop cluster is a group of computers called nodes that act as a single, centralised system working on the same task (Coursera, 2025). The Hadoop cluster in this study is made up of 2 Slave Nodes and one Master Node. It is possible to add more nodes to a cluster, which allows HDFS and MapReduce to provide a flexible setup with more computational power. Each node operates independently, with its own memory and storage (Simuni & Amaranatha, 2024). As displayed by the legend below **Figure 4.5**, each node has its own storage layer (HDFS), in grey, management layer (YARN), in yellow, and processing layer (MapReduce), in green. It does not share hardware resources with other machines, except for being connected through a common network.

The Master Node manages data distribution, allocates resources, and oversees parallel processing tasks (Hannan, 2016). The Slave Nodes execute the tasks assigned to them by the Master Node. The Government Ministries would then be responsible for loading data into the cluster and the OPM requesting and retrieving processed results. They act as an interface between the Hadoop cluster and external systems or applications, whereas the OPM would then be the host of the system.

The Hadoop cluster setup of the centralised data framework proposed in this study is illustrated in *Figure 4.6* below.

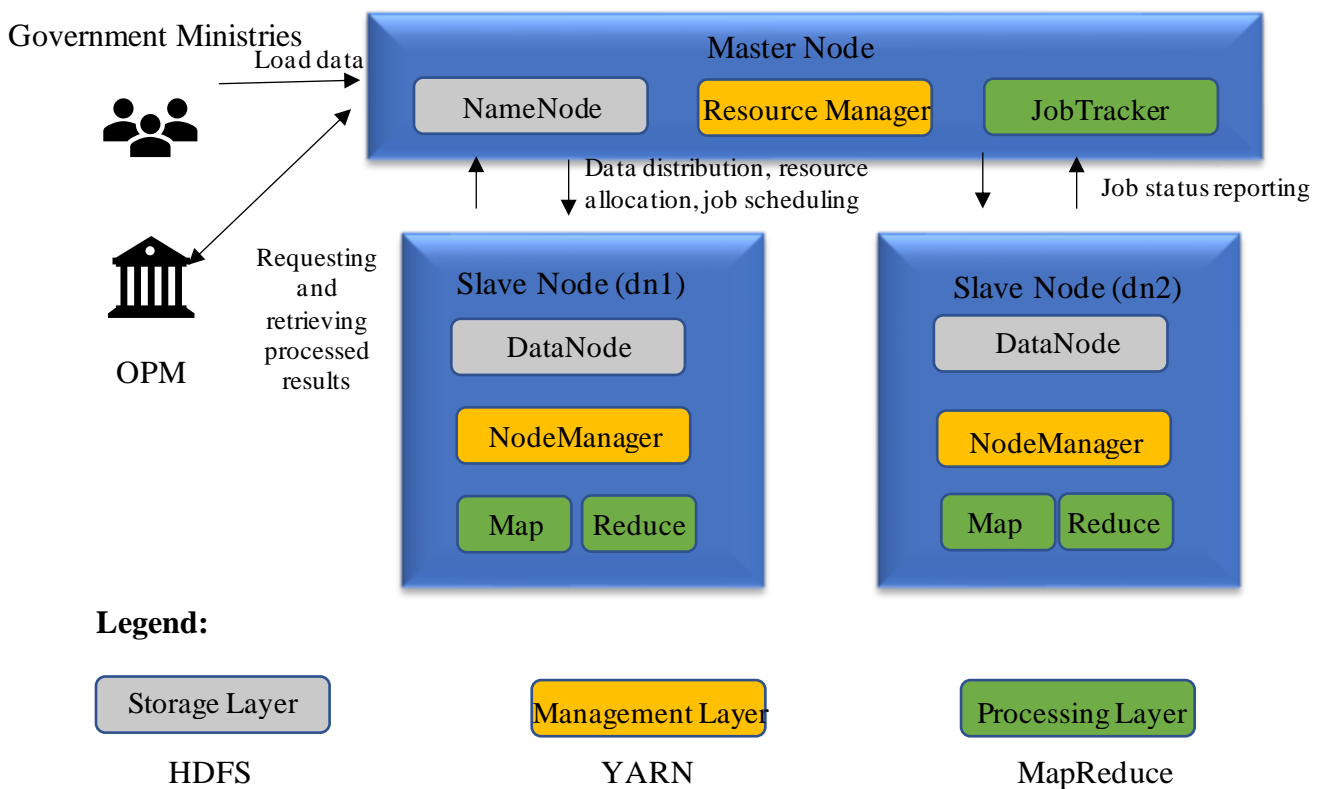


Figure 4.6: Hadoop Setup Visualisation

a) Storage Layer

Each node within the configured Hadoop cluster operates independently, with its own dedicated storage, memory, processing power, and network bandwidth (Greeshma & Pradeepini, 2016). When datasets from various ministries are uploaded, Hadoop automatically splits the data into manageable blocks and distributes them across the cluster using HDFS. To address potential hardware failures common in government IT infrastructure, HDFS is designed to replicate each data block three times across different nodes. This ensures data reliability and availability. The official site of Apache Hadoop, (2025a) outlines that the NameNode (master node) is responsible for managing the metadata of these blocks and their replicas, enabling efficient tracking, access, and coordination of ministry-specific data within the centralised framework.

b) Management layer

Effective coordination of cluster nodes is critical to ensuring that multiple applications and users can access and share resources seamlessly. Initially, Hadoop MapReduce was responsible

for both data processing and resource management. However, with the introduction of YARN, these functions are now separated, improving efficiency and scalability.

YARN acts as the central resource manager in the Hadoop ecosystem, assigning system resources across various data processing frameworks, not just MapReduce, this includes tools like Apache Pig, Hive, Giraph, and Zookeeper, (Apache Hadoop, 2025a). By enabling multiple analytical tools to run concurrently on the same cluster, YARN enhances Hadoop's adaptability to diverse workload demands (Apache Hadoop, 2025a). This makes it an ideal solution for Namibia's cross-ministerial data environment, where various analytical needs must be supported on a shared infrastructure without compromising performance or reliability.

c) Processing Layer

The processing layer is essential to turning raw information into actionable insights. This layer consists of computational frameworks that analyse both structured and unstructured datasets as they enter the cluster. As outlined in the MapReduce tutorial, Apache Hadoop, (2025c), data is subjected to operations such as mapping, splitting, sorting, merging, and reducing, breaking it down into smaller, manageable blocks for efficient processing, as illustrated in *Figure 4.7* below.

These operations are distributed across multiple nodes within the cluster, positioned close to the storage location of the data to reduce latency and optimise performance (Apache Hadoop, 2025c). In addition to the traditional MapReduce engine, modern computation frameworks such as Apache Spark, Storm, and Tez enhance the capabilities of the processing layer. These tools support real-time analytics, interactive querying, and flexible programming models, making them especially suitable for the dynamic and cross-ministerial data analysis needed for timely and informed decision-making in public service delivery.

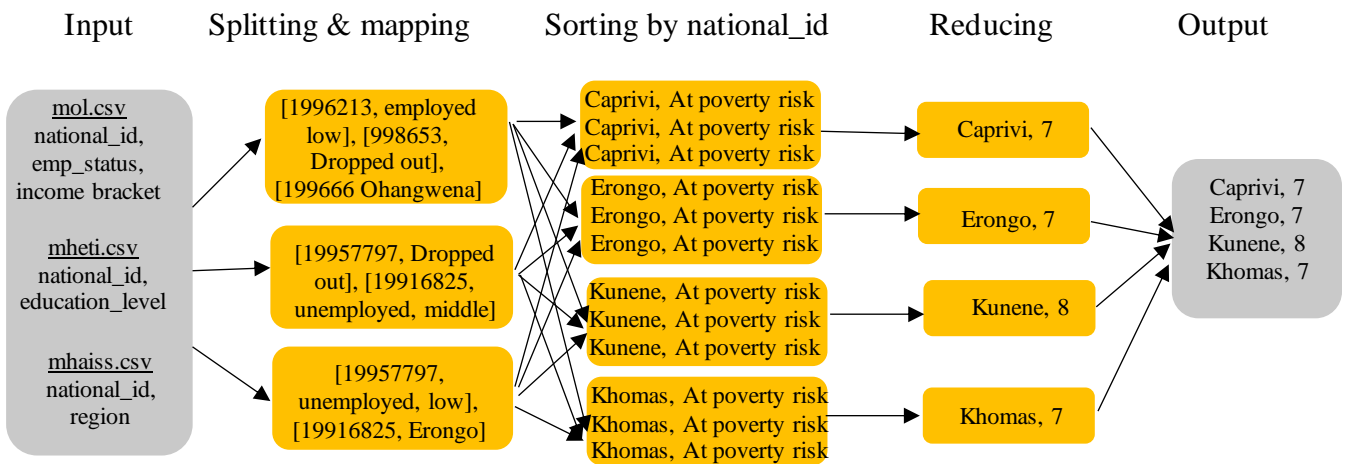


Figure 4.7: MapReduce Job Processing

In the centralised Hadoop framework, the master node, equipped with a JobTracker, acts as the central coordinator for processing tasks. When a client node at the OPM submits a job, the JobTracker consults the NameNode in HDFS to determine where the required data resides across the cluster.

Once the data location is identified, the JobTracker assigns the processing tasks to the closest available slave nodes, which are typically TaskTrackers located on the same servers as the DataNodes that store the target data. This proximity ensures data-locality, reducing network load and improving processing speed.

Each TaskTracker then carries out the map and reduce operations as instructed by the JobTracker. It also continuously reports its progress back to the master node. If a TaskTracker fails during execution, the JobTracker detects the failure and automatically reassigns the task to another nearby slave node, ensuring fault tolerance and uninterrupted data processing. This distributed and resilient design is crucial for efficiently handling large-scale, cross-ministerial analytics in a centralised system.

In the current administration, cross-ministerial data retrieval is a slow, manual process, often requiring days to request, receive, and reconcile data from different ministries. This prevents effective planning, rapid service delivery, and emergency response. The proposed Hadoop-based system, with YARN managing resources and MapReduce handling distributed computation, can perform the same operations in minutes, as demonstrated in the simulation job, which processed multiple CSV datasets across ministries in under three minutes.

This framework not only unifies fragmented data silos but also eliminates bottlenecks by automating data retrieval, supporting various file types, and enabling real-time analysis. With ministries submitting data to centralised HDFS directories and YARN ensuring efficient resource use, this system lays the foundation for a data-driven, responsive, and citizen-centric governance model in Namibia.

4.8 Chapter Summary

This chapter outlined the setup of a centralised Hadoop cluster hosted at the OPM, configured with one master and two slave nodes. Ministry-specific directories were created in HDFS, simulating real government data using CSV, PDF, and JPG files to reflect Namibia's diverse data landscape. This demonstrated Hadoop's ability to consolidate fragmented data silos under a consolidated system.

The use of MapReduce enabled efficient cross-ministerial analysis, completing tasks in minutes compared to the current manual, time-consuming processes. The visual architecture illustrated how government data flows into the centralised system, and into MapReduce Layer for data analysis. Overall, the chapter confirmed Hadoop as an effective framework that addresses the numerous issues posed by big data due to its ability to distribute processing and storage capacity among thousands of cluster nodes. It also demonstrated its capabilities in enhancing data accessibility, analysis, and administrative efficiency across ministries. Hadoop is a cost-efficient, open-source software, that is scalable and fault-tolerant; which makes it an ideal framework to support the diverse and versatile Namibian data ecosystem.

5. RESULTS AND EVALUATION

5.1 Introduction

This chapter evaluates the research outcomes against its objectives, using data collected through a questionnaire administered online via JotForm, interviews, and comparative analyses. The results are aligned with the research questions, focusing on data redundancy, delays in decision-making, data inconsistencies, and security risks.

