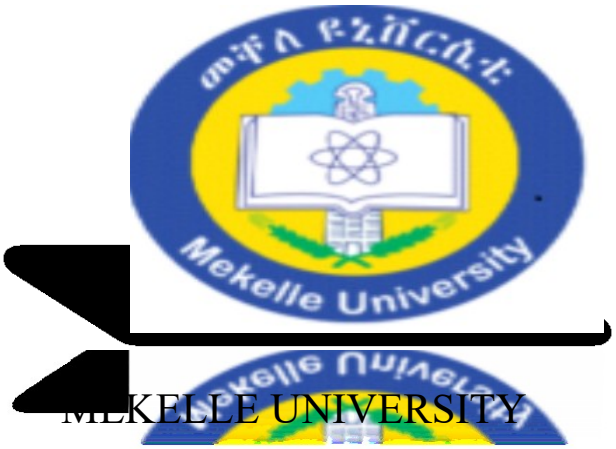


MEKELLE UNIVERSITY

ETHIOPIAN INSTITUTION TECHNOLOGY MEKELLE

SCHOOL OF EiT-M

DEPARTMENT OF INFORMATION TECHNOLOGY

Design and Implementation of Big Data Analytics Framework at Abyssinia Bank: A Case Study in the Oromia Special Zone, Amhara Region

By

Bayew Girma

ID.NO. EiT-M/PS255/10

A Thesis Submitted to the School of Computing EiT-M in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology

ADVISOR: Dr. Behailu Getachew (phd)

Mekelle, Ethiopia

August 2025

Mekelle University

Ethiopian Institution Technology Mekelle

Department of EiT-M

Design and Implementation of Big Data Analytics Framework at Abyssinia Bank: A Case Study in the Oromia Special Zone, Amhara Region

By

Bayew Girma

August 2025

Name and Signature of Members of the Examining Board

Dr. Behailu Getachew (PhD)	_____
Advisor	Signature
_____	_____
Examiner	Signature
Dr. Kindie Biredagn	
Examiner	Signature

DECLARATION

I declare that this thesis is my original work and has not been submitted for any Degree in any other University. I have undertaken the study independently with the guidance and support of the research advisor.

Signature: _____

Bayew Girma

This thesis has been submitted for examination with my approval as university advisor.

Advisor: _____

Dr. Behailu Getachew (PhD)

External Examiner: _____

_____

Internal Examiner: _____

_____

ACKNOWLEDGEMENTS

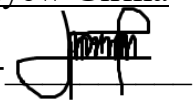
First and foremost, praise to my Almighty God for his blessings and guidance in giving me the strength, courage, patience and perseverance to start and finalize this study.

I am deeply grateful to my research advisor ***Dr. Behailu Getachew*** for his valuable comments and support. I sincerely acknowledge his invaluable help and constant encouragement throughout the course of this research. I am most grateful for his advice and suggested approaches that gave me direction of the study.

In addition, I acknowledge encouragement from the department's lecturer's and my Colleagues both at work and at school throughout the thesis work.

Finally, I most gratefully acknowledge My Family's and my friends for all their support throughout the period of this research.

Name:- Bayew Girma

Signature:- 

Date:- 14/09/2025

Abstract

The Ethiopian banking industry is one of the fastest growing in the economy, Abyssinia Bank is one of the first banks in Ethiopia and it is among the top list of banks in the country. Big data analytics (BDA) has become an increasingly popular topic over the years amongst academics and practitioners alike. Big data, which is an important part of BDA, was originally defined with three Vs, being volume, velocity and variety. The rapid emergence of big data presents significant opportunities for organizations to enhance decision-making processes through data-driven insights. This thesis explores the design and implementation of a big data analytics framework at Abyssinia Bank, focusing on its application in the Oromia Special Zone of the Amhara Region. This study aims to identify specific use cases where big data analytics can improve operational efficiency, customer service, and strategic decision-making. Through a mixed-methods approach that incorporates quantitative analysis of existing data and qualitative interviews with key stakeholders, this research will contribute to a deeper understanding of how big data can transform banking practices in emerging markets.

Keywords: *Big data, Big data management, Big data Governance, Big data Management framework*

Table of Content

Contents

List of Table

List of Figure

List of Acronym

Acronym	Description
ATM	Automated Teller Machine
BDA	Big Data Analytics
BDM	Big Data Management
BI	Business Intelligence
DI-BDAF-BoA	Design and Implementation of Big Data Analytics Framework at Abyssinia Bank
CIO	Chief Information Officer
DSRM	Design science research Methodology
DW&BI	Data ware house and business intelligence
IBM	International Business Machines
ICT	Information communication Technology
ITDR	Information Technology Design research
IT	Information Technology
MIS	Management Information System
NBE	National Bank of Ethiopia
VP	Vice President
BoA	Bank of Abyssinia

CHAPTER ONE

INTRODUCTION

1.1. BACKGROUND

The term Big Data refers to large growing data sets that include heterogeneous formats: structured, unstructured and semi-structured data. Big Data has a complex nature that requires advanced algorithms and powerful technologies. And in the case of big data applications, the traditional static Business Intelligence tools can no longer be effective. (Oussous, Benjelloun, Ait Lahcen, & Belfkih, 2018)

"It is a technology that helps to receive and store data generated from a variety of sources, including digital processes, social media, sensors, and mobile devices and use analytics to gain useful insight on customer behavior. The advancement made in both storage technology and computing power has made it feasible to collect and store this data (Yan, 2013)." (OCHIENG, 2015)

Big Data Definition What is big data? Big data is data that is characterized by five V's: Volume, Variety, Velocity, Veracity and Value. ¬ Volume - represents the amount of data collected by a company, with data sets ranging from terabytes to zeta bytes in size. ¬ Variety - different types of data are available from a range of data sources, both external and internal to the firm, from sources such as customer databases, social media, ATM transactions and the like. ¬ Velocity – refers to the speed or frequency at which data are gathered and processed. It could be in real time or near time sensors, sales transactions and social media posts. ¬ Veracity –the degree in which the data is trusted and protected from unauthorized access to make a decision. ¬ Value – refers to economically useful benefits that an organization obtained from big data. The usefulness of the data is measured by how predictive the data is and how useful the information is that the organization can derive from this data. Big data provides no value by itself. The value of big data comes from the ability to analyze data that was not previously available, or were too expensive to store or process, to provide new insights and improve the basis of decision making. As such data is either difficult or impossible to manage using traditional database or analytics tools, companies have begun exploring new technologies.

Big data management is about two things—big data and data management—plus how the two work together to achieve business and technology goals. Big data is mainly about the quantity of data, including massive data sets measured in tens of terabytes, or at times in hundreds of terabytes or peta bytes. Whereas big data management is the application of data management disciplines, tools, and platforms to the management of big data.

The banking sector has witnessed an exponential increase in the volume of data generated daily. To remain competitive, financial institutions must leverage Big Data analytics to uncover valuable insights from their data. Abyssinia Bank can enhance operational efficiency, improve customer satisfaction, and strengthen its risk management capabilities through a structured analytics framework. This paper aims to outline the design and implementation of a Big Data Analytics Framework utilizing Hadoop, addressing the needs and goals of Abyssinia Bank.

In today's data-driven environment, organizations face an overwhelming volume of data that, if harnessed correctly, can lead to significant improvements in decision-making processes. This proposal outlines the design and implementation of a Big Data Analytics Framework aimed at enhancing decision-making capabilities across the organization.

1.1.1. Banking Big Data Analytics

Big Data Analytics can become the main driver of innovation in the banking industry; it has many benefits if the banks start exercising Big Data management.

Examples of using Big Data in the industry of banking

Customer spending patterns

The banks have direct access to a wealth of historical data regarding the customer spending patterns. They know how much money you have been paid as a salary in any given month, how much has gone to your saving account, how much has gone to your utility providers, and so on. This provides a reach basis for further analysis. Applying filters such as festive seasons and macroeconomic conditions helps banking employees to understand whether the customer's income is steadily increasing and whether the expenditure remains sufficient.

This is one of the cornerstone factors for risk assessment, loan screening, mortgage valuation, and multiple financial products such as insurance cross-sales. (Fedak, 2018)

Transaction channel identification

The banks benefit greatly from understanding whether their customers withdraw the entire sum available on the payday in cash, or whether they prefer to keep their money on the credit / debit card. Clearly, one can approach the latter customers with the offers to invest in short-term loans with high payout rates, etc. (Fedak, 2018)

Customer segmentation and profiling

If the initial analysis of consumer spending habits and preferred transaction channels is complete, multiple correct profiles will segment the customer base. Easy spenders, cautious investors, prompt loan repayers, deadline rush returners ... Knowing all customers ' financial profiles helps the bank assess next month's expected expenditure and income and make a detailed plans to secure the bottom line and maximize income. (Fedak, 2018)

Big Data management will provide banks with in-depth insights into customer spending habits and patterns, simplifying the task of identifying their needs and wishes. By tracking and tracking each and every customer transaction, banks can categorize their customers based on different parameters, including commonly accessed services, preferred credit card spending, or even net worth. The advantage of customer segmentation is that it enables banks to target their customers better with related marketing campaigns tailored to suit their needs. (Mauricio, 2016)

Product cross-selling

Why not offer a better return on interest to cautious investors to stimulate them to spend more actively? Is it worth providing a short-time loan to an easy spender who already struggles to repay a debt Precise analysis of the financial backgrounds of the customers ensures that the bank can more efficiently cross-sell the auxiliary products and better engage customers with customized offers. (Fedak, 2018)

Fraud management & prevention

One of the biggest problems faced by the banking sector is fraud. And Big data management Will enable banks to ensure that no unauthorized transactions are made, providing a level of safety and security that will raise the safety standard for the industry as a whole. (Mauricio, 2016)

Knowing an individual's normal spending habits helps raise a red flag when something unusual is happening. If a prudent investor who wants to pay with his card attempts to withdraw all the money from his account via an ATM, this could mean that fraudsters robbed the card and used it.

A call from a bank demanding a clearance for such an action helps to understand quickly whether the cardholder does not know of a valid claim or a fraudulent conduct. Analysis of other kinds of transactions helps to cut down the risk of fraudulent actions greatly. (Fedak, 2018)

Risk assessment, compliance & reporting

A similar risk assessment technique may be used when trading stocks or screening a loan applicant. Understanding a client's expenditure patterns and past credit history can help quickly assess the risks of issuing a loan. Big Data algorithms can also help resolve compliance, audit and reporting problems to streamline processes and remove the overhead for managers. (Fedak, 2018)

Banks now have access to millions or even billions of customer needs, and they can now use Big Data Management to more meaningfully cater for them. Based analytics packages can synchronize with your Big Data systems in real time, creating dynamically actionable insight. Big Data management will expand the banking sector in a way that will enable them to earn revenue from mores by reducing costs. And by cutting down on unnecessary costs, the banking industry can provide customers with exactly what they’re looking for, instead of irrelevant information. (Mauricio, 2016)

1.2. Overview of Abyssinia Bank in Amhara Region Oromo Special Zone

Bank of Abyssinia (BOA) is a share company of private individuals who amass experience and success in different areas including business, entertainment and education.