The chapter also assesses the feasibility of the proposed Hadoop framework, addressing the shortcomings that the OPM currently faces in its administrative processes as a result of government ministries' data management practices.

5.2 Data analysis overview

The data analysis in this study employed both qualitative and quantitative approaches to assess the feasibility and effectiveness of a centralised Hadoop-based database framework for the OPM. Data was collected through questionnaires and interviews with officials across various government ministries to understand the current data management practices, challenges in inter-ministerial coordination, and attitudes toward centralised database systems. The responses revealed significant fragmentation, reliance on manual processes, and delayed service delivery due to siloed data.

To validate the proposed centralised database framework, a Hadoop cluster with three data nodes was configured, and simulated datasets were used for three ministries, MHAISS, MoL, and the MEAC. Each dataset contained 1,000 records, with fields such as *national_id*, *region*, *employment_status*, *education_level*, and *household_income*, reflecting the types of data that would typically exist in these ministries. The aim was to assess poverty distribution across Namibia's regions by joining these datasets on the *national_id* field and applying logical conditions, such as low income, low education level, and unemployment, to identify vulnerable populations. A MapReduce job was executed to perform this cross-ministerial analysis efficiently. The results provided region-specific insights into poverty, which could be used to inform planning and targeted interventions, such as drought relief programs, currently coordinated by the OPM. This demonstrates Hadoop's potential to support real-time, data-driven policymaking across government sectors.

5.2.1 Additional benefits of Hadoop

Hadoop's architecture is designed to support a cluster of multiple computers, even when those machines have different computational capacities, to strengthen the overall computational power of the cluster. This flexibility is one of Hadoop's major strengths. In a traditional setup, similar to the current government ministries setup, where a single computer processes large datasets, performance is constrained by the limits of that one machine's CPU, memory, and storage. However, in a Hadoop cluster, tasks are distributed across several nodes, allowing each machine, regardless of its individual performance, to contribute to the processing workload. This means that even older or less powerful machines can be utilized effectively alongside high-performance nodes. Hadoop's fault tolerance and data replication mechanisms ensure that failures on less performant nodes do not disrupt the overall process, while its YARN resource manager allocates jobs based on each node's capacity. As a result, this cluster setup not only maximises resource utilisation, but also reduces processing time, increases scalability, and minimises the cost of hardware upgrades, all of which make it ideal for public sector institutions where infrastructure may be heterogeneous and budgets limited.

5.3 Results of the study

5.3.1 Findings on Current Data Management in Namibian Ministries

(RQ1: What are the current data storage and management practices across Namibian government ministries?)

The research findings indicate that government ministries manage data independently, utilising various storage formats, including spreadsheets, proprietary databases, HTML, and physical books and documents. The Directorate of National Examinations and Assessments (DNEA) within the Ministry of Education, Arts, and Culture is responsible for administering national exams in schools. An interview with two administrative officers revealed that each collects data separately, one using Excel sheets and the other a physical book, which means that there is no standardised method of data capturing. In terms of their job duties, one of the administrative officers stated that they need to interact with at least four other ministries to carry out their daily tasks. To acquire the necessary information, they use either a telephone or face-to-face communication, which can take weeks to deliver services. The research also found that there is no backup of the information they record in books, and the only way to keep it safe is to store it in a secure place.

A senior education officer within the same department stated that they use the Education Management Information System (EMIS) to collect candidate information from all schools,

using an online registration form. However, there are challenges such as slow internet connectivity and outdated systems. They use the PDF format to share information across ministries, and it takes about a week to attend to citizens' requests. Regarding measures for protecting sensitive information, internal policies are in place, including the use of passwords and backups, which are maintained at the OPM. According to an accountant from the same department, they used to store information in HTML format, which is captured manually, and they have one database which is integrated with other systems. They need to at least interact with two other ministries to perform their daily tasks, which they request in the form of submission letters, which are in Word format.

An IT specialist from the Ministry of Labour and Public relations stated that their ministry has a Namibia Integrated Employment Information System (NIEIS) in place, that captures names, addresses and qualification of job seekers, with a unique reference number for each applicant that is solely used on their system – meaning that it would be invalid elsewhere. According to a Senior Fleet Officer from the Finance and Administration department, they use Excel to store information, and emails to request information, which is shared in PDF format. They also have two databases within their department which are not integrated, and have to interact with many ministries to get their work done on a daily basis. In terms of data security, they have their server with a backup kept at the OPM.

An Analyst Programmer from the Ministry of Agriculture, Fisheries and Water and Land Reform, IT department, stated that they use digital forms to capture information, which is saved in PDF format, but shared in Word via email. They interact with several other ministries to complete their daily tasks, and it takes them days to respond to information requests within the ministries as well as citizens' requests.

The research found that citizens' records exist in more than one system, for instance, birth registration is done at the Ministry of Health and Social Services and the Ministry of Home Affairs and Immigration; death registration is done at the Ministry of Health and Social Services, the Ministry of Safety and Security, and the Ministry of Home Affairs and Immigration. At the same time, divorce registration is handled by the Ministry of Justice, while the Ministry of Justice and the Churches conduct marriage registration. Most of the ministries' data backups are kept at the OPM.

The World Bank Group (2016), analyses report, highlights that, six databases are maintained by the Ministry of Home Affairs and Immigration (MHAIS): the National Population Register System (NPRS), the Automated Fingerprint Identification Systems (AFIS), the IMAGO

database, the border control database, and a database containing 4.3 million PDF records and amendments concerning births that took place before 2012. From February 2016, MHAI incorporated a citizenship database running on Oracle, with border control and NPRS databases backed up at the OPM. Further, they procured a new blade server to address inadequate server capacity for civil registration and identification.

A study conducted by the (Namibia Statistics Agency, 2023), estimated that 45.5% of the sectors lack an Integrated Management System (IMS), have insufficient databases, as well as limited or no internet connectivity.

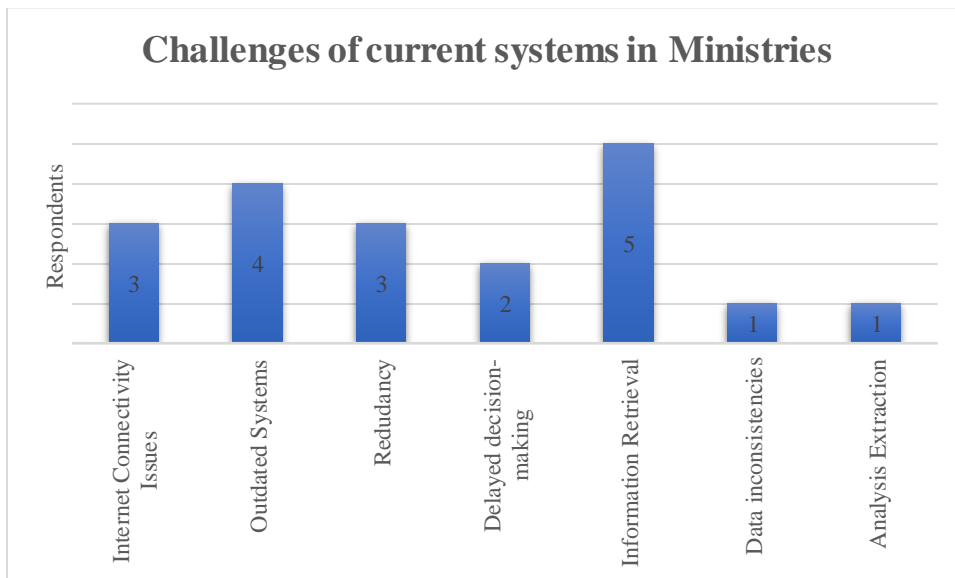


Figure 5.1: Challenges of Current Systems in Ministries

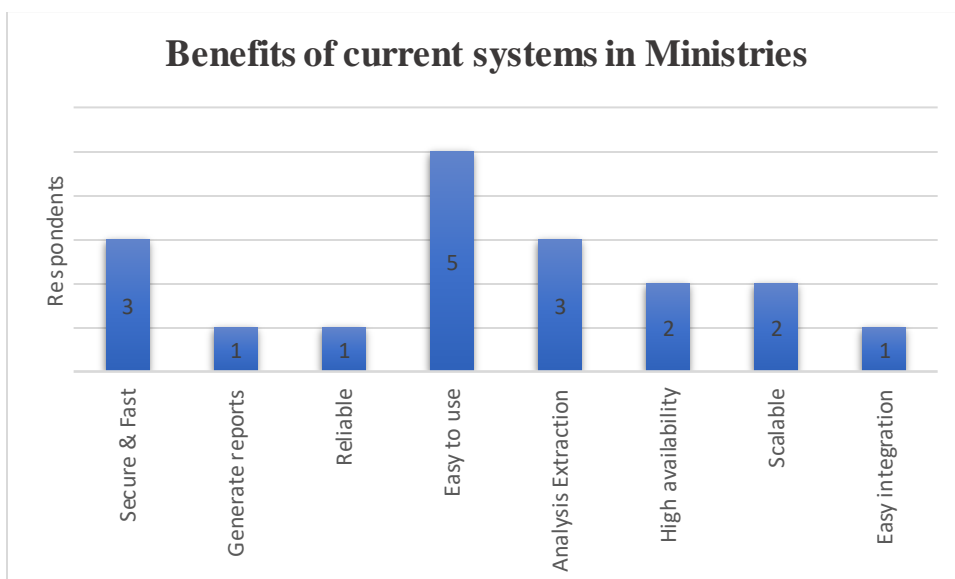


Figure 5.2: Benefits of Current Systems in Ministries

The findings show that, respondents' main challenge with current government systems is information retrieval. This can be addressed by Hadoop HDFS, which allows structured, semi-structured, and unstructured data from all ministries to be stored in a centralised system, with each ministry having its accessible repository. Data requests will no longer be manual, as all records will be digitised and accessible from the OPM system.

According to the respondents, the major benefits of the current systems are ease of use and analysis extraction, which can be attributed to the fact that they have been in use for a long time and the staff have become proficient in using them. However, this does not necessarily mean they are efficient. Hadoop will, however, require upskilling and staff training to ensure they are equipped to operate, manage, and extract meaningful insights from the system effectively, bridging the gap between current familiarity with legacy systems and the increasing demands of a modern, scalable data infrastructure.

5.3.2 Findings on Inefficiencies within the OPM

(RQ2: What inefficiencies arise from the lack of centralised data integration at the OPM?)

As stated in the OPM's Customer Service Charter, they have approximately ten different departments, with additional divisions and sub-divisions as needed. These departments each have a role to play in administering and coordinating the work of the Cabinet. They offer services to all Government Offices, Ministries, and Agencies (OMAs), Regional Councils, and Local Authorities (RCs & LAs), as well as the General Public, OPM staff, and other relevant entities, resulting in a vast customer base.

In respect of the above, the study identified the following inefficiencies as a result of the absence of a centralised database within the OPM:

a) Redundancy

The OPM maintains various databases for each department to store information, and these databases are not interconnected. They act in isolation and keep duplicate information concerning governments and citizens that already exists in every other department separately, and the same information somehow sits in another ministry's database. According to the questionnaire, some of the databases listed are kept are NPRS, ISAS, DRMCOMIS, EDRMS, IFMS, HRMS, e-Service, e-Governance, SQL Database, MariaDB, CBMIS, Biometric

System, and e-Declaration. Some Ministries also have their backups kept at the OPM, on their servers for security purposes.

According to the customer service charter, when a citizen needs services from the OPM, they are required to provide their full name, postal address, email, and telephone number, which is information that should already exist within the government systems. It should be easy to retrieve that with the use of an identification number or a citizen's mobile number, as it would be accessible from a centralised system. Most government offices also maintain attendance books of all customer visits, which, by the end of the year, would be difficult to track. This allows for better decision-making and resource allocation by identifying the most visited days and months, as well as the services that are most highly requested.

These poor data management practices create silos of overlapping and redundant information, that becomes very costly to store with the increasing volumes of data captured and also not being leveraged to improve customer service. In an era of advancing technology, there is an increasing demand for big data warehouses to enable governments to make smarter decisions and deliver up-to-standard customer service. These traditional databases will not help the OPM keep up with the growing data needs and leverage them to optimise their administrative efficiency; hence, there is a need for a better solution to centralise all these databases into one unified system. Hadoop is one of the latest technologies that is free, scalable, and supports big data analytics, providing a holistic overview of the national state of affairs to inform decision-making.

The respondents have outlined some of the challenges and benefits of the current systems in the OPM, as displayed by the figures below.

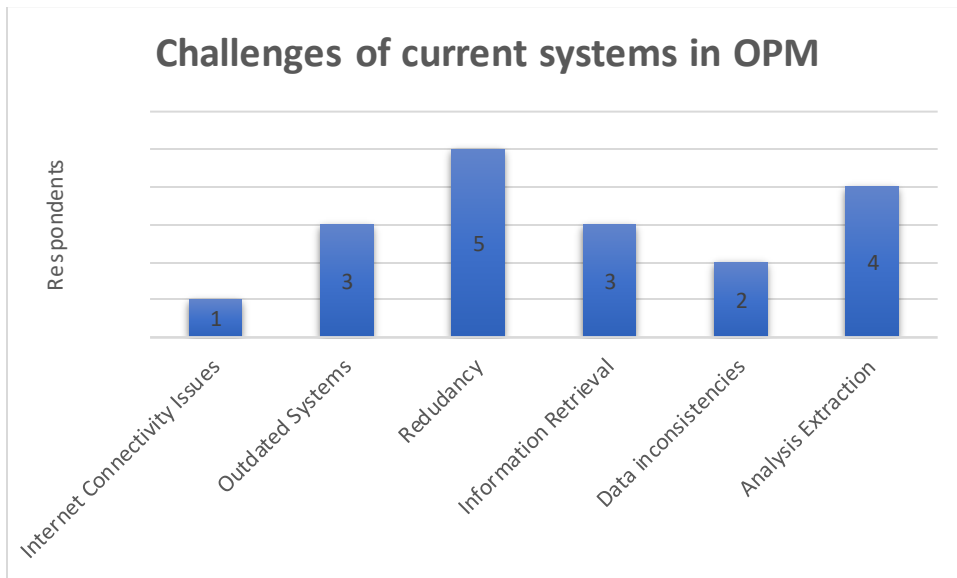


Figure 5.3: Challenges of Current Systems in the OPM

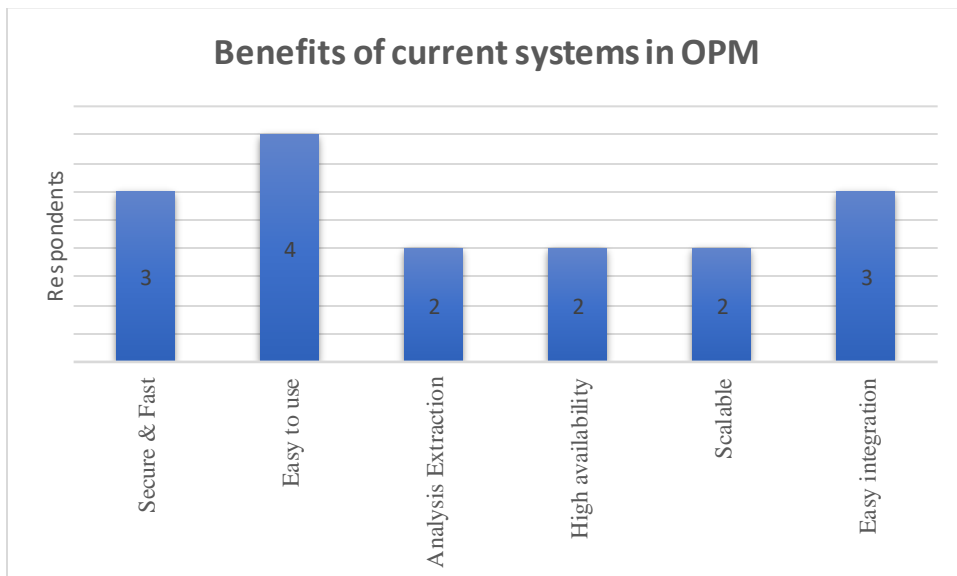


Figure 5.4: Benefits of Current Systems in the OPM

It is evident from the responses that the primary concern with the current systems is data redundancy, and the majority of the OPM staff find them easy to use, citing this as a significant benefit, as they have operated these systems for a while and have become proficient in them.

b) Delayed Decision-Making

According to the respondents, more than 50% stated that they take days to respond to customers' queries. 100% of the respondents indicated that they interact with all OMAs in their daily operations to execute their tasks, with more than 80% using email to request the necessary information.

Decision-making at OPM is significantly slowed due to the manual retrieval and processing of data from multiple ministries. Without direct access to a centralised government database, OPM officials must request reports, wait for responses, and manually compile the information, which delays critical interventions.

For example, during drought relief planning, the OPM relies on multiple reports from different ministries, which take weeks to compile, delaying aid distribution. Additionally, in cases where a patient is unconscious and transferred from one state hospital to another without their physical record, the doctor would not have access to the patient's medical history. Davies et al., (2020), highlighted that Namibia's lack of adaptation to integrate services amongst ministries results in poor service delivery without sufficient coordination across government agencies and across governance scales (from the national and regional to the local).

c) **Data Inconsistencies**

According to the respondents, information is stored in various formats, including PDF, Word, HTML, Excel, PowerPoint, and Notepad. These variations in the information captured in each department make it challenging to extract meaningful analysis from such data, and most databases are not even compatible with some of these formats. To achieve a realistic overview of the current status of affairs, the OPM needs to consider a system that can handle all data formats currently in use across all departments and government ministries. This is crucial in extracting meaningful analysis from historical data that will inform future trends, because every piece of information counts, regardless of its format, to make accurate and well-informed decisions.

Hadoop is primarily designed for handling structured and semi-structured data, but it can also store unstructured data, such as PDFs, Word documents, and images (JPGs). Data inconsistencies due to different data formats in Hadoop are handled through a combination of its ecosystem tools and user-defined processes, extracted, normalised and processed. As a result, no data formats will be discarded due to incomparability issues.

d) Security Risks

According to the responses from the questionnaire, information requests are done via email, applications, or in person, in PDF, Excel or Hard copy formats, across departments and ministries in the proportion as shown below:

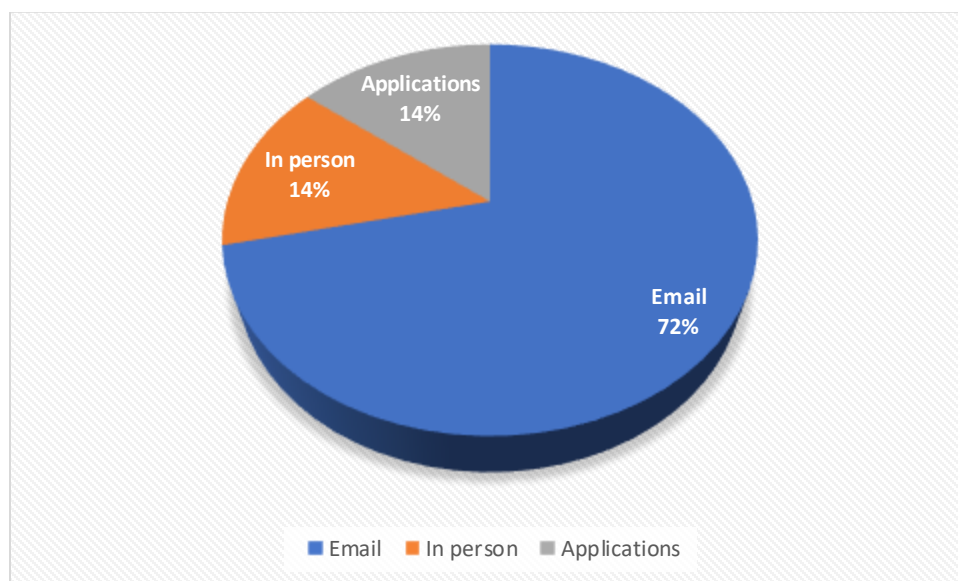


Figure 5.5: Channels of Communication

Sharing sensitive information over email or in person can be very risky, especially when handling citizens' sensitive information. The risks are such that attachments may contain viruses, there is little control over forwarding or unauthorised sharing, files can be tampered with, and no audit trail or encryption. However, role-based access control is ensured over files stored in Hadoop, and an audit log is available that tracks who accessed what, when, and what actions were performed. This will enable, for instance, finance departments to stop requesting citizens' income info via email (PDFs). Instead, a secure web form will be used to feed data into Hadoop via Apache NiFi, with access to that data encrypted, role-restricted, and fully auditable.

5.3.3 Benefits of a centralised database with Hadoop for administration

(RQ3: How can a Hadoop-based system enhance data integration and decision-making?)

The research findings reveal that the adoption of a Hadoop-based centralised system can significantly enhance the OPM’s ability to integrate data across ministries, enabling more efficient and informed decision-making. Below is a structured discussion of each major administrative challenge, along with a proposed solution based on Hadoop.

i) Fragmented and Siloed Data Across Ministries

All government ministries operate as separate entities, building up stacks of data silos, with independent data storage practices and minimal interoperability.

HDFS allows for the integration of datasets from various ministries into a single, centralised system. In the simulation, ministry datasets were stored in HDFS directories and accessed uniformly via MapReduce jobs. This eliminated the need for manual data consolidation, creating a unified database that promotes cross-sectoral governance.

ii) Manual and Time-Consuming Data Processing

Data collecting and reporting within ministries is primarily manual, and inter-ministerial queries often require more time and human intervention.

MapReduce enables efficient processing of large datasets. In the case of the simulated poverty analysis job, the system processed over 3,000 records (from 3 different ministries) in just 2 minutes and 57 seconds, as displayed by the *Figure* below.

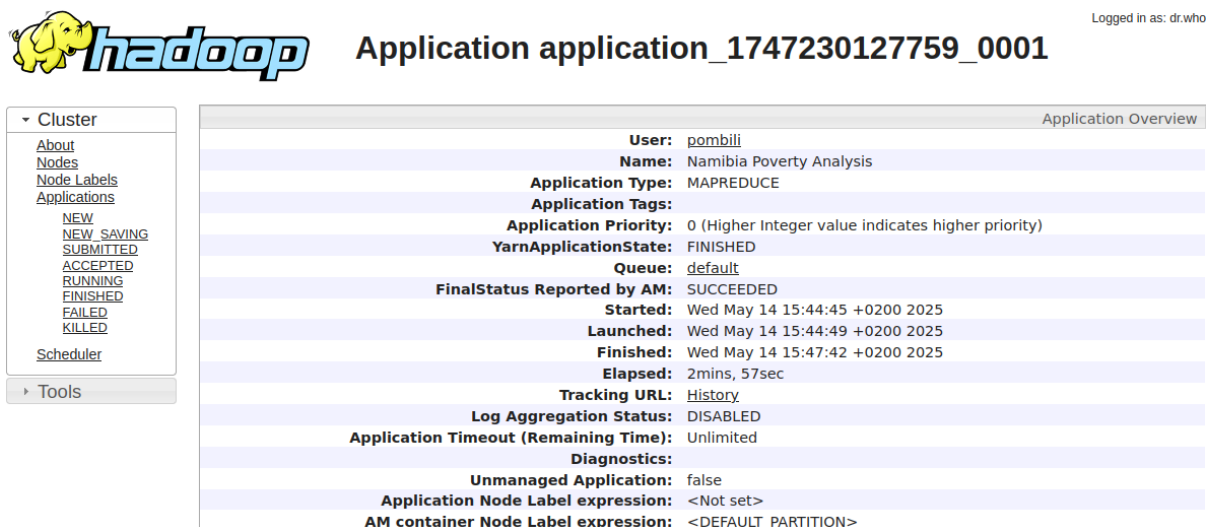


Figure 5.6: MapReduce Job Processing Time

This demonstrates that Hadoop can facilitate the automation of repetitive analytical tasks, thereby reducing the delays associated with decision-making. In almost real time, administrative tasks that once took weeks to complete can now be completed.

iii) Inconsistent Data Standards

Ministries employ various formats and software systems to store their data, leading to incompatibility and inadequate data sharing.