The BOA established in 1905, was given a 50-years concession by the Emperor Menelek II. In Amhara Reign Oromo Zone one of BOA in Kemisse, Cheffa, Senbete,Bati and Wolledi.

Table 1 Current States of Abyssinia Bank

Total Capital	ETB 17.5 billion
Deposits	ETB 192.5 billion
Loans & advances	ETB 23.9 billion
Number of Depositors	ETB 2.8 million
Number of Mobile banking users	ETB 2,450,000
Profit before tax ETB	ETB 5.2 billion
Number of branches	ETB 929
ATM	ETB 1757
Virtual Banking Center	ETB 73
Customer	ETB 11,510,000
Employee	ETB 11,065

1.3. Statement of the Problem:

As more and more fields involve Big Data problems, ranging from the global economy to society administration and from scientific researches to national security, we have entered the era of Big Data. (C.L. Philip Chen & Chun-Yang Zhang, 2014)

Experts in the financial industry define big data as a tool that enables an organization to create, manipulate and manage very large data sets within a given timeframe and the storage required to support the data volume, characterized by the variety, volume, and velocity. The big data revolution taking place in and around the 21st century has found a resonance with the banking firms, taking into account the valuable data they have stored for many decades.

In the information age, almost every big company encounters Big Data problems, data that are used within the banking sector continue to grow, while the sources of the data also continue to increase. The increase in size and velocity constitutes big data in the environments.

The high velocity and increasing size of the big data lead and contribute to banking sector challenges, such as inaccuracy and inconsistency of customer data. In the last decade, when it comes to operations and service delivery, the banking sector has grown by leaps and bounds.

Major challenges in implementing big data analytics include legacy systems integration, cyber security concerns, data quality issues, skill gaps in data science, and ethical considerations around data usage.

1.4. Research Questions:

1. What is the current status of Design and Implementation of Big Data Analytics Framework to Enhance Decision-Making at Abyssinia Bank?
2. What is the awareness level of Big Data analytics framework?
3. What appropriate Big Data Analytics framework can be developed for Abyssinia Bank

1.5. Objectives:

1.5.1. General Objective

The primary goal of this framework is to enable Abyssinia Bank to leverage big data analytics to enhance decision-making, optimize operations, and improve customer satisfaction.

1.5.2. Specific Objective

1. **Improve Customer Insights:** Analyzing customer data to deliver personalized banking experiences.
2. **Enhance Risk Management:** Utilizing predictive analytics to identify and mitigate risks and fraudulent activities.
3. **Optimize Operational Efficiency:** Streamlining processes by identifying inefficiencies and improving resource allocation.
4. **Data-Driven Decision Making:** Providing executives with actionable insights for strategic planning.

5. Evaluate the proposed framework.
6. Assess the current practice of BDM and awareness level of staffs at the bank.
7. Review Big Data Management related literature including Big Data management frameworks.
8. Identify design areas and requirements for Big Data management for the bank.
9. Develop Big Data management framework for the

1.6. Significance of the study

The significance of this research is in three ways, in that; it contributes from practical, methodological, and theoretical perspectives. Abyssinia bank will consume the benefits of the proposed framework by providing an efficient and effective data driven services and products to its customers. From a methodological perspective, the study provides a guide in the form of a framework, on how Big Data management can be conducted on bank big data. From a theoretical perspective, the study adds to existing literature, particularly from big data analytics and banking perspectives.

1.7. Scope

The scope of the study encompasses suggestion and development of Big Data management framework using a design science research approach. The study is intended to develop a framework which will be used for showing the ways on utilizing the vast amount of data incorporated in the bank's database in a productive way.

1.8. Methodology

The study follows the design science research methodology for the overall course of the study. The method will be applied to assess the current big data management practices, success factors, problems that impede on the implementation of big data management at Abyssinia Bank and to design the proposed big data management framework for the bank. The study will be conducted using survey questionnaire and interviews as a method of data collection and design science research methodology as a research paradigm.

The researcher, on the research methodology part of the study has discussed What, How, and why is each step of the research process done in detail.

1.9. Limitation and Delimitation of the study

The result of the research would be more comprehensive if it covers the entire Banks in Ethiopia. But due to COVID-19 pandemic, Ethiopian War (In Tigray, Amhara, Oromia and others) moving from place to place and gathering data was an impossible task, as a result, the researcher decides to conduct the thesis on a single bank.

The study is delimited to Abyssinia Bank. Sub branch in Oromo special zone is a location where major data services and equipment and IT personnel office are located. Such employees can have better knowledge about the study than other employees who work at branch level.

1.10. Organization of the Thesis

The study is organized in six sections. The first section introduced by giving an overview of the study and the problem statement, the research questions are included. Chapter two covers literatures on the concept of big data, big data management and big data usage in the banking area; and related works done on big data management frameworks in order to understand how it is implemented or used. Chapter three describes the research methodology and strategy followed to identify the potential big data management framework that could be provided to banking sector. It begins by describing the research approach. Primary data collection and analysis technique were described as information acquisition method. Chapter four presents the results of the survey questionnaire and interview described in chapter three. The data collection method utilized was a structured questionnaire and semi structured interview that provided both control and direction through a list of questions. In Chapter five, backgrounds of a framework for building the proposed big data management framework for BoA, the detail elements of the proposed big data management framework are explained. Finally, in chapter six conclusions about the research and suggestions for future research direction were presented.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. WHAT IS BIG DATA?

Big data refers to massive, complex datasets generated at high speed from numerous sources, including structured, unstructured, and semi-structured data like text, images, and videos. The term big data was coined in 2012 and has since become one of the most trending topics in technology, business and management. This interest for big data and associated topics is a result of the significance big data is believed to have on our society in the near future. Supposedly big data will change the way people live their everyday lives, their work and the way their homes are connected. But first what do we mean by big data? In the proceeding sessions I will show the definition of the term big data given by different scholars.

Big data is a rather new term, as indicated by Google trends (2014), however it is still surprising that up until now there is no clear or uniform definition of big data. Opinion leaders and companies working in the field have their own opinion and definitions of big data at this point. Clear is however that big data embodies an accumulation of different technologies under the heading "Big Data". (Blasiak, (2014))

(Laurila, (2017).)States "Even though data volumes have been skyrocketing throughout the 21st century, big data as a term became widespread as recently as 2012 (Gandomi & Haider, 2015 Due to the sudden increase of interest, defined by Ward and Barker (2013) as exponential since 2011, and the ambiguity of the concept. The rapid adoption by the public and private sectors and the rapid growth of big data technology meant that the academic environment had little room for debate to mature (Gandomi & Haider, 2015). From various fields, early big data literature led to several interpretations that were vague and sometimes contradictory. (Ward & Barker, 2013)."

We are living in the digital world and the data and the technology are integral part of the system. The technology has enabled us to use the transaction online while at the same time it has generated enormous amount of data which is somewhere eating up the storage space. At one side the technology is gearing up to provide more space and creating cloud technology to provide and meet up the requirement of the massive data which is being generated while at the same time others are busy in finding ways to use this data for their businesses and make it a business. (Chandani, (2015).)

(Chandani, Mita, Neeraja, & Om, 2015) Defined big data as big data is the data which is huge in quantity and which is captured by IT system in prevalence; it is too big and complex to be analyzed using the traditional software.

The quantum and the speed at which data is being generated is tremendous; but if evaluated and applied correctly, it could help the company a long way by giving them a deep insight into a vantage situation thereby enabling a better decision making, as the name would itself suggest, big data is an enormous or huge data-set, with a massive and complex volume so as to make it extremely difficult to process in the way traditional datasets are being managed as of today. The huge dataset poses excessive challenges in terms of analyzing, capturing, storing, sharing, visualizing, presenting and securing, as it is unwieldy. To understand more on the nature of big data, it is often characterized by having: 1. Too much volume 2. Too much velocity (with which it comes or the speed with which it keeps coming to an organization such as a Bank). Too much variety, as in today's context all sort of data with different formats, in the form of messages, images, clips, pdf, XML, etc., keep coming-up.

"One such definition for big data was provided by Oracle: big data is the derivation of value from traditional relational database driven business decision making, augmented with new sources of unstructured data (Ward & Barker, 2013). The core idea behind this definition is inclusion: including additional data sources for augmentation of existing operations is the core of big data. A more concrete quantification of big data was provided by Intel: big data organizations "generate a median of 300 terabytes (TB) of data weekly" (Ward & Barker, 2013). Another definition is provided by Microsoft: "Big data is the term increasingly used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information" (Ward & Barker, 2013). While there were and still are many individuals and organizations trying to fine-tune the definition for big data, one definition has been cited and expanded on more than any other: the three V's." (Laurila, 2017)

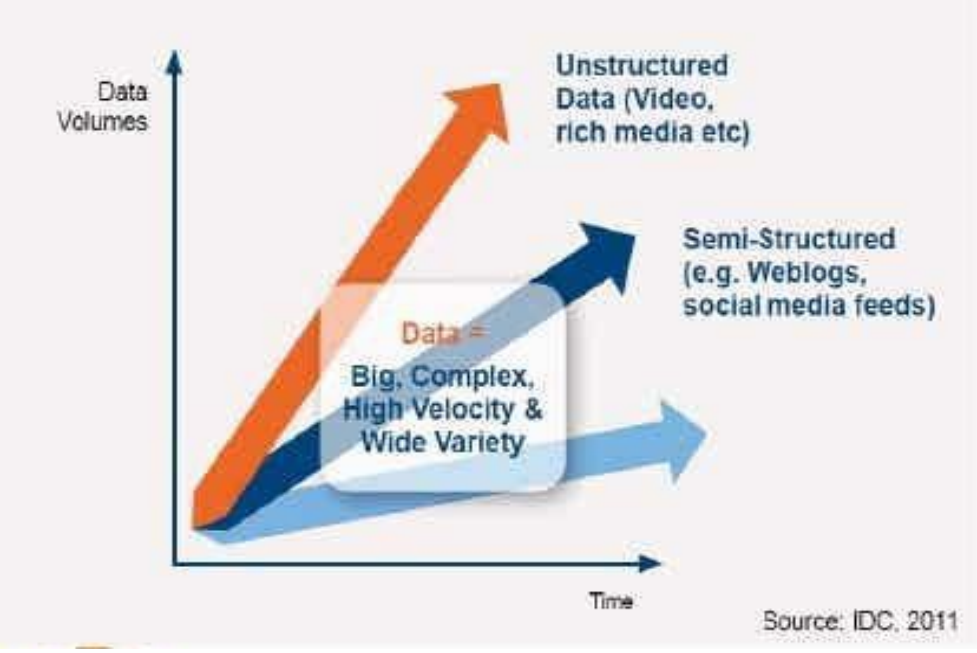


Figure 1Definition of big data (Aliu, 2019)

(Laurila, 2017) stated the common way of defining big data is through three V's "Today, the most common way of defining big data is through the three V's, data volume, data variety and data velocity. First introduced by Doug Laney in 2001 in his research note "3D Data Management: Controlling Data Volume, Velocity, and Variety", the three V's has become what is closest to the industry standard way of defining big data. The reason that the three V's has become such a prevalent way of defining big data, is that it helps you understand that it is not only the size of the data set that constitutes big data. (Russom, 2011)"

On the other hand, (Aliu, 2019) characterizes and defines big data through five V's as follows

I. **Volume**: The quantity of analyzed data exceeds the capability of conventional analytics and statistical modeling tools. Big data itself relates to voluminous size hence size is germane to determining value out of data. Pence (2014) saw the sheer volume of data stored explode and IBM predicts 35 zettabytes will be stored by 2020. The Economics and Business Research Centre and Business Research (2013) is of the view that as data volumes increase, there is a need for greater sophistication of technologies to harness the benefits that can be derived from it.