Mockaroo-generated CSVs were standardised using a common identifier (*national_id*) and consistent attribute formatting. This standardisation allowed the Mapper class to emit key-value pairs that were later joined in the Reducer stage. The exercise demonstrated that it is possible to create a standardised data schema across ministries and that Hadoop provides the necessary infrastructure to support and implement these standards at scale.

iv) Limited Insight into Multi-Dimensional Problems (e.g., Poverty)

The OPM lacks an integrated mechanism for evaluating complex socio-economic indicators that require data from multiple ministries.

The research implemented a cross-ministerial poverty analysis that combined data on education, employment, and citizenship. The logic encoded in the Reducer file classified individuals by poverty status based on their educational attainment, employment status, and citizen records. The outputs, when visualised in Power BI, provided the OPM with helpful information, including the ability to identify areas with high poverty numbers, enabling data-driven resource allocation and policy creation.

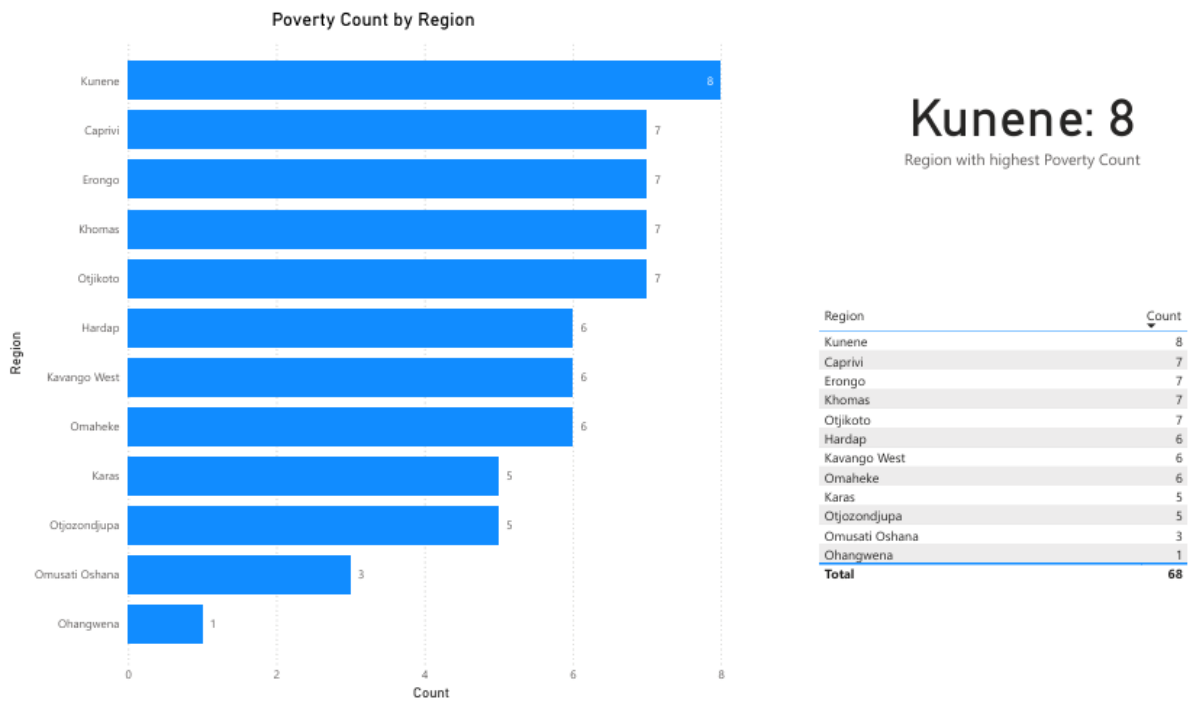


Figure 5.7: Regional Poverty Analysis Visualisation

As shown in the **Figure** above, the analysis indicates that Kunene has the highest poverty rate.

v) Redundant and Duplicated Data Storage

Ministries often collect and store overlapping data independently, resulting in redundancy and inflated data management costs.

By establishing a single repository in HDFS, data was stored once and accessed by all processing jobs. This reduces duplication, streamlines updates, and simplifies version control. A centralised system enforces consistency, ensures data accuracy, and lowers the cost of maintaining multiple disjointed systems.

5.4 Implementation of a Hadoop-based centralised system

(RQ4: How effective is a Hadoop-based centralised database system in processing and analysing cross-ministerial data?)

5.4.1 Evaluation of the proposed system

The proposed system was evaluated against the following criteria and the responses received from the questionnaire:

Table 5.1: Proposed Systems Evaluation

Evaluation Metric	Current Systems	Proposed Centralised System
Adequacy	Data stored in silos, making integration difficult.	Centralised storage enables seamless access to all government data.
Coverage	Only individual ministries have access to their data.	All ministries can access and contribute to a unified database.
Efficiency	Data requests take weeks or months to process.	Real-time data retrieval and analytics enable faster, more informed decision-making.
Productiveness	Data redundancy leads to high storage costs and repetitive tasks.	Eliminates duplicate datasets, optimising resource allocation.
User Friendliness	Ministries use various systems, which require training.	A standardised interface ensures easy adoption, and uniformity.

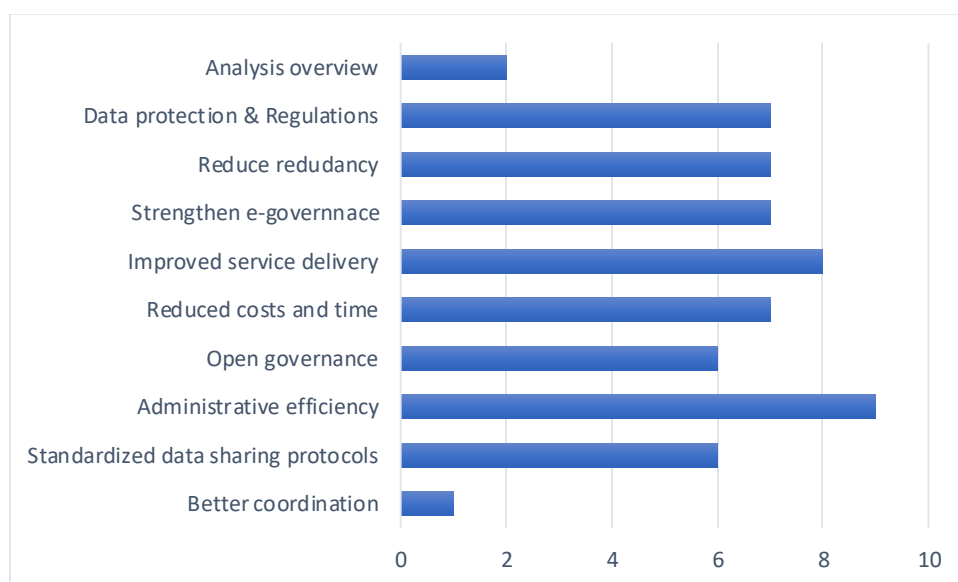


Figure 5.8: Benefits of Centralising Data

As displayed by the *Figure* above, majority of respondents have identified administrative efficiency as a significant benefit of a centralised database, which is the key instrument that will ensure the OPM achieves its role of driving innovation and a citizen-centric Public Service. The respondents were further requested to give their rating on a scale of 1 to 5 in favour of a centralised database, and the *Figure* below shows the outcome:

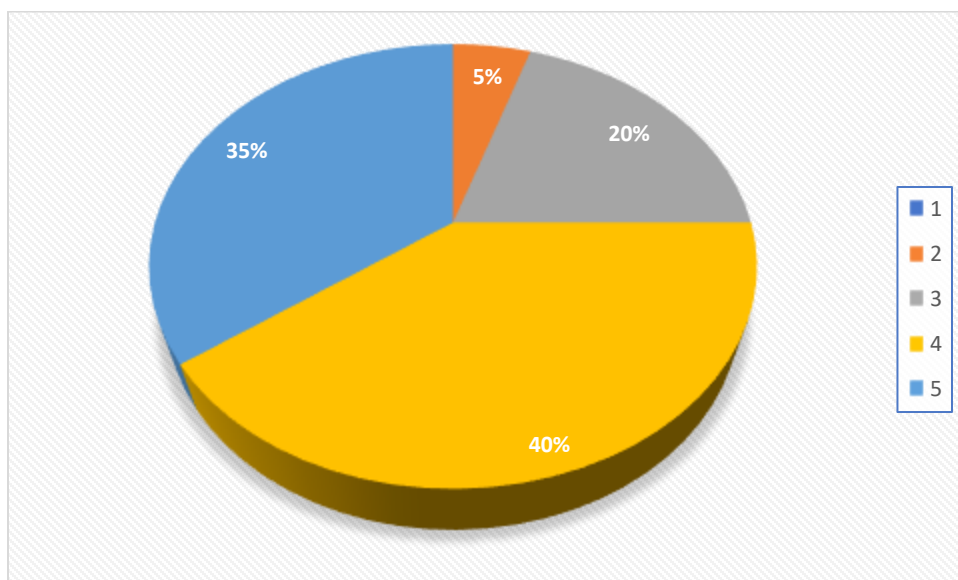


Figure 5.9: Ratings on Centralised Database System Implementation

Displayed by the *Figure* above, 40% of the respondents have rated a 4 and 35% a 5, which brings it to a total of 75%, who are in favour of a centralised system, and 20% undecided, and this can be attributed to the significant benefit of the system which is administrative efficiency.

5.5 Chapter Summary

This chapter outlined the key findings from the research, highlighting the inefficiencies in the Namibian government's data management system and the potential of a Hadoop-based solution. This research has demonstrated not only the technical feasibility of using Hadoop for centralised data integration in a developing-country context, but also its potential to transform governance. The OPM, as the coordinating entity of Namibia's public administration, stands to benefit immensely from adopting such a system. It would empower senior leadership with accurate, timely, and actionable insights, reduce redundancies, and pave the way for effective e-governance.

Additionally, the success of comparative case studies such as Estonia's X-Road and Kenya's Huduma Centres suggests that Namibia is well-positioned to implement similar innovations, especially with the foundational infrastructure piloted in this study.

6. CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

This chapter presents research questions alongside the key findings from the study, as well as recommendations directly informed by the results. The goal is to demonstrate how to operationalise the outcomes of this research to enhance data governance, administrative efficiency, and evidence-based decision-making through the adoption of a centralised Hadoop-based system.

6.2 Findings of the Study

The study investigated the current state of data management practices across Namibian government ministries and how the implementation of a Hadoop centralised database could solve administration inefficiencies at the OPM through data integration. As a result, public service delivery can be enhanced through big data analysis, which would be made possible by the proposed centralised database framework.

The research questions and the associated key findings are as follows:

RQ1: *What are the current data storage and management practices across Namibian government ministries?*

The study revealed that Namibian government ministries use a variety of storage formats, including SQL databases, Excel sheets, CSV files, PDFs, and manual file records. Most ministries use a combination of manual and electronic systems for storing documents, which are stored in separate locations. For example, the Ministry of Health and Social Services at the district level uses a manual system for data entry. In contrast, at both the regional and national levels, both manual and electronic means are employed. This results in siloed and fragmented data systems across all government institutions. To address this, the study presented Hadoop as a suitable solution, offering a unified data storage framework that can store both unstructured and structured files in various formats within HDFS, all within the same directory.

RQ2: *What inefficiencies arise from the lack of centralised data integration at the OPM?*

Findings from the study show that the absence of a centralised repository significantly limits the OPM's ability to oversee and coordinate government activities effectively, and it restricts access to integrated data across the entire government system. This results in data duplication, delays in service delivery, and administrative inefficiencies, ultimately hindering the OPM's capacity to fulfil its mandate of coordinating, enhancing e-governance and managing

government performance effectively. The study has demonstrated the capabilities of Hadoop in automating analysis tasks (e.g., poverty analysis by region), thereby removing dependence on manually aggregated spreadsheets or delayed reports. Hadoop also supports parallel processing, which ensures that even as datasets grow, analysis remains efficient. These capabilities of Hadoop would ultimately improve the OPM's administrative processes.

RQ3: *How can a Hadoop-based system enhance data integration and decision-making?*

Data inconsistencies and standardisation across government ministries pose a significant challenge to centralised data integration, consistency, and automation. Simulated CSVs generated using Mockaroo were standardised with a common identifier, *national_id*. This demonstrated how schema alignment and pre-processing can allow otherwise incompatible datasets to be joined and analysed. By establishing a single repository in HDFS, data was stored once and accessed by all processing jobs. This reduces duplication, streamlines updates, and simplifies version control. A centralised system enforces consistency, ensures data accuracy, and lowers the cost of maintaining multiple disjointed systems.

RQ4: *How effective is a Hadoop-based centralised database system in processing, integrating, and analysing cross-ministerial data for public sector governance?*

The study demonstrated that HDFS and MapReduce framework enabled rapid integration of disparate datasets (e.g., from Labour, Health, Higher Education, Home Affairs), producing accurate cross-ministerial insights in seconds. These capabilities are essential for improving the OPM's capacity for coordinated, timely, and evidence-based decision-making.

6.3 Recommendations

The following recommendations are derived directly from the study's findings and are limited to issues of data management within ministries in the context of a Hadoop-based framework:

6.3.1 OPM

The study revealed that fragmented datasets across ministries prevent the OPM from obtaining timely, integrated insights (e.g., for poverty analysis or emergency responses). To address this:

- **Deploy a Centralised Hadoop Cluster:** Establish and maintain a Hadoop-based data lake at the OPM to serve as the central repository for all government data.
- **Standardise Identifiers:** Mandate the use of *national_id* across all ministries to support seamless integration and cross-sector analysis. Mandate all ministries to submit

data in open, machine-readable formats (e.g., CSV, JSON, XML). Distribute pre-configured data templates tailored to each ministry's function.

- **Enable Secure Access and Analytics:** Implement Role-Based Access Control using Apache Ranger. Develop a user-friendly Hadoop data portal for ministries to submit data requests, access dashboards, and track queries.
- **Streamline Data Pipelines:** Use Apache Sqoop to schedule routine data extractions from ministerial databases into HDFS. Embed ETL pipelines for near-real-time or batch updates.
- **Monitor and Evaluate Usage:** Track portal usage logs, data sharing requests, and inter-ministerial analytics to ensure accountability and optimise performance.

6.3.2 MICT

Since connectivity and ICT infrastructure are prerequisites for centralised data systems, and the 2009 Telecommunications Policy for Namibia stressed the need for robust ICT, and to safeguard the centralised repository, the MICT should:

- **Enforce Data Protection and Privacy Standards:** Finalise and implement Namibia's Data Protection Bill to govern access, storage, and sharing of sensitive citizen data
- **Upgrade Government Portal:** Transform the current portal into an integrated e-services platform that offers online forms, real-time status tracking, submission, and feedback.
- **Support ICT Infrastructure:** Provide secure, high-speed connectivity between ministries and the OPM to enable real-time/scheduled data transfer.
- **Align with e-Government Strategy:** Ensure the Hadoop framework complements Namibia's e-Government Strategic Action Plan (2014-2018) by enforcing interoperability standards.

6.3.3 Department of Public Enterprise

The findings indicate that historic data exists in multiple formats across ministries. These datasets, along with the existing computer infrastructure, should be preserved and integrated as part of the central Hadoop system by:

- **Digitising and Migrating Historical Records:** Convert legacy data records into digital formats and upload to Hadoop-compatible repositories.
- **Optimising ICT Resources:** Leverage existing ICT infrastructure within government institutions by repurposing available computer systems as Hadoop nodes. These nodes

can connect directly to the central Hadoop framework managed at the OPM, enabling continuous data streaming, integration, and analytics.

- **Appointing Data Staff:** Deploy trained Data Capturers at all offices for accurate, daily entry. Assign Regional Data Security & Compliance Officers to monitor data integrity and policy adherence.

6.3.4 Government Ministries

The findings showed that ministries rely on multiple formats (Excel, PDFs, CSVs, and manual files), which prevents interoperability and introduces duplication and delays.

Ministries are therefore encouraged to:

- **Use Data Collection Templates:** Adopt standardised templates for daily operations (e.g., certificates, medical records, employment stats) to ensure consistency. Store data in open, machine-readable formats (CSV, JSON, XML) to enable Hadoop processing.
- **Contribute to Shared Repository:** Upload datasets into ministry-specific directories within HDFS to ensure accessibility and integration.
- **Strengthen Data Reporting:** A report by the Namibia Statistics Agency (NSA) highlighted that mortality and causes of death reports only contained data up to 2021, with many cases classified under “ill-defined causes” due to incomplete or inconsistent manual records (NBC, 2024). This delay limited timely policy responses to public health issues. Therefore, ministries are encouraged to stream data into a centralised Hadoop system, to improve accuracy, and strengthen timely evidence-based decision-making.
- **Implement Customer Feedback Loops:** Mandate short, digital review surveys (max five questions) across ministries to gather citizen input for continuous service improvement. Stream this data to Hadoop for service trend analysis.

6.3.4 National Data Governance Authority (To be established)

The study confirmed inconsistencies and privacy risks when datasets are handled manually:

- **Enforce Compliance Standards:** Oversee compliance, standardisation, and ethical use of integrated data within the Hadoop platform.
- **Audit and Approve Integrations:** Vet and authorise any data systems integrated into the central Hadoop cluster.
- **Educate and Advocate:** Conduct awareness campaigns for public servants and citizens on data protection rights, responsibilities, and the benefits of the system.

6.3.5 IT Departments across Ministries

Lack of skilled personnel for managing integrated systems, could hinder adoption of integrated systems. Therefore:

- **Integrate with Central System:** Configure and setup regional data hubs that integrates with the centralised Hadoop infrastructure and comply with standard data formats
- **Train and Support Staff:** Offer hands-on training for ministry IT teams on Apache Hadoop, Sqoop, data pipelines, and portal usage.
- **Support Ongoing Maintenance:** Establish an inter-ministerial tech support unit to troubleshoot, update, and scale the system as needed.

6.4 Contribution

6.4.1 Practical Impact

The study proposes and demonstrates a Hadoop-based centralised database framework capable of aggregating fragmented datasets across government ministries. By using an open-source ecosystem (Hadoop, HDFS, Java MapReduce, and Power BI), the framework provides:

- A cost-effective alternative to proprietary enterprise data warehouses,
- Scalability for handling high-volume, high-variety government data, and
- A customizable architecture that aligns with Namibia's digital infrastructure maturity.

This technology is practical for resource-constrained governments like Namibia, where licensing costs have historically limited digital transformation initiatives.

6.4.2 Academic contribution

In academia, the study advances the discourse on:

- Applied action research methodologies in the context of African governance systems,
- The suitability of big data architectures like Hadoop in low- and middle-income countries (LMICs),
- The integration of ICT and policy studies, by showing how technical prototypes can inform governance reform.

By focusing on Namibia as a case study, the research provides context-specific insights that apply to other sub-Saharan African governments facing similar digital governance challenges.

6.4.3 Policy formulation

Namibia does not currently have an implemented Data Protection and Privacy Act, although such legislation has been drafted. This research acknowledges this legal void and contributes by:

- Highlighting the urgency of enacting and enforcing data protection legislation to accompany any centralised data initiative,
- Proposing data governance best practices such as encryption and access controls, that can be applied even before formal legislation is in place,
- Presenting a trial case for ethical and secure handling of sensitive citizen data using open technologies, which may inform Namibia's upcoming policy direction.

This contribution is significant as it balances the vision for e-governance with the need for privacy, consent, and citizen trust.

6.5 Limitation of the study

The study was unable to capture the state of data management practices across all 21 government ministries due to time constraints. Additionally, the responses received from the questionnaires were too few to provide a comprehensive view of the inefficiencies present across different government ministries. They may not fully represent ministerial leadership priorities.

While simulated data provided practical insights, real-world implementation may involve additional complexities, such as data privacy laws, integrating legacy systems, and institutional resistance.

6.6 Future Work

While this study demonstrated the feasibility and benefits of implementing a centralised database using Hadoop to support inter-ministerial data integration and socio-economic analysis, several areas remain open for further exploration and development.

First, future work should explore scaling the Hadoop implementation beyond the three-node prototype cluster to a complete government-wide infrastructure. This would involve performance benchmarking, infrastructure stress testing, and assessing cost implications for large-scale deployment.

Second, real-time data processing capabilities, utilising tools such as Apache Kafka or Apache Flink, should be explored to support time-sensitive government services and emergency responses (e.g., disaster relief, pandemic monitoring, and resource allocation).

Future studies should also assess citizen engagement and adoption of the centralised digital platform. This includes designing and evaluating user experience, accessibility, and trust-building mechanisms, particularly in rural areas where digital literacy may be limited.

In addition, researchers should investigate the long-term legal, ethical, and policy implications of large-scale data sharing and centralised storage, particularly regarding privacy, consent, and data sovereignty. Cross-country comparisons, such as with Estonia or Rwanda, may provide valuable policy insights.

Ultimately, the revival and integration of the X-Road data exchange layer within the Hadoop ecosystem presents a promising direction for enabling secure, automated communication between government systems. A technical feasibility study and pilot implementation could provide a roadmap for national deployment.

By addressing these areas, future work can help refine and scale the proposed framework into a robust, secure, and inclusive foundation for digital governance in Namibia.

6.7 Conclusion

According to the World Bank's GovTech metadata of March 2023, 11 African countries had operational integrated government frameworks, 13 had drafts, while 30, which is more than 50%, do not have systematic integration of digital government services, resulting in fragmented service development and duplicated integration costs for public service providers. According to the African Union, (n.d.), their vision is to create an inclusive and integrated digital society and economy in Africa that will improve the quality of life for its people, boost and diversify the continent's economy, and establish Africa as a producer and consumer on the world market. To align with this digital transformation by 2030, Namibia needs to act now and work towards its in-house digital transformation by leveraging the latest technologies, such as Hadoop.

The Office of the Prime Minister, 2014, eGovernment Strategic Action Plan calls for inter-ministerial data sharing, which sets out the vision for digital transformation of public services. This identifies integration of government systems, data sharing, and citizen-centric digital services as core goals.