II. **Velocity**: It refers to speed at which data is accessed. While conventional Business Intelligence (BI) applications use historical data from weeks, months, or quarters, big data relies on real-time information to provide quick insights. According to Harvard Business Review (2012), the rate or speed by which data is produced is as important as the other characteristics as this provides rapid vision that can in turn provide competitive advantage and confirmed to be beneficial to organizations.

III. **Variety**: This represents information expansion resulting into heterogeneous data types (textual, numeric, etc.), encoding, formats, semantics, structures (structured, semi-structured, unstructured), syntax etc. Usually, data can be separated into structured, semi-structured and unstructured data. According to Gartner (2013), unstructured data represents 85% of data while structured data is described as data grouped into rows and columns, making it easy to query and obtain information for an organization's Operational needs. On the other hand, semi-structured data is a blend of both structured and unstructured data.

IV. **Veracity**: This is tantamount to the quality of data. It involves the Origin, authenticity, availability, accountability and trustworthiness of data.

V. **Value**: This refers to the ability to convert a tsunami of data into business. Customer data and requests have different values than data generated by GPS devices. Data can be transactional, descriptive, statistical, hypothetical or historical. Each of them has different value, hides different information (Andrej, 2013)

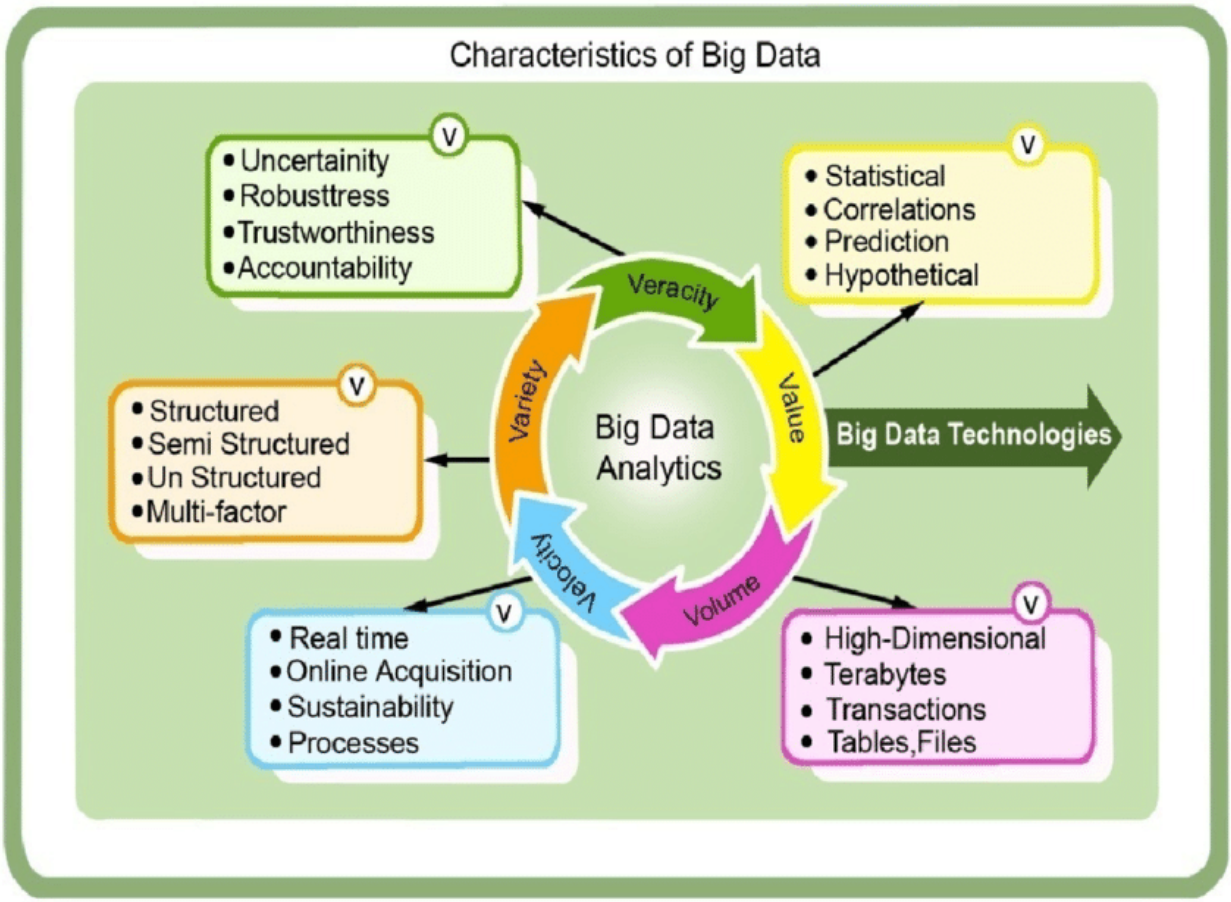


Figure 2 5V's of big data (Aliu, 2019)

2.2. Bank Big data

Although not a new concept, big data is now gaining the world's attention like never before. Some call it the "new oil", given its growing reputation as a valuable, largely untapped resource. Indeed, today we see data unleashed across many different ways of life; as a growing global consensus believes it could transform the way the world works dramatically. (Barnes, 2019)

(Gagnon, 2019)A perfect example of how technology has revolutionized customer service is the banking industry. Gone are the days when on a Saturday morning, consumers had to wait in line just to deposit their paycheck. Customers can now use their cell phone to check their account balance, check deposit, pay bills and move money, they don't even need to leave the building of their home.

The financial-services industry is already one of the world's most data-driven Banks currently have huge amounts of customer data at hand including compliance checks by KYC (know your customer), customer activity at ATMs (automated teller machines), point-of - sale purchases and online banking profiles. Today, data and data analysis are increasingly serving as the basis not only for more understanding of customers but also to improve internal processes, such as operations and compliance, as well as greatly boost scalability. (Barnes, 2019)

These self-service features are fantastic for customers, but they're one of the main reasons traditional banks struggle to compete with similar businesses and financial institutions online only. Since customer activity now occurs mostly online, certain in-person services known to be provided by brick-and - mortar banks are no longer relevant to customer needs. (Gagnon, 2019)

This is where the adoption of big data strategies and instruments becomes so important for the banking sector. The banks can use personal and transactional information to create a 360-degree view of their customers in order to: (Gagnon, 2019)

- Track customer spending patterns
- Segment customers based on their profiles
- Implement risk management processes
- Personalize product offering
- Incorporate retention strategies
- Collect, analyze, and respond to customer feedback

The ability to access, analyze, and manage vast volumes of data while rapidly evolving the Information Architecture has long been critical to financial services companies as they improve business efficiency and their performance. Recently, bank profitability has been on the rise, especially in regions of the world where economic conditions are good. Financial services organizations will continue to focus on revenue growth and higher margins through operational efficiency, better risk management, and improved customer intimacy. Banks will also develop new revenue streams by entering new markets and service areas. (Oracle, 2015)

Using information technology to enhance the customers' experience in both retail and business banking can help grow interest-based and fee-based revenue.

Many larger financial organizations are gravitating towards expanding portfolios for wealth management to ensure lower risk and consistent fee-based revenues. Differentiated Services, cross-selling and up-selling initiatives, and expansion into emerging global wealth-management markets are on the rise. Analytics and control of knowledge play a central role in ensuring that these strategies are properly executed. (Oracle, 2015)

The ultimate goal of a Big Data pipeline is to facilitate analytics on the available data. Big Data analytics provides the ability to infer actionable insights from massive amounts of data and can assist with information discovery during the process. It has become a core component that is being deployed and used by entities operating at various spheres of the financial field. (Aliu, 2019)

Big Data mining provides a huge opportunity to stand out from the competition, as banks are now being digitized. All banking transactions are a bulk of data, so the industry is sitting on massive information stores. Banks can advance or reinvent virtually every aspect of banking by using data science to collect and analyze big data. Big data analytics has thus become a key innovation driver in financial sector. (Aliu, 2019)

(Oracle, 2015)As financial services firms embark on a quest to achieve a deeper understanding of consumers and their household needs to provide reliable and differentiated services, the volume of data is growing, data collection is more frequent, and data diversity is becoming more complex. Today, these data sources can include:

Traditional enterprise data from operational systems related to customer touch points such as:

- ATMs
- Call Centers
- Web-based and mobile sources
- Branches / Brokerage units
- Mortgage units
- Credit cards
- Debt cards

Financial business forecasts from various sources such as:

- News
- Industry data
- Trading data
- Regulatory data
- Analyst reports (internal and competing banks)

As the rate that this data is generated increases, business analysts who crave such data rapidly consume it. Information discovery tools allow them to quickly combine different data sets leading to better insight. They also want more data to be processed at higher levels and stored longer, and want the rising amounts of data to be analyzed more quickly. Solutions "Big Data" help financial services and banking institutions respond to those needs. (Oracle, 2015)

Oracle puts Big Data and analytic capabilities as part of a “next-generation” architecture that can meet the needs of the dynamic financial services and banking industries.

Every day, financial institutions find new ways to leverage the power of big-data analytics in banking a journey of exploration powered by technical innovation. Machine learning and artificial intelligence (AI) models are two such innovations, combine big data and automation to optimize data quality management and customer segmentation, reduce errors, and make it easier for banks to make groupings and review product data and customer preferences. (Gagnon, 2019)

Companies in the consumer banking and financial services industry typically have data warehouses and business intelligence tools for reporting on and analyzing customer behavior to better anticipate their needs, and for optimizing operations. By deploying Big Data Management Systems that include data reservoirs (featuring Hadoop and / or NoSQL databases), greater benefits in these areas can be achieved as the business gains more predictive capabilities and becomes more agile. The addition of Big Data systems enables organizations to gain much higher levels of insight into data faster and enables more effective decision making.

2.3. What is Big Data management

Big data management is important to companies and other organizations that have big data to manage, but big data is still relatively new. (Russom, 2013)

Today, we gather and store data from a multitude of sources such as biological science and analysis, sensors and cell phones and their applications, social media activity, mobile devices and, to name a few, automated sensors. Technology is also paving the way for new and better hardware. However; Big Data is driving the development of storage hardware, and network infrastructure, with all its computing and storage needs, the most important infrastructure aspect of Big Data analytics is storage. (Adam, A. Majid, & Jasni , 2014)

Big data management is about two things—big data and data management—plus how the two work together to achieve business and technology goals. (Russom, 2013)

(Russom, 2013)Defines both the terms and later coined them to get better understanding of the concepts. Below are the definitions given by him.

Big data is mainly about the quantity of data, including massive data sets measured in tens of terabytes, or at times in hundreds of terabytes or peta bytes. We spoke about very large databases (VLDBs) before the word big data became common parlance. VLDBs usually contain exclusively structured data, managed in a database management system (DBMS). In many organizations, big data and its management follow the VLDB paradigm. (Russom, 2013)

He adds to the above definition of big data, in addition to very large data sets, big data can also be an eclectic mix of structured data (relational data), unstructured data (human language text), semi-structured data (RFID, XML), and streaming data (from machines, sensors, Web applications, and social media). In this report, the term multi-structured data refers to data sets or data environments that include a mix of these data types and structures. (Russom, 2013)

Second (Russom, 2013) defines Data management. He articulates data management Includes data collection and storage, plus processing and delivery — whether traditional, new big data, or both. Processing may be extensive, particularly if data is repurposed for use that differs from its or Data management is a broad activity covering a variety of data disciplines, including data

warehousing, data creation, data consistency, data management, content management, event planning, database management, and so on. (Russom, 2013)

Lastly (Russom, 2013) defines Big data management (BDM). This is where data management disciplines, tools, and platforms (both old and new) are applied to the management of big data (in the base definition or the extended one). (Russom, 2013)

Traditional data and new big data can be quite different in terms of content, structure, and intended use, and each category has many variations within it. To accommodate this variety, BDM's software solutions aim to provide various types of tools and frameworks for data management, as well as diverse user skills and practice. (Russom, 2013)igin (as is common in business intelligence [BI], data warehousing [DW], and analytics).015)

2.4. How is Banks Operating towards Big Data Management?

Generally banks are shifting to be data driven service providers and they are taking many steps to achieve this data-driven services and products. Unrelenting pressure from non-traditional players is driving financial services agencies to transform themselves digitally. In order to become data-driven businesses, banks and insurance companies need to address three key trends in data management: data volume, ubiquity and user demands. (Egetoft, 2019) Mobile apps and devices generate massive volumes of data from new sources, such as images, audio, and video. Combined with emerging business models and value chain players that accelerate the digitalization of financial services, this new data offers opportunities for companies to gain additional insight and value. (Egetoft, 2019) Today, data is everywhere. And financial services firms need to collect it all: customer records, financial transactions, product and service purchasing history, consumer journeys, marketing campaigns, business inquiries, market reports, social media networks, Internet of Things (IoT) networks, device logs, and text messages (including emails and SMS), plus other newer outlets. (Egetoft, 2019)

(Egetoft, 2019) User demand for this data is rising. In Today's financial services company, it is important to recognize that every employee is truly a user of analytics who needs:

- Decision support, allowing users to base decisions on empirical evidence rather than gut fee Trust in the security and accuracy of data

- The ability to proactively anticipate and influence business outcomes by paying attention to new and increasingly forward-looking signals
- Self-service access to data and easily usable analytics tools
- Speed and intelligent information equivalent to what users experience with personal consumer technology.

These expanding data sources and volumes create a new challenge: an increasingly complex data management landscape of hundreds of silos at the enterprise. It's not unusual for firms to deploy multiple data lakes, data warehouses, operational applications, mobile apps, online apps, call centers, IoT sensors, and analytics solutions. Data can be located in hybrid environments, on-premises, and in the cloud. (Egetoft, 2019)

Companies need to combine their existing and new data into one single data universe to reduce complexity. Universal data helps companies improve visibility, providing insights that can enhance efficiency, automation and growth. By transforming the data into insights, organizations can become smart businesses. (Egetoft, 2019)

For many financial services enterprises, however, a single data universe is still an aspiration. More often, data resides in multiple siloed environments. Because data is not meaningfully connected across these silos, it has become less accessible – compromising insight into customers, partners, products, sales channels, and financial performance. (Egetoft, 2019) If data is the new oil, then the banks are the Exxon Mobil of the financial world. (Gascoigne, 2019)As mass amounts of data meet ever-increasing regulation in the world of finance, sophisticated data management has never been more important. How a bank handles this complex problem will make or break its position as a global player. (Mizuno, 2015)

With digital transformation being a must for banks wishing to stay in business over the long haul, banks must address a variety of challenges – e.g., rising data volumes, data omnipresence and user data demands – to become the data-driven enterprises they need to be. Huge volumes of data are flowing in into a number of formats from mobile apps and devices like images, audio, and video, and data is all over the place. Banks are swamped with data on customers, from financial transactions, customer purchase histories, marketing campaigns, social media streams, third-party sources, text messages and more. Users themselves – i.e., bank employees – are

potential analytics users who need trustworthy, secure data they can use for decision support and analysis, with self-service access to both data and easy-to-use analytics.

(SAS, 2020) The quantity of data banks hold on individuals is so huge it needs “warehouses” or “lakes” to store or make sense of it. (Gascoigne, 2019) While data is so important for making decisions which yields greater benefits (Gascoigne, 2019) argues that most banks are not using their data efficiently by saying, for most chief data officers within the banking industry, simply consolidating existing data is enough of a task to keep them busy. But, according to a recent white paper, why data culture matters, from consultants McKinsey, the number-one takeaway when looking to improve is that your “data culture is decision culture… The fundamental objective in collecting, analyzing and deploying data is to make better decisions”. Continuing his argument he added “There are those who would argue a regulated industry such as banking has other commitments, with compliance playing a large part in banking data management. But Rob Casper, chief data officer at J.P. Morgan Chase, says: “If you simply rely on having huge quantities of data in a data lake, you’re kidding yourself. Volume is not a viable data strategy. The most important objective is to find those business problems and then dedicate your data management efforts towards them.”” (Gascoigne, 2019) Sadly, for banks’ customers, the industry has had little need to make those efforts so far. Hans Tesselaar, executive director at BIAN, a not-for-profit association that was established to promote banking interoperability, points out that historically there was little or no competitive pressure on banks. (Gascoigne, 2019) In reality, financial services organizations have not been very good at taking advantage of the mounds of data at their fingertips. Much of their data are stored in ancient, siloed systems that were not built to handle the large volumes of modern times, nor the complex queries needed in today's financial environment. It can be painfully slow, complicated and costly to move data across the pipeline to extract insights both in terms of infrastructure and manpower. (The financial brand, 2020)

With the continual introduction of new banking regulations by global regulatory bodies, the burden of fully complying has increased considerably. Some regulations require banks to prepare new data sets while liquidity regulations, such as the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR), require banks to provide their balance sheets with data source.

New regulatory concepts, such as frameworks for risk appetite and stress tests, will inevitably require banks to further improve their data management systems. (Mizuno, 2015)

In response, most major banks are currently building platforms for data management building large-scale data infrastructure that addresses every potential business and regulatory need. While a multi-purpose data platform is a step in the right direction, due to their sheer scope and "do everything" strategy, these ambitious projects sometimes fail on that very task. (Mizuno, 2015) Specifically, Banks in Japan usually handle data management using an Information Management System (MIS). An MIS is a collection of IT platforms from aggregating data to reporting to senior management that are used for all stages of the process. All banks need to develop that as a basis for a risk tolerance system. (Mizuno, 2015) Other firms like SQream, the developer of the SQream DB data analytics acceleration platform, has produced a free step-by-step guide entitled 'Best Data Management Practices for Financial Services Data Professional.' (The financial brand, 2020)The guide provides financial services data professionals with a clear list of steps they can take to optimize data ingestion, manage their data more efficiently and ultimately gain better business intelligence. And others like (SAS, 2020) SAS (Analytics software and solutions) provide a single, unified data management framework and automated processes – from data collection and aggregation to data validation and cleansing. Their banking and financial analytics software offers the most complete, sophisticated data management capabilities, including:

- Data quality and integration. Consolidate data from internal and external sources, track data lineage, and reduce errors and inconsistencies with embedded data quality tools.
- Data preparation. Prepare data for analytics and reporting in a self-service environment without coding or help from IT. Spend less time preparing data and more time analyzing it.
- Data governance. Spend less time managing data with a consistent set of policies and processes that ensures your data's conformity with established policies and regulatory standards.
- Personal data protection. Satisfy your customers' needs for personal data security. Fully understand and monitor the state of your data.
- Data federation and stewardship. Eliminate the complexities of data integration with a virtual data environment that simplifies administration and security while delivering a complete data picture in a business-friendly format.

2.5. The Concept of Framework and its Representation

As concisely explained by (Hevner, et al., March 2004), the major output of an IS research is an artifact which is intended to address an organizational problem. These IT artifacts can broadly have classified as constructs, methods, models and instantiations.

According to (Hevner, et al., March 2004), constructs represent vocabulary and symbols; models represent abstractions and representations; methods incorporates algorithms and practices; instantiations stands for implemented and prototype systems. Framework is a model artifact that provides a broad overview or skeleton of interlinked items which helps as a guide to achieve a specific objective.

2.5.1. Big Data Framework

Frameworks provide structure. The core objective of the Big Data Framework is to provide a structure for enterprise organizations that aim to benefit from the potential of Big Data. In order to achieve long-term success, Big Data is more than just the combination of skilled people and technology – it requires structure and capabilities. (Big Data Framework, 2018)

Many organizations struggle to embed a successful Big Data practice in their organization. The structure provided by the Big Data Framework provides an approach for organizations that takes into account all organizational capabilities of a successful Big Data practice. (Big Data Framework, 2018)

2.5.2. Why Big Data Framework is important?

Big Data is currently one of the most demanded niches in the development and supplement of enterprise software. The high popularity of Big Data technologies is a phenomenon provoked by the rapid and constant growth of data volumes. (Ilyukha, 2020) The main benefits of applying a Big Data framework include: The Big Data Framework provides a structure for organizations that want to start with Big Data or aim to develop their Big Data capabilities further, the Big

Data Framework includes all organizational aspects that should be taken into account in a Big Data organization, the Big Data Framework is vendor independent.