Nam-X is one of the e-governance projects initiated to enable data exchange between ministries and improve interoperability. Nam-X focuses on structured data exchange, whereas Hadoop system handles unstructured/semi-structured data (PDFs, images), which makes it a viable alternative or complementary framework to Nam-X, and also extends to big data capabilities.

With a strong legal framework in place, ministries can securely digitise their services and stream data into Hadoop while maintaining trust, transparency, accountability, and efficiency. The integration of government systems promotes the adoption of e-government services by lowering costs, enhancing resource utilisation, eliminating redundancies and inefficiencies, and improving overall public service performance. The citizen is always the centre of focus in every government's successful service delivery programmes.

Indonesia's Presidential Decree No. 39/2019 (One Data Indonesia) highlights how Hadoop-driven governance can standardise fragmented datasets and improve decision-making in the absence of mature laws. Namibia can learn from these laws.

There has been a national record of attempts to digitise government but falling short due to lack of integration, data silos, or technical limitations. This study proposes a next-generation solution aligned with national development objectives (e.g., Vision 2030, Harambee Prosperity Plan) with cost savings because it is an open-source framework.

REFERENCES

- Abuljadail, M., Khalil, A., Talwar, S., & Kaur, P. (2023). Big data analytics and e-governance: Actors, opportunities, tensions, and applications. *Technological Forecasting and Social Change*, 193. <https://doi.org/10.1016/j.techfore.2023.122612>
- African Union. (2020). *The Digital Transformation Strategy for Africa (2020-2030)*. www.au.int
- Al-Sai, Z., & Albualigah, L. (2017). *Big Data and E-government: A review*.
- Apache Hadoop. (2025a, August 20). *Apache Hadoop Yarn*. <https://Hadoop.Apache.Org/Docs/Stable/Hadoop-Yarn/Hadoop-Yarn-Site/YARN.Html>.
- Apache Hadoop. (2025b, August 20). *HDFS Architecture*. <https://Hadoop.Apache.Org/Docs/Stable/Hadoop-Project-Dist/Hadoop-Hdfs/HdfsDesign.Html>.
- Apache Hadoop. (2025c, August 20). *MapReduce Tutorial*. <https://Hadoop.Apache.Org/Docs/Stable/Hadoop-Mapreduce-Client/Hadoop-Mapreduce-Client-Core/MapReduceTutorial.Html>.
- Asenahabi, B. M. (2019). Qualitative research, Mixed method research. In *International Journal of Contemporary Applied Researches* (Vol. 6, Issue 5). www.ijcar.net
- Axel Domeyer, Solveigh Hieronimus, Julia Klier, & Thomas Weber. (2021). Government data management for the digital age. *Mckinsey & Company*.
- Bhushan Jadhav, Archana B. Patankar, & Sonali B. Jadhav. (2018). *A Practical approach for integrating Big data Analytics into E-governance using hadoop*. IEEE.
- Center for African Studies. (2023). *E-government in Namibia*. E-Governance Knowledge Hub.
- Chen, J., Lv, Z., & Song, H. (2019). Design of personnel big data management system based on blockchain. *Future Generation Computer Systems*, 101, 1122–1129. <https://doi.org/10.1016/j.future.2019.07.037>
- Coursera. (2025, May 27). *What Is a Hadoop Cluster*. <https://Www.Coursera.Org/Articles/Hadoop-Cluster>.
- Davies, J. E., Spear, D., Ziervogel, G., Hegga, S., Ndapewa Angula, M., Kunamwene, I., & Togarepi, C. (2020). Avenues of understanding: mapping the intersecting barriers to adaptation in Namibia. *Climate and Development*, 12(3), 268–280. <https://doi.org/10.1080/17565529.2019.1613952>
- DLA Piper. (2025). *Data Protection Laws of the World*.
- Dolan, J., Satapathy, S., & Sabiti, B. (n.d.). *Data can drive shared prosperity for governments, businesses, and citizens. Unlocking it requires trusted data exchange*.
- El Mohajir, Mohammed., Al Achhab, Mohammed., Chahhou, Mohamed., & Eddine El Mohajir, Badr. (2016). *Designing A Health Data Management System Based Hadoop-Agent*. IEEE.
- Garzon, C. (2024, November 7). *Data Engineering Tools Comparison: Snowflake vs Redshift vs BigQuery*. DataEngineerAcademy.
- Giest, S. (2017). Big data for policymaking: fad or fasttrack? *Policy Sciences*, 50(3), 367–382. <https://doi.org/10.1007/s11077-017-9293-1>
- Gomathy, C. K. (2022). *The Study of Hadoop Ecosystem*. <https://doi.org/10.55041/IJSREM16965>
- Greeshma, L., & Pradeepini, G. (2016). Big Data Analytics with Apache Hadoop MapReduce Framework. *Indian Journal of Science and Technology*, 9(26). <https://doi.org/10.17485/ijst/2016/v9i26/93418>
- Güvenoğlu, E., & Razbonyalı, C. (2016). Traditional Data Storage Methods and the Big Data Concepts. *International Research Journal of Engineering and Technology*. www.irjet.net

- Hannan, S. A. (2016). An overview on Big Data and Hadoop. In *Article in International Journal of Computer Applications International Journal of Computer Applications* (Vol. 154, Issue 10). <https://www.researchgate.net/publication/372631347>
- Hong, W., Zhang, Y., & Lu, Y. (2018). *The Development of E-government under the Influence of Big Data*.
- Jadhav, B. (2018). Opportunities and Challenges in integrating Cloud Computing and Big Data Analytics to E-governance. In *International Journal of Computer Applications* (Vol. 180, Issue 15).
- Martinez-Mosquera, D., & Luján-Mora, S. (2019). Framework for big data integration in e-government. *DYNA (Colombia)*, 86(209), 215–224. <https://doi.org/10.15446/dyna.v86n209.77902>
- MICT. (2021). *Data Protection Bill final draft bill*.
- Mohammed, A., & Saleh, B. M. (2017). Centralized Database: A Prerequisite for Security and Sustainable Development in Nigeria. *International Journal of Innovative Research in Computer Science & Technology*, 5(1), 209–213. <https://doi.org/10.21276/ijircst.2017.5.1.7>
- Namibia Statistics Agency. (2023). *Namibia 2023 Population and Housing Census Report*.
- Nashilongo Gervasius. (2020). *Data Protection and Privacy in Namibia*.
- Nengomasha, C., Nengomasha, C. T., & Nyanga, E. H. (2012). Managing semi-current records: A case for records centres for the public service of Namibia. In *Journal for Studies in Humanities and Social Sciences* (Vol. 1, Issue 2). <https://www.researchgate.net/publication/309195503>
- Nengomasha, C. T., Abankwah, R., Uutoni, W., & Pazvakawambwa, L. (2018). Health information systems in Namibia. *Information and Learning Science*, 119(7–8), 358–376. <https://doi.org/10.1108/ILS-03-2018-0015>
- Office of the Prime Minister. (2014). *eGovernment Strategic Action Plan for the Public Service of Namibia*.
- Office of The Prime Minister. (2023). *Customer Service Charter*.
- Omar, H. K., & Jumaa, A. K. (2019). Big Data Analysis Using Apache Spark MLlib and Hadoop HDFS with Scala and Java. *Kurdistan Journal of Applied Research*, 4(1), 7–14. <https://doi.org/10.24017/science.2019.1.2>
- Oyeniran, O. A., Oyeniran, S. T., & Oyeniyi, J. (2019). *A Robust National Centralized Database System for Identity Management and Security Control in Nigeria*. <https://www.researchgate.net/publication/337635606>
- Poladi, H. (2023). *Unraveling the Power of Apache Hadoop: A Comprehensive Guide*. <https://www.researchgate.net/publication/379372817>
- Rogger, D., & Schuster, C. (2023). *The Government Analytics Handbook: Leveraging data to strengthen public administration*.
- Rutto, D., & Yudah, O. (2015). Huduma Centres Revolutionizing Service Provision in Kenya: Lessons for the Information Profession. *International Journal of Science and Research (IJSR) ISSN*, 6. <https://doi.org/10.21275/ART20174146>
- Saleh, Z. I., Obeidat, R. A., & Khamayseh, Y. (2013). A Framework for an E-government Based on Service Oriented Architecture for Jordan. *International Journal of Information Engineering and Electronic Business*, 5(3), 1–10. <https://doi.org/10.5815/ijieeb.2013.03.01>
- Scherenberg, F., Hellmeier, M., & Otto, B. (2024). Data Sovereignty in Information Systems. *Electronic Markets*, 34(1). <https://doi.org/10.1007/s12525-024-00693-4>
- Sileyew, K. J. (2019). *Research Design and Methodology*. <https://doi.org/10.5772>

- Silva, Y. N., Almeida, I., & Queiroz, M. (2016). SQL: From traditional databases to big data. *SIGCSE 2016 - Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 413–418. <https://doi.org/10.1145/2839509.2844560>
- Simuni, G., & Amaranatha, A. (2024). *Hadoop in Enterprise Data Governance: Ensuring Compliance and Data Integrity*.
- Somasundaram, P. (2020). Cloud Storage Strategies for High-Performance Analytics: An In-Depth Look at Databases, Data Warehouses, and Object Storage Solutions. *International Journal of Science and Research (IJSR)*, 9(7), 2004–2009. <https://doi.org/10.21275/SR24203232850>
- Vassil, K. (2015). *Digital Dividends Estonian e-Government Ecosystem: Foundation, Applications, Outcomes Estonian e-Government Ecosystem Foundation, applications, outcomes*.
- World Bank Group. (2016). *Namibia Identity management system Analysis*.
- World Health Organization. (2020, April 14). *WHO updates COVID-19 dashboard with better data visualization*. https://www.who.int/news-room/feature-stories/detail/who-updates-covid-19-dashboard-with-better-data-visualization?utm_source=chatgpt.com.

APPENDICES

Appendix A: Clearance Certificate



FACULTY RESEARCH ETHICS COMMITTEE (F-REC)
DECISION/FEEDBACK ON THE RESEARCH PROPOSAL

Dear Pombili Haitamba

RESEARCH TOPIC: LEVERAGING BIG DATA TO IMPLEMENT A CENTRALIZED DATABASE ACROSS NAMIBIAN GOVERNMENT MINISTRIES.

Supervisor (if applicable): Dr Richard Malikwata

Qualification registered for (if applicable): Master of Data Science

(Reference number of applications: FACULTY RESEARCH ETHICS COMMITTEE REGISTRATION NUMBER: **FREC - 28/24**)

Re: Ethical screening application No: **FREC - 28/24**

The Faculty of Computing and Informatics Ethics Screening Committee of the Namibia University of Science and Technology reviewed your application for the above-mentioned research. The research as set out in the application has been:

Approved	X
----------	---

(Indicate with an X, and N/A if not applicable and proceed)

We would like to point out that you, as a researcher, are obliged to maintain the ethical integrity of your research, adhere to the ethical guidelines of NUST, and remain within the scope of your research proposal and supporting evidence as submitted to the F-REC. Should any aspect of your research change from the information as presented to the F-REC, which could affect the possibility of harm to any research subject, you are under the obligation to report it immediately to your supervisor or F-REC as applicable in writing. Should there be any uncertainty in this regard, you must consult with the F-REC.

We wish you success with your research and trust that it will make a positive contribution to the quest for knowledge at NUST.

Any ethical issues that need to be highlighted?	Why are these issues important?	What must/could be done to minimize the ethical risk?
No	N/A	N/A

Recommendation: The application is approved.