It can be applied to any organization regardless of choice of technology, specialization or tools, the Big Data Framework provides a common reference model that can be used across departmental functions or country boundaries, the Big Data Framework identifies core and measurable capabilities in each of its six domains so that the organization can develop over time. (Big Data Framework, 2018)

2.5.3. Top Big Data frameworks

Nowadays, there's probably no single Big Data software that wouldn't be able to process enormous volumes of data. Special Big Data frameworks have been created to implement and support the functionality of such software. They help rapidly process and structure huge chunks of real-time data. (Ilyukha, 2020)

There are many great Big Data tools on the market right now and some of the top listed frameworks are the following; (Ilyukha, 2020)

Hadoop: Apache Hadoop was a revolutionary solution for Big Data storage and processing at its time. Most of Big Data software is either built around or compliant with Hadoop. It's an open-source project from the Apache Software Foundation.

Hadoop is great for reliable, scalable, distributed calculations. However, it can also be exploited as common-purpose file storage. It can store and process petabytes of data. This solution consists of three key components:

- HDFS file system, responsible for the storage of data in the Hadoop cluster;
- MapReduce system, intended to process large volumes of data in a cluster;
- YARN, a core that handles resource management.

Hadoop can store and process many petabytes of info, while the fastest processes in Hadoop only take a few seconds to operate. It also forbids any edits to the data, already stored in the HDFS system during the processing.

Spark: It's an open-source framework, created as a more advanced solution, compared to Apache Hadoop. The initial framework was explicitly built for working with Big Data. The main difference between these two solutions is a data retrieval model.

Hadoop saves data on the hard drive along with each step of the MapReduce algorithm. While Spark implements all operations, using the random-access memory. Due to this, Spark shows a speedy performance, and it allows to process massive data flows. The functional pillars and main features of Spark are high performance and fail-safety.

It supports four languages:

- Scala;
- Java;
- Python;
- R.

It has five components: the core and four libraries that optimize interaction with Big Data. Spark SQL is one of the four dedicated framework libraries that is used for structured data processing. Using DataFrames and solving of Hadoop Hive requests up to 100 times faster.

Spark has one of the best AI implementation in the industry with Sparkling Water 2.3.0. Spark also features Streaming tool for the processing of the thread-specific data in real-time. In reality, this tool is more of a micro-batch processor rather than a stream processor, and benchmarks prove as much.

Spark is often considered as a real-time alternative to Hadoop. It can be, but as with all components in the Hadoop ecosystem, it can be used together with Hadoop and other prominent Big Data Frameworks.

Hive: Apache Hive was created by Facebook to combine the scalability of one of the most popular Big Data frameworks. It is an engine that turns SQL-requests into chains of MapReduce tasks.

- The engine includes such components as:
- Parser (that sorts the incoming SQL-requests);

- Optimizer (that optimizes the requests for more efficiency);
- Executor (that launches tasks in the MapReduce framework).

Hive can be integrated with Hadoop (as a server part) for the analysis of large data volumes. it remains one of the most used Big data analytics frameworks ten years after the initial release.

Storm: Apache Storm is another prominent solution, focused on working with a large real-time data flow. The key features of Storm are scalability and prompt restoring ability after downtime. You can work with this solution with the help of Java, as well as Python, Ruby, and Fancy.

Storm features several elements that make it significantly different from analogs. The first one is Tuple a key data representation element that supports serialization. Then there is Stream that includes the scheme of naming fields in the Tuple. Spout receives data from external sources, forms the Tuple out of them, and sends them to the Stream.

There is also Bolt, a data processor, and Topology, a package of elements with the description of their interrelation. When combined, all these elements help developers to manage large flows of unstructured data.

Speaking of performance, Storm provides better latency than both Flink and Spark. However, it has worse throughput. Recently Twitter (Storm's leading proponent) moved to a new framework Heron. Storm is still used by big companies like Yelp, Yahoo!, Alibaba, and some others.

Flink: Apache Flink is a robust Big Data processing framework for stream and batch processing. First conceived as a part of a scientific experiment around 2008, it went open source around 2014. It has been gaining popularity ever since.

It's a choice for simplifying an architecture where both streaming and batch processing is required. It can extract timestamps from the steamed data to create a more accurate time estimate and better framing of streamed data analysis. It also has a machine learning implementation ability.

As a part of the Hadoop ecosystem, it can be integrated into existing architecture without any hassle. It has the legacy of integration with MapReduce and Storm so that you can run your existing applications on it. It has good scalability for Big Data.

Flink is a good fit for designing event-driven apps and also has connectivity with a popular data visualization tool Zeppelin.

2.6. Awareness of Big Data

We are in a flood of data today. In a broad range of application areas, data is being collected at unprecedented scale. Decisions that previously were based on guesswork, or on painstakingly constructed models of reality, can now be made based on the data itself. Such Big Data analysis now drives nearly every aspect of our modern society, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences. (Fahim, 2017)

There are three ways to use big data analytics in a corporate strategy: as a core product or service offering (e.g., Cloudera), as a supporting product or service (e.g., Progressive), or as a key organizational capability. If an organization is employing one of the first two approaches, the use of big data analytics should be conspicuous. If they are trying to build a big-data-driven organization, that message may not be so obvious. In this case, organizational change management is extremely important to the strategy, and it begins with a very basic step that's often taken for granted: awareness. (Weathington, 2014)

Many projects fail, not for technical reasons, but because the natural resistance to change of stakeholders has been treated as a secondary factor, or even ignored. Big Data does not escape this observation. Its newness and relative complexity only increase the risk of failure. (agiledss, 2020) If an organization is using big data analytics as a key capability to support its corporate strategy, their first order of business is to ensure the entire organization is aware of what they are doing. (Weathington, 2014)

(Weathington, 2014) suggests, the best way to measure awareness is with a corporate scorecard. A scorecard is a highly-visible and accessible dashBoArd (e.g., front page on the corporate intranet) that clearly communicates strategic intentions, easily gives everyone immediate access to salient strategic metrics, and subtly (if not overtly) communicates the value placed on metrics and analysis.

2.7. Related Works

In this section, the study describes the available literature on best practice related to big data management frameworks development.

Unfortunately, local researcher on the assessment of framework of big data management implementation for different sectors including banking sector could not be found.

A study conducted in Oman, which aims to explore big data governance frameworks using literature review method, showed different big data governance frameworks, their shortcomings and proposes a new framework which has eight components. The study reveals there are very few studies on Big Data governance frameworks and the objective of the study was proposing a new framework. These components of the framework are the organizations' structure identification, stakeholders' identification, identifying the scope of Big Data, setting the policies and standards, optimizing and computing, measuring and monitoring quality, data storage, communication and data management. Big Data governance must produce quality output. To achieve this quality data, the data must be cleaned before analysis. Preparation and analysis are very important for quality data. The data is stored in a secured location, while at the same time being accessible when required. Finally, the outputs are communicated and delivered to the clients. (Al-Badia, Ali Tarhinia, & Khana, 2018)

A study by (Yang, et al, 2019) proposes a framework of big data governance that guides organizations to make better data-informed business decisions within the related regularity framework, paying close attention to data security, privacy, and accessibility. They started their study by presenting the challenges of the existing data governance mechanisms to handle big data management and some of the challenges listed are the traditional governance models are unable to handle structured and unstructured big data, insufficient knowledge on big data by the business managers, big data security and privacy, organizations resistance to use data in goal setting and decision-making, Lack of required tools to generate insight from the data in timely manner, shortage of the required skilled people on big data and then they proceed to proposing the framework by using literature review methodology. The researchers insist that data governance for big data environment must focus on three areas namely on Data architecture, Data quality and Data security, privacy, and ethical considerations.

Based on this theme they proposed a big data governance framework for cyber security which has three domains (Planning, Governance and Evaluation) and thirteen components.

A study conducted by (MBALUKA, 2013) using descriptive survey on 43 commercial banks and 1 mortgage finance company in Kenya, which focused on the extent of big data management in commercial banks, benefits of big data management, challenges in big data management and the effects of big data management on business value has revealed that banking sector in Kenya is in the very early stages of the big data management initiatives and banks are using various big data management techniques and tools.

The other finding of the study is banks with big data management in operation shows that they already have hands-on experience with big data techniques and tools.

From the study the researcher finds out that it is obvious that five key business areas that present low-risk opportunities for tangible performance using big data include: fraud detection, consumer sentiment, intelligent forecasting, customer profiling and target marketing, and customer service. (MBALUKA, 2013) and finally the researcher expects over the next five years, there will be notable and significant progress toward big data deployments within financial services in Kenya.

The main research gaps for BoA regarding big data analytics are the lack of studies on its specific application in the Ethiopian context, the limited research on its effectiveness in improving local banking challenges like fraud and customer service, and the absence of comparative analyses between Abyssinia Bank and other banks in the region regarding data-driven strategies. Research is needed to assess factors specific to its operating environment, such as local regulatory hurdles, infrastructure limitations, and the need for employee skill development for big data implementation.

Table 2 Summary of related works:

Authors	Objective	Methodology	Findings
Al-Badia, Ali Tarhinia, & Khana, 2018 **(Exploring Big Data Governance Frameworks)**	To explore big data governance frameworks, their shortcomings and proposes a new framework	literature review	Proposes a new framework which has eight components
Yang, Li, Elisa, Prickett, & Chao, 2019 **(Towards Big data Governance in Cybersecurity)**	To assess the challenges of the existing data governance mechanism for big data management	literature review	Proposed a big data governance framework for cyber security which has three domains and thirteen components
MBALUKA, 2013 **(BIG DATA MANAGEMENT AND BUSINESS VALUE IN THE COMMERCIAL BANKING SECTOR IN KENYA)**	To evaluate the influence of big data management on business value in commercial banks	Descriptive survey	The banks are at the very early stages of the big data management initiatives

2.8. Conceptual Framework of the Study

The conceptual framework of this research is derived from the related works and different studies conducted in the area of big data management. From the related works it has clearly seen that corporate strategy of an organization, policy, technology, people, regulations and competitions from the external environment are the main enablers of BDM. Moreover (Vossen, 2014) has depicted the various perspectives of big data as technological dimension, legal dimension, economical dimension and organization dimension. Another study by (Lineth Rodríguez, Mihalis Giannakis, , & Catherine da Cunha, 2018) which investigates the enablers of Big Data Analytics on sustainable supply chain has described the enablers, as the factors that motivate the adoption of sustainable supply chain analytics and they classified them as internal and external drivers.

The internal drivers include the commitment of top management, involvement of employees, resources, expertise and training of people, whereas the external drivers are related to the global context in which a firm works and the drivers include competition, regulatory, society/public demands and customer demand.

For this study purpose, the conceptual framework has four dimensions; Strategy, Technology, Environment and People, these dimensions are developed by benchmarking the above works discussed and by modifying to the study objectives and by integrating related attributes under the same category. The arrows in the diagram show interactions between the variables. As depicted in the diagram big data management is impacted by strategically, technological, environmental and people aspect of an organization. The conceptual framework guides the study to gather the relevant data based on these four dimensions.

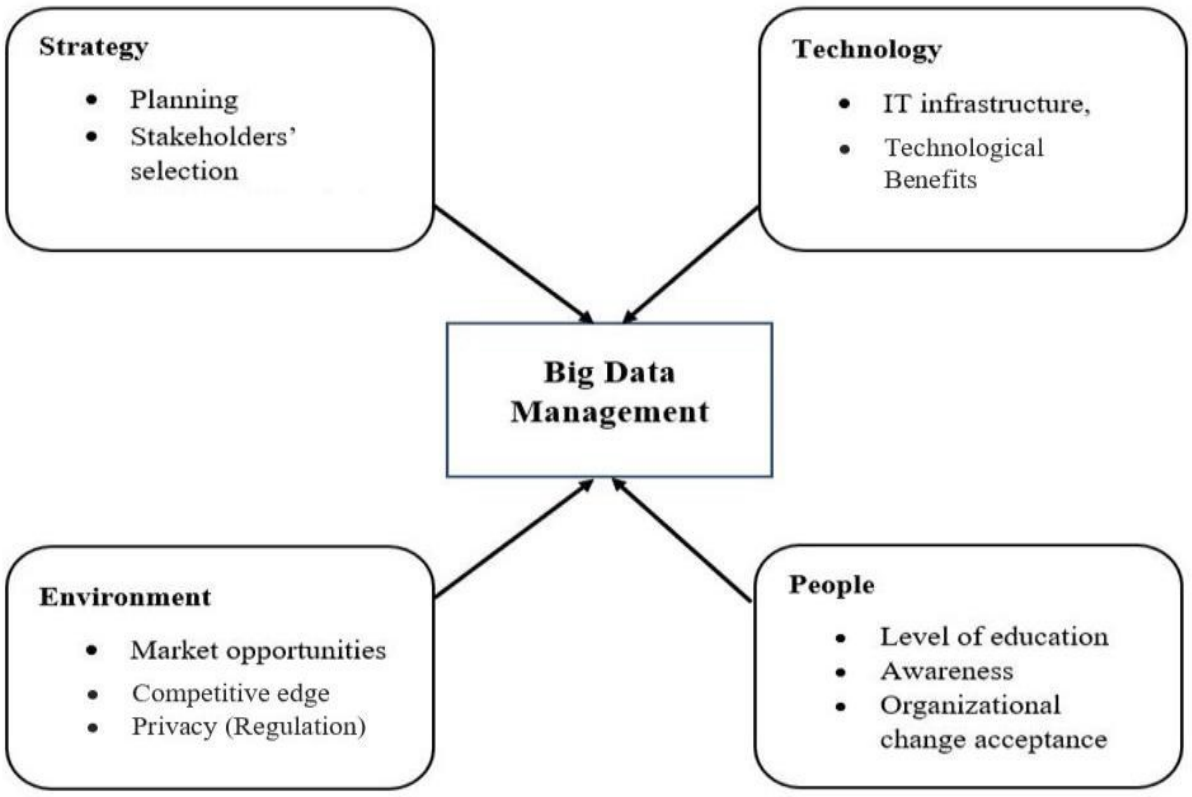


Figure 3 Conceptual framework to analyze the determinants of Big data management

CHAPTER THREE

3. RESEARCH METHODOLOGY

3.1. INTRODUCTION

This research has attempted to identify Big Data management related issues in general and specifically at BOA context. In the previous chapter, the concept of Big Data and Big Data management and its usefulness were discussed briefly. Although to get a holistic view multiple sources of data were used. This chapter explains the methodology that will be used in the study. This chapter will focus on the research design, data collection and data analysis methods followed.

The study follows the Information Tecnology Design Research (ITDR) approach. The method is applied to assess the current big data management practices, success factors, problems that impede on the implementation of big data management in the bank and to design the proposed big data management framework for the BOA Bank S.C.

The purpose of this study is to explore and examine the current big data management practices at BOA and to propose a solution for the big data management task.

The study was conducted using survey questionnaire and interviews as a method of data collection and design science research methodology as a research paradigm.

The collected data is used to investigate the current big data management practices carried out within the bank, gathering requirements from the employees for the new big data management framework and exploring its benefits. The output of the survey is an input for design science research methodology. Based on the gathered relevant information through survey which is an input for design science research methodology, the design science research method is used to design a framework for the big data management that is believed to be helpful for the banks' big data management tasks.

The below section will address the method followed for this research in detail.

3.2. Design Science Research Methodology

Within the field of the information Technology most studies have two paradigms. These are the paradigm of behavioral science, and the paradigm of design science.

The paradigm for behavioral science seeks to develop and validate theories to explain or predict human or organizational behavior. The philosophy of design science aims to expand the limits of human and organizational capability by making new and creative objects. Both paradigms are foundational to the IS discipline, positioned as it is at the confluence of people, organizations, and technology. In the design-science paradigm, knowledge and understanding of a problem domain and its solution are achieved in the building and application of the designed artifact. (Hevner, Salvatore T. March, Jinsoo Park, & Sudha Ram , March 2004)

The methodology of design science research (DSRM) integrates the principles, methods and procedures needed to perform such research and meets three objectives: it is consistent with previous literature, it provides a nominal process model for conducting DS research, and it provides a mental model for presenting and assessing DS research in IS. (Peffers, Tuure Tuunanen, Marcus A. Rothenberger, & Samir , 2007-8)

3.2.1. Design Science Research Process Models

According to (Peffers, et al., 2006), for IS researchers in which Design Science research offers an important paradigm for conducting applicable, yet rigorous, research, i.e., research that is closer to IS's applied raison d'être. A conceptual process could help researchers with a conceptual process for successfully carrying out Design Science research and a mental model for its presentation. For the research community, including editors and reviewers, such a process and mental model could help them to recognize such research and to respect its objectives, processes, and outputs.

Different researchers have adopted different design science research process models. In this study, the researcher compares and contrasts the design science research processes and adopted the appropriate relevant process model that is fitted to the research context.

Table 3 shows different design science research process via the researchers based on the common elements of the design science research methodology respectively with their time sequence.

Table 3 Design science process elements from IS other disciplines and synthesis objectives

Objectives for a design science research process model	Archer (1984)	(Takeda et al. 1990)	Eekels and Roozenburg (1991)	Nunamaker et al (1991)	Walls et al (1992)	(Rossi et al. 2003)	(Hevner et al. 2004)
1. Problem identification and motivation	Programming Data collection	Problem enumeration	Analysis	Construct a conceptual framework	Meta-requirements Kernel theories	Identify a need	Important and relevant problems
2. Objectives of a solution			Requirements				Implicit in "relevance"
3. Design and development	Analysis Synthesis Development	Suggestion Development	Synthesis, Tentative design proposals	Develop a system architecture Analyze and design the system. Build the system	Design method Meta design	Build	Iterative search process Artifact
4. Demonstration			Simulation, Conditional prediction	Experiment, observe, and evaluate the system			
5. Evaluation		Confirmatory evaluation	Evaluation, Decision, Definite design		Testable design process/product hypotheses	Evaluate	Evaluate
6. Communication	Communication						Communication

According to (Peffers, et al., 2006) of the above comparison of the design science research process models, the researcher has selected the guidelines of (Hevner, et al., March 2004). It is because of the suitability of the model that is appropriately fit this research context. The researcher can simply understand the guidelines and know how to apply those guidelines when and where he will use and in which circumstance of the research context.
The main reason behind to use (Hevner, et al., March 2004) design science research process model is that they established seven guidelines in order to assist researchers, reviewers, editors, and readers to understand the requirements for effective design-science research.

(Hevner, et al., March 2004) design science research process model also provides a suitable guideline to use anywhere in the research project by the researcher free judgment. Design science is inherently a problem-solving process. The fundamental principle of design-science research from (Hevner, et al., March 2004) seven guidelines are derived is that knowledge and understanding of a design problem and its solution are acquired in the building and application of an artifact. That is, design-science research requires the creation of an innovative, purposeful artifact.

Therefore, this study adopted the research procedures from (Hevner, et al., March 2004) in order to design the big data management framework for BOA. as shown in Figure 4.

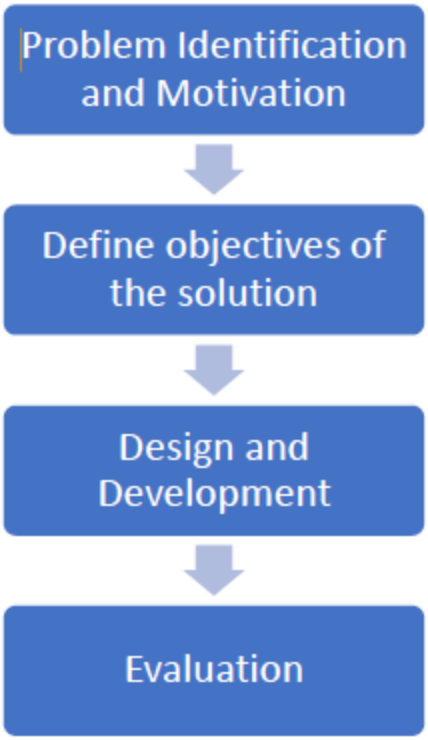


Table 4 DSR process adopted from (Hevner, et al., March 2004)

3.3. Problem Identification and Motivation

In this stage the study briefly defined the specific study problem and justifies the value of the solution. In defining the study problem, developed an artifact that can effectively provide a solution. The value of the solution also justified motivating the researcher and the audience of the research to pursue the solution, to accept the results, and to understand the reasoning associated with the researcher's understanding of the problem. These activities include knowledge of the state of the problem and the importance of its solution.

3.4. Source of Data and Data Collection Technique

For conducting this research, the researcher used descriptive survey method. This method was selected for this particular study because it was found an appropriate technique for collecting vast information and opinion from respondents. It is also relevant to gather detail description of existing condition and practices of data management in the bank.

Problem Identification and Motivation

3.4.1. Source of Data

The primary data sources used in this study are IT department staffs including VP of IT, IT managers and who have decision power related to data management and also IT officers. This is because, IT departments manage all the information systems functionalities including the database while the database experts or system administrators make sure that the systems are functioning as per the required policy, procedures, bank's requirement, etc., for the collection of the primary data the researcher uses interview and questioners.

In addition, secondary sources of data such as relevant best practices in big data management policy, standard and procedure documents, articles and the internet are reviewed.