Sincerely,

Prof. Suama L. Hamunyela
Chairperson: Faculty Ethics Screening
Committee Tel: +264-61-207-2922
CC: Co-supervisor:



Appendix B: Letter of Permission from OPM



REPUBLIC OF NAMIBIA

OFFICE OF THE PRIME MINISTER

Tel No: (061) 287 9111
Fax No: (061) 234 296

Private Bag 13318
WINDHOEK

Enquiries: M. Nekoro
061 – 287 2328

4 July 2024

Pombili Haitamba
P O Box 210
Ongha

Dear Ms. Haitamba,

RE: PERMISSION TO CONDUCT ACADEMIC RESEARCH STUDY AT THE OFFICE OF THE PRIME MINISTER

1. Reference is made to your letter dated the 20 June 2024, requesting for permission to conduct research at the Office of the Prime Minister.
2. Kindly note that your request to conduct academic research within the Office of the Prime Minister titled "Leveraging big data to integrate a centralized database within Namibia's Office of the Prime Minister (OPM)" has been approved.
3. You are hereby urged to uphold information obtained in this study, especially for the Office of the Prime Minister with confidentiality and to share a copy of the dissertation/ report with the Office of the Prime Minister.

Yours Sincerely

I-BEN NATANGWE NASHANDI
EXECUTIVE DIRECTOR
OFFICE OF THE PRIME MINISTER



All official correspondence must be addressed to the Executive Director

Appendix C: Research Questionnaire

Topic: Designing a Centralised Database using Hadoop for Integrating Big Data Analysis in the Namibian Office of the Prime Minister (OPM)

Introduction:

Hadoop is an open-source framework that provides a cost-effective and efficient way to manage and analyse big data across clusters of computers using simple programming models. This study looks into the feasibility of a Hadoop framework at the Office of the Prime Minister (OPM) for data integration across all government ministries in comparison with the current data management practices to eliminate inefficiencies in their administrative operations.

Part A: Demographic Information

Position/Role:

Government Ministry:

Department:

Part B: Data Management Systems

1. **How is information stored within your ministry?** *Excel/PDF/Word/HTML/Other*
2. **Does your ministry currently use any special computer systems for capturing data?** *Yes/No*. If no, skip to question 3
 - i) **How is the data captured on these systems** (i.e., scanner, digital forms, manually)?
 - ii) **What are some of the challenges with these systems?** (Choose all that apply)
Redundancy/Delayed decision making/Data inconsistencies/Outdated systems/Internet connectivity issues/Slow to retrieve information/Scalability Issues due to high volumes of data/Difficult to extract analysis.
 - iii) **What are the benefits of using these systems?**
Easy to use/Secure and fast/Easy to extract analysis/Easy to integrate with other systems/High availability/Can handle the increasing volumes of data
3. **How many databases does the ministry have at the moment and what are they each used for?**
4. **List the names of databases that you know of and what they are each used for?**
5. **Are they integrated?**
6. **How many ministries/departments do you need to constantly interact with on a daily basis to achieve your daily tasks?**
7. **How do you request information from other ministries/departments** (email, phone call, face-to-face)? *Email/Face-to-Face/Phone call/Other*
8. **What format do you use to share information across ministries/departments** (pdf, word, manual)? *Pdf/Manual/Word/Other*
9. **What is the average time set to respond to information requests within your ministry** (hrs, days, weeks)
10. **What is the average time set to respond to information requests to the public** (hrs, days, weeks)
11. **What data security measures are in place to protect sensitive information within your ministry's data systems?** (Choose all that apply)

Passwords/Two-factor-authentication/Smart-cards/Firewalls/Antivirus/Data masking/Audit logs/Policies/Data encryption/Time based restrictions/Biometric/Other

12. Does everyone within your ministry/department have equal access to the information?

13. Are there any backups of data? If yes, where is it kept?

14. What are challenges that you foresee in implementing centralised data systems? (Choose all that apply)

Data integration issues/Single point of failure/Scalability issues/Security risks/Lack of skilled staff/High maintenance/Regulatory complexity/Other

15. In your opinion, what are the potential benefits of centralising government data across ministries? (Choose all that apply)

Administrative efficiency/Open governance/Enhanced data protection and regulations/Standardised data sharing protocols/Reduced costs and time/Improved service delivery/Strengthen e-governance/Better analysis overview/Reduce redundancy/Other

16. On a scale of 1-5, how would you recommend a centralised database across all government ministries

Appendix D: Raw Data

Records from Mheti (10)

national_id	year_of_graduation	education_level	institution_attended	current_enrollment_status	region
19957797	2009	Secondary		Dropped Out	Erongo
19967213	2013	Tertiary	IUM	Dropped Out	Erongo
19976244	2008	Primary		Dropped Out	Kavango West
19996666	2009	Tertiary	IUM	Dropped Out	Omaheke
19908904	1992	Secondary		Dropped Out	Otjozondjupa
19910825	1997	Secondary		Dropped Out	Caprivi
19977866	2009	Tertiary	VTC	Dropped Out	Ohangwena
19931635	2016	Tertiary	VTC	Dropped Out	Hardap
19989680	2013	Secondary		Dropped Out	Hardap
19964888	1996	Tertiary	IUM	Dropped Out	Erongo

Records from mol (10)

national_id	employment_status	job_sector	income_bracket	contract_type	field_of_study	region
19957797	Self-employed	Education	High	Temporary	Law	Caprivi
19967213	Employed	Technology	Low	Full-time	Engineering	Omusati Oshana
19976244	Self-employed	Education	Low	Part-time	Medicine	Caprivi
19996666	Unemployed	Services	High	Temporary	Medicine	Otjikoto
19908904	Unemployed	Other	Low	Part-time	Computer Science	Erongo
19910825	Unemployed	Other	Low	Part-time	Law	Khomas
19977866	Employed	Other	Low	Full-time	Business	Omusati Oshana
19931635	Self-employed	Other	Low	Part-time	Education	Omaheke
19989680	Unemployed	Agriculture	Low	Full-time	Engineering	Erongo
19964888	Self-employed	Education	High	Full-time	Law	Caprivi

Records from mhaiss (10)

year_of_birth	citizenship	Region	Gender	national_id	Ethnicity
1995	Namibian	Kavango West	Male	19957797	Damara
1996	Namibian	Erongo	Non-binary	19967213	Damara
1997	Namibian	Caprivi	Male	19976244	Kavango
1999	Namibian	Khomas	Female	19996666	Damara
1990	Namibian	Hardap	Female	19908904	White
1991	Namibian	Otjikoto	Female	19910825	Herero
1997	Namibian	Khomas	Female	19977866	White
1993	Namibian	Erongo	Male	19931635	Ovambo
1998	Namibian	Omaheke	Female	19989680	Coloured
1998	Namibian	Hardap	Female	19964888	Nama

Appendix E: Hadoop 3-node Cluster Configurations

These codes were adopted from:

https://www.tutorialspoint.com/hadoop/hadoop_multi_node_cluster.htm

Prerequisites

The following prerequisites should be met before proceeding with the setup,

- A virtual machine with at least 4 GB of RAM, 100 GB of disk space, and a multi-core CPU.
- A Linux-based operating system – *Ubuntu 22.04*
- SSH installed and configured for password less login
- Java Development Kit (JDK) installed (Hadoop requires Java 8 or later)

Steps for Setup and Configuration

Step 1: Java Development Kit (JDK)

Java version 8 or higher was Installed, which is a necessary component of Apache Hadoop. It is required to download and install a Java Development Kit on all the nodes in the network where Hadoop was installed, to be able to create applications that run on Hadoop.

Firstly, hostname configuration was done on all nodes, to ensure, each node communicates with the other by name.

This was done by accessing, the `/etc/hosts` file on all the nodes by running the command below
nano /etc/hosts

Then, edited by adding the following lines

192.168.0.11 master

192.168.0.12 dn1

192.168.0.13 dn2

By doing so, the nodes were then renamed as master, dn1, dn2, and thereafter saved.

Then the following command was used to install Java.

sudo apt update && sudo apt install openjdk-11-jdk

Now, to verify that the correct version is installed, the following command was run to ensure the version is 8 or later to avoid runtime issues and compatibility bugs.

java -version

Hadoop is written in Java and requires the Java Development Kit (JDK) to compile, run, and manage its components (like HDFS, MapReduce, and YARN). The JAVA_HOME environment variable specifies the location of the Java installation. Hadoop needs this path to execute Java commands correctly. So, it is important to locate the JAVA_HOME directory by running the following command *dirname \$(dirname \$(readlink -f \$(which java)))*.

Running the command as shown below, displays the location of java, that is saved on your machine as it will be required later.

```
pombili@master:~/Desktop$ dirname $(dirname $(readlink -f $(which java)))  
/usr/lib/jvm/java-11-openjdk-amd64
```

Step 2: Install and configure SSH

Next was the SSH (Secure Shell) installation which is vital for Hadoop as it enables secure communication between nodes in the Hadoop cluster. Without it, Hadoop cannot manage distributed nodes effectively. It also ensures data integrity, confidentiality, and allows for efficient distributed processing of data across the cluster. This was done by using the following commands

```
sudo apt install openssh-server openssh-client
```

Now to configure password-less SSH access for the newly created Hadoop user, an SSH keypair was first generated: *ssh-keygen -t rsa*

The generated public key is then copied to dn1 and proper permissions set by running the following commands:

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorised_keys  
chmod 640 ~/.ssh/authorised_keys ssh-copy-id pombili@dn1
```

The last command was then used to copy to dn2 by replacing dn1 with dn2 in the command line. The system will require an authentication host to add RSA keys to dn1 & dn2, by typing yes and hit Enter to authenticate the master node

To ensure that password-less connection has successfully been configured, the following SSH command lines are used to access the dn1 and dn2 from the master nodes to see if the login in will require a password or not.

```
ssh dn1
```

```
ssh dn2
```

Step 3: Install Hadoop on Ubuntu

Once Java is successfully installed and ssh is setup, to install Hadoop, Apache Hadoop and all its related components, including Hive, Pig, Sqoop, etc were downloaded from Hadoop's official website, ensure to download the binary archive (not the source).

The following command was used to download Hadoop 3.3.4 binary archive, (the latest version at the time):

```
wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz
```

Hadoop is distributed as a compressed tarball (.tar.gz) to make downloading and sharing easier.

Extract the file using this command. The command below is used to unpacks Hadoop's binaries, libraries, config files, etc., into a usable directory, and renamed to remove version information

```
tar xzvf hadoop-3.3.4.tar.gz
```

```
mv hadoop-3.3.4 hadoop
```

Next, Hadoop and Java Environment Variables were configured by opening the ~/.bashrc file using the command: `sudo vi ~/.bashrc`

The below lines were added to the file, of which the first line contains the JAVA_HOME environment variable location that was configured during the Java installation.

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

```
export HADOOP_HOME=/home/pombili/hadoop
```

```
export HADOOP_INSTALL=$HADOOP_HOME
```

```
export HADOOP_MAPRED_HOME=$HADOOP_HOME
```

```
export HADOOP_COMMON_HOME=$HADOOP_HOME
```

```
export HADOOP_HDFS_HOME=$HADOOP_HOME
```

```
export HADOOP_YARN_HOME=$HADOOP_HOME
```

```
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
```

```
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

```
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

The above configuration is added to the current environment by running the following commands, to activate the environment changes and verify that Hadoop is correctly configured in the system path

```
source ~/.bashrc
```

```
echo $HADOOP_HOME
```

Next the `hadoop-env.sh` file is opened in the vi editor, to configure environment variables for Hadoop, especially the `JAVA_HOME`, this is to ensure that Hadoop knows which Java version to use when running its components.

```
vi $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

The below lines are added to the file, to set the java path.

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

Step 4: Configuring Hadoop

Next is the configuration of Hadoop files available under every directory.

Firstly, the namenode and datanode directories should be created inside the user home directory which are essential components in the HDFS. The namenode directory stores HDFS metadata like file structure, block locations, and permissions while the datanode directory stores the actual data blocks of files. These were created by running the following command to create both directories: `mkdir -p /home/pombili/hdfs/{namenode, datanode}`

Then the `core-site.xml` file accessed by running the following command

```
vi $HADOOP_HOME/etc/hadoop/core-site.xml
```

The lines below are added between the two `<Configuration>` `</Configuration>` tags

```
<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
  <description>The default file system URI</description>
</property>
```

Then the `mapred-site.xml` file accessed by the following command:

```
vi $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

The following lines are added between the two `<Configuration>` `</Configuration>` tags

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

Then the `hdfs-site.xml` file using the following command

```
vi $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

The following lines are added between the two `<Configuration>` `</Configuration>` tags

```
<property>
  <name>dfs.replication</name>
  <value>2</value>
</property>
<property>
  <name>dfs.name.dir</name>
  <value>file:///home/pombili/hdfs/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>file:///home/pombili/hdfs/datanode</value>
</property>
```

Then the `yarn-site.xml` file using the following command

```
vi $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

The following lines are added between the two `<Configuration>` `</Configuration>` tags

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
```

Lastly, the workers file on the master node and the data nodes roles are defined to ensure that the nodes register their assigned roles

```
vi $HADOOP_HOME/etc/hadoop/workers
```

The following lines are added to the file

```
master
dn1
dn2
```

Then, all these configurations that were done on the master node, are copied to both dn1 and dn2 respectively using the commands below.

```
scp ~/.bashrc pombili@dn1:~/.bashrc
```

```
scp $HADOOP_HOME/etc/hadoop/* pombili@dn1:$HADOOP_HOME/etc/hadoop/
```

Step 5: Starting Hadoop Cluster

Once the setup is complete, the Hadoop cluster can now be started, but before that the namenode need to be formatted for the new hadoop user to ensure that all required files are created with the right permissions, allowing HDFS to run smoothly.

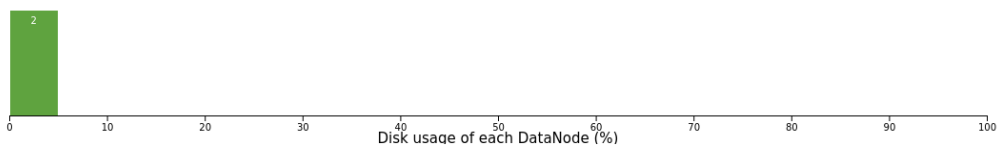
The following command is used to format the Hadoop namenode: `hdfs namenode -format`

Then the Hadoop cluster can be started with the following command

```
start-all.sh
```

```
pombili@master:~/Desktop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as pombili in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [master]
Starting datanodes
Starting secondary namenodes [master]
Starting resourcemanager
Starting nodemanagers
```

Once all the services started, Hadoop can be accessed at: <http://localhost:9870>, and the Hadoop application page is available at <http://localhost:8088>.



In operation

DataNode State: All | Show: 25 entries | Search:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
✓ [default-rack/dn2:9866] (192.168.0.13:9866)	http://dn2:9864	1s	6m	24 KB	14.83 GB	19.02 GB	0	24 KB (0%)	3.3.4
✓ [default-rack/dn1:9866] (192.168.0.12:9866)	http://dn1:9864	0s	6m	24 KB	14.13 GB	19.02 GB	0	24 KB (0%)	3.3.4

Figure 6: Hadoop Cluster

The figure above shows that 2 datanodes are currently up and live on the Hadoop cluster. We the go ahead and setup the government ministries directories on Hadoop and upload data in different structures for processing.

Creating Ministries directories

The directories were created in hdfs environment in Hadoop for all 21 government ministries. These directories are where the various structured and unstructured files will be stored, within their respective ministries (the ministries are named as their corresponding abbreviations).

```
hdfs dfs -mkdir -p / pombili /O_P_M
hdfs dfs -mkdir -p / pombili /O_J
hdfs dfs -mkdir -p / pombili /M_A_W_L_R
hdfs dfs -mkdir -p / pombili /M_o_D
hdfs dfs -mkdir -p / pombili /M_o_E
hdfs dfs -mkdir -p / pombili /M_E_F_T
hdfs dfs -mkdir -p / pombili /M_o_F
hdfs dfs -mkdir -p / pombili /M_F_M_R
hdfs dfs -mkdir -p / pombili /M_H_S_S
hdfs dfs -mkdir -p / pombili /M_H_E_T_I
hdfs dfs -mkdir -p / pombili /M_H_A_I_S_S
hdfs dfs -mkdir -p / pombili /M_I_T
hdfs dfs -mkdir -p / pombili /M_I_R_C_O
hdfs dfs -mkdir -p / pombili /M_I_C_T
hdfs dfs -mkdir -p / pombili /M_o_J
hdfs dfs -mkdir -p / pombili /M_O_L
hdfs dfs -mkdir -p / pombili /M_M_E
hdfs dfs -mkdir -p / pombili /M_P_E
hdfs dfs -mkdir -p / pombili /M_S_Y_N_S
hdfs dfs -mkdir -p / pombili /M_O_W
hdfs dfs -mkdir -p / pombili /M_U_R_D
```

Appendix F: Java Codes

Mapper Classes

```
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import java.io.IOException;

public class EducationMapper extends Mapper<Object, Text, Text, TaggedRecordWritable>
{
    @Override
    protected void map(Object key, Text value, Context context) throws IOException,
    InterruptedException {
        if (value.toString().contains("national_id")) return;
        String[] parts = value.toString().split(",");
        if (parts.length >= 1) {
            String nationalId = parts[0];
            context.write(new Text(nationalId), new TaggedRecordWritable("mheti",
            value.toString()));
        }
    }
}
```

```
public class HomeAffairsMapper extends Mapper<Object, Text, Text, TaggedRecordWritable> {

    @Override

    protected void map(Object key, Text value, Context context) throws IOException,
    InterruptedException {

        if (value.toString().contains("year_of_birth")) return;

        String[] parts = value.toString().split(",");

        if (parts.length >= 5) {

            String nationalId = parts[4];

            context.write(new Text(national_id), new TaggedRecordWritable("mhaiss", value.toString()));

        }

    }

}
```

```
public class LabourMapper extends Mapper<Object, Text, Text, TaggedRecordWritable> {

    @Override

    protected void map(Object key, Text value, Context context) throws IOException,
    InterruptedException {

        if (value.toString().contains("national_id")) return;

        String[] parts = value.toString().split(",");
```

```

    if (parts.length >= 1) {
        String national_id = parts[0];
        context.write(new Text(national_id), new TaggedRecordWritable("mol", value.toString()));
    }
}
}

```

Reducer Class

```

import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
import java.io.IOException;

public class PovertyAnalysisReducer extends Reducer<Text, TaggedRecordWritable, Text, Text> {
    @Override
    protected void reduce(Text national_id, Iterable<TaggedRecordWritable> values, Context context)
    throws IOException, InterruptedException {
        String mhaiss = "", mheti = "", mol = "";
        String region = "";

        for (TaggedRecordWritable val : values) {
            switch (val.getTag()) {
                case "mhaiss":
                    home = val.getRecord();
                    region = mhaiss.split(",")[2]; // Extract region from home affairs data
                    break;
                case "mheti":
                    mheti = val.getRecord();
                    break;
                case "mol":
                    mol = val.getRecord();
                    break;
            }
        }

        if (!mhaiss.isEmpty() && !mheti.isEmpty() && !mol.isEmpty()) {
            String[] eduFields = mheti.split(",");
            String eduLevel = eduFields[2].trim();
            String enrollment_status = eduFields[4].trim();

            String[] labourFields = labour.split(",");
            String income = labourFields[4].trim();

            boolean likelyPoor = eduLevel.equalsIgnoreCase("Tertiary") &&
                enrollmentStatus.equalsIgnoreCase("Dropped Out") &&

```

```

        income.equalsIgnoreCase("Low");

        if (likelyPoor) {
            context.write(new Text(region), new Text("At Risk of Poverty"));
        }
    }
}
}

```

Driver Class

```

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.MultipleInputs;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.io.Text;

public class PovertyAnalysisDriver {
    public static void main(String[] args) throws Exception {
        if (args.length != 4) {
            System.err.println("Usage: <mheti> <mhaiss> <mol> <output>");
            System.exit(2);
        }

        Configuration conf = new Configuration();
        Job = Job.getInstance(conf, "Namibia PovertyAnalysis");
        job.setJarByClass(PovertyAnalysisDriver.class);

        MultipleInputs.addInputPath(job, new Path(args[0]), TextInputFormat.class,
HomeAffairsMapper.class);
        MultipleInputs.addInputPath(job, new Path(args[1]), TextInputFormat.class,
EducationMapper.class);
        MultipleInputs.addInputPath(job, new Path(args[2]), TextInputFormat.class,
LabourMapper.class);

        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(TaggedRecordWritable.class);

        job.setReducerClass(PovertyAnalysisReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(Text.class);

        FileOutputFormat.setOutputPath(job, new Path(args[3]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}

```

Appendix G: Poverty Analysis Output File

Omusati Oshana At Risk of Poverty
At Risk of Poverty
Erongo At Risk of Poverty
Omusati Oshana At Risk of Poverty
Caprivi At Risk of Poverty
Otjikoto At Risk of Poverty
Kavango West At Risk of Poverty
At Risk of Poverty
Khomas At Risk of Poverty
Omaheke At Risk of Poverty
Otjikoto At Risk of Poverty
Kunene At Risk of Poverty
Kavango West At Risk of Poverty
Omaheke At Risk of Poverty
Otjikoto At Risk of Poverty
At Risk of Poverty
Erongo At Risk of Poverty
Omaheke At Risk of Poverty
Hardap At Risk of Poverty
Otjikoto At Risk of Poverty
Otjikoto At Risk of Poverty
Otjozondjupa At Risk of Poverty
Kavango West At Risk of Poverty
Otjozondjupa At Risk of Poverty
Ohangwena At Risk of Poverty
Kavango West At Risk of Poverty
Erongo At Risk of Poverty
Hardap At Risk of Poverty
Kavango West At Risk of Poverty
Omusati Oshana At Risk of Poverty
Hardap At Risk of Poverty
At Risk of Poverty
Hardap At Risk of Poverty
Caprivi At Risk of Poverty
Kunene At Risk of Poverty
At Risk of Poverty
Kunene At Risk of Poverty
Kunene At Risk of Poverty
Otjikoto At Risk of Poverty
Omaheke At Risk of Poverty
Erongo At Risk of Poverty
At Risk of Poverty
Kunene At Risk of Poverty
Otjikoto At Risk of Poverty
Karas At Risk of Poverty
Khomas At Risk of Poverty
Kavango West At Risk of Poverty

Caprivi At Risk of Poverty
Kunene At Risk of Poverty
Erongo At Risk of Poverty
Omaheke At Risk of Poverty
Karas At Risk of Poverty
At Risk of Poverty
Karas At Risk of Poverty
Otjozondjupa At Risk of Poverty
At Risk of Poverty
Khomas At Risk of Poverty
At Risk of Poverty
Omaheke At Risk of Poverty
Kunene At Risk of Poverty
Karas At Risk of Poverty
Otjozondjupa At Risk of Poverty
Otjozondjupa At Risk of Poverty
Caprivi At Risk of Poverty
Caprivi At Risk of Poverty
Khomas At Risk of Poverty
Khomas At Risk of Poverty
Hardap At Risk of Poverty
Erongo At Risk of Poverty
At Risk of Poverty
Caprivi At Risk of Poverty
Khomas At Risk of Poverty
Kunene At Risk of Poverty
Khomas At Risk of Poverty
Erongo At Risk of Poverty
Karas At Risk of Poverty
Hardap At Risk of Poverty
Caprivi At Risk of Poverty