3.4.2. Instruments of Data Collection

Generally, two types of instruments, namely: questionnaire and interview were employed for the data collection. The primary data was collected through questionnaire (structured) and interview (semi-structured).

3.4.3. Questionnaire

A questionnaire was designed based on the four categories; Strategy, Technology, People and Environment. The questionnaire was adapted from related works and mainly from (Marta, 2019) and some modification was done in line with the objectives of the research project. The questionnaire was prepared in English; the self-administered questionnaire was distributed and collected by the investigator.

The questions items are closed ended and are focused on practices and status of big data Management process. The questioners were prepared and distributed to IT staffs working other IT professionals at BOA in Oromia Special Zone Kemissie,Wolledi, Chefa and Bati Twon.

The questionnaire developed for IT staffs had 37 questions in five categories. The first section dealt with the socio demographic attributes of the respondents; the second section dealt with strategy of the bank towards big data management.

The third section inquired about the technological readiness of the bank on big data technologies, the fourth section deals with the people aspect of the bank on which the questions assess the staff's awareness regarding big data management and the last section dealt with the business environment of the bank currently operating.

The measuring scale for each question is a five-point Likert scale, i.e. Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree

3.4.4. Interview

The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. (Cothari, 2004) The interview is an important data gathering technique involving verbal communication between the researcher and the subject. Interviews are commonly used in survey designs and in exploratory and descriptive studies. (Nigel Mathers, Nick Fox, , & Amanda Hunn , 2002)

There are different types of interviews, which includes structured interview, unstructured interview and semi structured interview. Semi-structured interviews involve a series of open-ended questions based on the topic areas the researcher wants to cover. (Nigel Mathers, Nick Fox, , & Amanda Hunn , 2002) Accordingly, this activity was undertaken with purposely selected sample, key officials and experts working in IT department who have different responsibility, expertise and experiences about data management. Thus, interview guides which comprises of semi structured questions were developed. Open-ended questions are used in organizational research to explore, explain, or reconfirm existing ideas. (Dean C. Vitale, Achilles A. Armenakis, & Hubert S. Feild, 2008)

The main purpose of the interview session is to supplement and increase the validity and reliability of the information obtained through the questionnaire. The interview questions focused on the current practice of big data management, requirements for big data management, awareness level of big data technologies and problems that impede application of big data technologies at BOA.

The main purpose of the interview session is to supplement and increase the validity and reliability of the information obtained through the questionnaire.

The interview questions focused on the current practice of big data management, requirements for big data management, awareness level of big data technologies and problems that impede application of big data technologies at Choose a Sampling Technique for Research, 2016) Accordingly different IT staffs including IT support staffs, e-banking officers, network admins, system admins and database managers in the bank were selected for survey questionnaire and interviews. The selected IT experts were included for the survey questionnaire and interview because it is believed that those people have a good exposure with data(big) management tasks. Twenty-two individuals were selected as respondents for the survey questionnaire and three experts are selected for the interview.

Table 5 Target population for survey questionnaire

Respondents	Total population	Target population
IT support staffs	13	9
Database officers	4	4
E-banking officers	5	2
Network admins	4	2
System admins	3	3
Data warehouse and BI officers	3	2
Total Target population		22

Table 6 Target population for interview

Respondents	Total population	Target population
Manager	1	1
Manager	1	1
VP	1	1
Total Target population		3

3.5. Define the Objectives of the Solution

In this stage the study aims to identify the requirements for developing the proposed big data management framework from knowledge of the state of problems and current solutions.

The researcher infers the objectives of a solution from the problem definition and knowledge of what is possible and feasible to align with the bank ‘s requirements. The qualitative objectives are used, which described how a proposed framework is expected to support solutions to problems not previously addressed. The researcher used the framework to inferred objectives rationally from the problem specification. In order to develop the proposed framework, researcher started with the literature review to gather relevant requirements and aspects of existing big data management frameworks.

After data was collected from respondents, it should be narrated and summarized. It involves data preparation, analysis, and finally data interpretation. Qualitative and quantitative data analysis method is used to analyze the data collected from the questionnaire and interview. The analysis technique used in this study is thematic coding analysis technique in which the result of interview was grouped in to main categories. Quantitative data mostly collected through survey questionnaire is analyzed using descriptive statistics techniques. Finally, appropriate generalization is made and presented accordingly for the collected data by way of narrating and interpreting the situations.

3.6. Design and Development

The conceptual models of DSRM must produce a viable artifact in the form of models, methods or frameworks. This stage determines the desired functionality and its architecture and then creating the actual artifact. The knowledge of theory that can be brought to bear in a solution is required to move from objective of a solution to design and development of the proposed framework. A framework is a model artifact that provides a broad overview or skeleton of interlinked items which helps as a guide to achieve the specific objective. After all the above stages are completed the appropriate and suitable framework was developed based on previously developed framework with possible modification by determining the functionality and each component to solve the existing problems at BOA. The software used for developing the framework is E-draw-Max. E-draw Max is selected because the software has rich features and easy to understand.

Based on the required knowledge of a solution identified in stage one (problem identification and motivation) and stage two (defining the objective of the solution) with the literature review, questionnaires and expert interviews analysis to develop the proposed framework with the process selected. After determining the big data management framework architecture, the proposed big data management framework for BOA is developed. The developed framework shows each component with its functionalities to solve the existed problems at the bank.

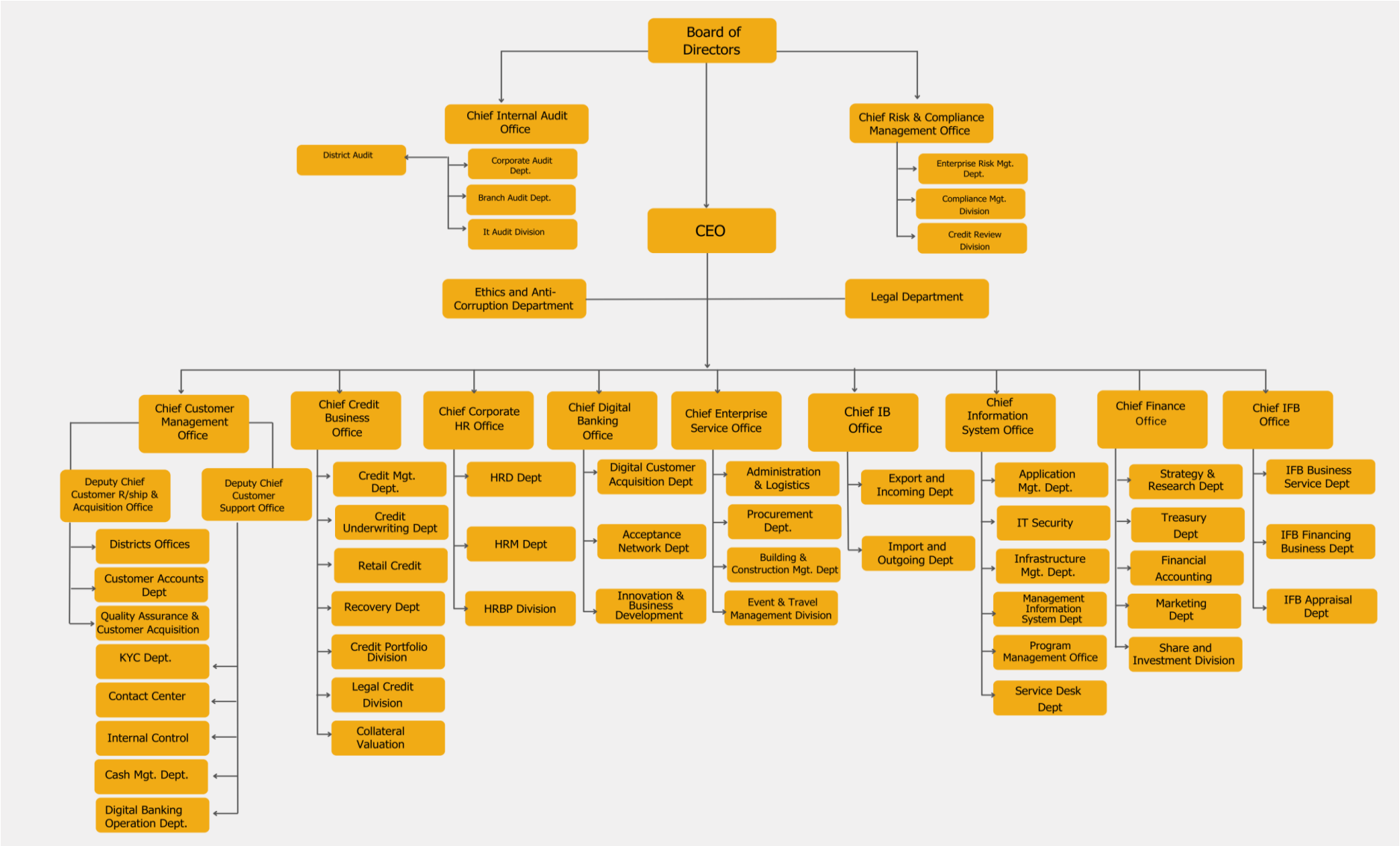


Figure 4:- The Organizational Design of BoA

3.7. Evaluation

This phase aims to observe and measure how well the framework supports a solution to the problem. Depending on the nature of the problem venue and the artifact, evaluation could take many forms, including comparison of artifact ‘s functionality with the solution objectives this study took the qualitative evaluation analysis techniques which include a comparison of the proposed framework functionality with the solution objectives from activity two above, objective qualitative performance measures, such as business impacts, and business continuity extents of use and user satisfaction. The proposed big data management framework will be evaluated by experts who have a direct relation with information technology infrastructure, and as an additional evaluation mechanism the proposed framework is compared with the ISO data governance framework.

3.8. Validity and Reliability

The most widely used measures of quality of research are reliability and validity. During selection of instruments of data collection like questionnaire and interview, detail discussions was made with both bank officers and some researchers. A pilot test is conducted with five staffs before distributing the questionnaire to all the respondents, accordingly, ambiguous or vague wording was avoided to ensure that respondents would read and answer the question consistently on different occasions in the same context. Additionally, constructs and items of the instruments were adopted after deep analysis of literature was conducted and modified to the context of the study. So, the reliability and validity of the study was considered from crosschecking of data from different sources.

Reliability is one of the characteristics and /or quality of measure of constructs. “Reliability concerns the extent to which a measurement of a phenomenon provides stable and consist result (Carmines and Zeller, 1979). Reliability is also concerned with repeatability. For example, a scale or test is said to be reliable if repeat measurement made by it under constant conditions will give the same result (Moser and Kalton, 1989)” (Taherdoost, 2016). There are many types of reliability: internal consistency reliability, test retest reliability, internal reliability (Cronbach’s alpha) and parallel form reliability. For this study internal reliability (Cronbach’s alpha) is used

to justify reliability and internal consistency of the questionnaire.

Table 7 Reliability of the questionnaire (Cronbach's alpha)

Reliability Statistics	
Cronbach's Alpha	N of Items
.931	32

As we can see from the above table, the Cronbach's alpha (internal consistency) of the questionnaire displayed for the 32 items of the questionnaire was 0.931, which indicates high internal reliability of items of the questionnaire which also indicate the survey is reliable since it is greater than 0.7.

3.9. Data Analysis

Data analysis involves critical thinking. In this research the data analysis is done after collecting all data from the respondents. Thus, the analysis of the study followed the objective of the research. Mixed method data analysis involves qualitative and quantitative data analysis that are combined, connected, or integrated. In view of this, the data analysis began after organizing the data collected through primary data collection methods. The data gathered through interview are analyzed and interpreted thematically based on the objective of the research. Quantitative data mostly collected through survey questionnaire is analyzed using descriptive statistics techniques. Frequency, average and percentage values are used for discussion of the data collected from the survey.

CHAPTER FOUR

4. DATA PRESENTATION AND ANALYSIS

This section describes results on Big data management practices at BOA. In this chapter, both quantitative and qualitative data that is obtained from survey and interview is presented and analyzed. Quantitative data collected from the respondents using structured questionnaire were organized in to tables and figures. Qualitative data obtained from interview is presented and analyzed with each respective quantitative response. To verify or support the validity of the responses, qualitative and quantitative data has been also triangulated with its corresponding conceptual and empirical data. lastly the summary of the results will be presented. It is based on the questionnaire and interview results that the framework was developed.

4.1. Data Analysis

Survey questions was prepared by adopting from literatures and modified according to the researcher's case. To undertake the survey for the study, a total of 22 questionnaires were distributed. Of the total distributed questionnaires, 21 (95.5%) were complete and returned back for analysis. Interview questions were also prepared by adopting from literatures and modified according to the researcher's case. The main purpose of the interview session is to supplement and increase the validity and reliability of the information obtained through the questionnaire. The interview questions focused on the current practice of big data management, requirements for big data management, awareness level of big data technologies and problems that impede application of big data technologies at BoA.

The interview was conducted by the researcher after relevant questions to the objective of the study were prepared. The participants of the interview are the CIO (VP) of the bank, the data warehouse and business intelligence manager and the IT systems manager. An interview finding is broadly categorized in to 4 sections. This includes: strategy, technology, environment and people aspect towards big data management.

4.1.1. Socio-Demographic Characteristics of Respondents

Among 21 respondents 13(61.9%) were males and the remaining 8(38.1%) are females. From the total, 18(85.7%) had first degree, 3(14.3%) had second degree. In terms of job position, there are 9(42.9 %) IT application analysts, 2(9.6%) of them are digital channel admins, 2(9.6%) of them are Data warehouse and BI officers, 3(14.3%) are database admins, 3(14.3%) are system admins and 2(9.6%) of them are network admins. Most of the respondents 9(42.9%) had worked for less than or equal to 5 years.

Table 8 Socio Demographic characteristics of the respondent n= 21

Characteristics	Frequency	Percent	Cumulative %
Gender			
✓ Male	13	61.9	61.9
✓ Female	8	38.1	100.0
Educational level			
✓ First degree	18	85.7	85.7
✓ Post graduate	3	14.3	100.0
Job Position			
✓ Data warehouse and BI manager	✓ 1	✓ 4.8	✓ 4.8
✓ Data warehouse and BI officer	✓ 1	✓ 4.8	✓ 9.5
✓ Database administrator	✓ 1	✓ 4.8	✓ 14.3
✓ Digital Channels administrator	✓ 1	✓ 4.8	✓ 19.0
✓ IT Application Analyst	✓ 6	✓ 28.6	✓ 47.6
✓ Network Administrator	✓ 1	✓ 4.8	✓ 52.4
✓ Principal Digital Channel Admin	✓ 1	✓ 4.8	✓ 57.1
✓ Principal IT application analyst	✓ 1	✓ 4.8	✓ 61.9
✓ Sr Network administrator	✓ 1	✓ 4.8	✓ 66.7
✓ Sr, System Administrator	✓ 1	✓ 4.8	✓ 71.4
✓ Sr. Database administrator	✓ 2	✓ 9.5	✓ 81.0
✓ Sr. IT Application Analyst	✓ 2	✓ 9.5	✓ 90.5
✓ System Administrator	✓ 2	✓ 9.5	✓ 100.0

Experience ✓ Less than 5 year ✓ From 5-10 years ✓ From 11-15 years ✓ More than 15 years	9 7 2 3	42.9 33.3 9.5 14.3	42.9 76.2 85.7 100.0

4.1.2. Big Data

Big data is relatively a new and emerging technology, in order to assess whether the respondents are familiar with topic they are asked whether the heard about big data. From the total respondents most of them 19(90.5%) have heard about big data.

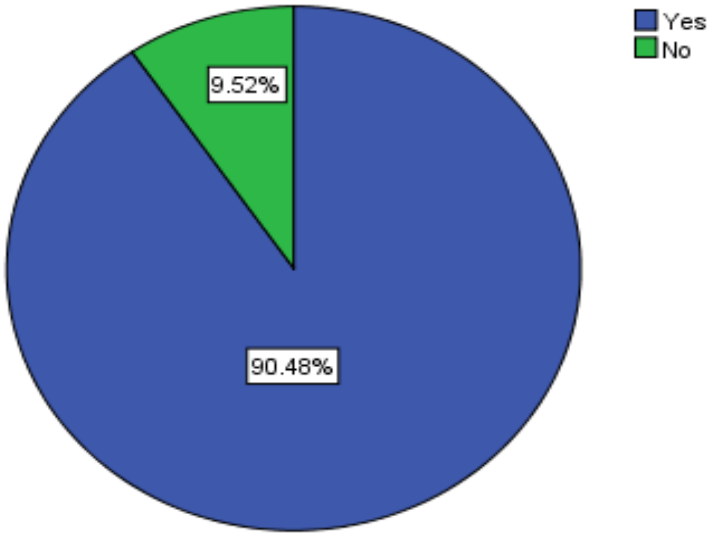


Figure 5 Respondents answers on whether they have heard about big data, BoA, Kemissie, Cheffa, Bati and Wolledi Amhara, Ethiopia 2024, n=21

4.1.3. Strategy

The second part of the survey begins by asking about the strategy of the bank on big data technology. The aim of these questions is to assess the strategically direction and vision of the bank towards big data management. The questions are directed towards assessing the plan of the bank on big data technology, the stakeholder's selection and hiring of the bank.

The strategy plays a great role towards big data management, without prior planning and preparedness it will be a difficult task working towards big data technology. Keeping in mind this notion, eight (8) questions are provided for the respondents.

Planning is the preliminary task for undertaking different activities. Regarding planning and vision of the bank towards big data management the respondents said that the bank had plans and visions for data analytics and business intelligence (BI) tasks and also they have said the bank is doing a road map for data analytics and the bank had a concerned directorate (MIS) for this task.

The results on table 8 indicate 43% of the respondents agree and 19% of the respondents even strongly agree on the bank's had a vision and mission about implementing big data technology. Therefore, 62% of the respondents believed on the bank have a strategy towards implementing big data technology.

In relation to staffing of ICT experts, the results show that 43% of the respondents agree on the bank's hiring strategy of ICT experts to promote the strength and efficiency of the bank, and 28.6 % are neutral. Regarding security capabilities of the bank for adopting big data technology, the respondents, from the interview session, said that the bank have data security policy, controls and governance schemes so, they believe that the bank have security capability to adopt the technology. In addition, they believe the tools for the technology have also security capabilities' and adopting the technology without securing it is impossible.

Overall, the mean distribution of responses found to be 2.34 which is near to the value of 'Agree' on the strategy of the bank. Thus, more than half of the respondents in the survey have believed that bank had strategies in place for implementation of big data technologies.

Table 9 Strategy towards big data management, n= 21

Strategy	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Mean Score**
	N (%)	N (%)	N (%)	N (%)	N (%)	
The bank has a clear mission, vision about implementing Big Data technology	4 (19.0%)	9 (42.9%)	6 (28.6%)	1 (4.8%)	1 (4.8%)	2.33

The bank policies are in place to promote and manage use of Big Data Management in the organization	4 (19.0%)	9 (42.9%)	6 (28.6%)	1 (4.8%)	1 (4.8%)	2.33
The bank has a direction, commitment and plan towards Big Data Management	3 (14.3%)	9 (42.9%)	6 (28.6%)	2 (9.5%)	1 (4.8%)	2.48
The bank has future development plan for basic ICT and information infrastructure to implement Big Data Management	3 (14.3%)	11 (52.4%)	6 (28.6%)	1 (4.8%)	0	2.23
The bank implements strategic plan related to ICT provisioning/supply and support	4 (19.0%)	14 (66.7%)	2 (9.5%)	1 (4.8%)	0	2
The bank has future development plan and respond towards the dynamic ICT advancement	3 (14.3%)	15 (71.4%)	2 (9.5%)	0	1 (4.8%)	2.09
The bank's ICT HR strategy related to hiring experienced and professional experts' strength the efficiency of the bank.	1 (4.8%)	9 (42.9%)	6 (28.6%)	4 (19.0%)	1 (4.8%)	2.76
The bank feel it is a strategic necessity to practice big data management to compete in the market place	2 (9.5%)	10 (9.5%)	6 (28.6%)	2 (9.5%)	1 (4.8%)	2.52
Total Mean Score						**2.34**

4.1.4. Technology

Big data encompasses broad range of technology and process that enable a big data technology to operate and provide services; the aim of this variable is used to assess the technological readiness and the IT infrastructure of the bank. The ten (10) questions incorporated on the technology section are aimed at assessing the benefit of the technology, the bank's IT infrastructure and the security capabilities to adopt the technology.

As shown on table 9 all the respondents agreed on big data technology is advantageous for the bank, 57.1% 'Strongly Agree' and 42.9% select 'Agree'. Twenty (95.3%) respondents believed using big data technology helps the bank in providing valuable information for decision making and improved operation.

Most of the respondents 42.9% were neutral on answering big data technology is a complex technology or not.

28.6% of the respondents strongly agree and 47.6% of the respondents agree on big data technology will be compatible with the bank's existing IT infrastructure and IT service management.

The respondents answered big data technology, if adopted, can and must be compatible with the bank's current operating practices and infrastructure, and they also add big data technology needs an additional infrastructure other than the existing one but the current data can be an input for big data technology.

Regarding to the approaches and tools used for managing the banks' data the respondents answered, the bank is currently using relational databases which are oracle and SQL database servers. In addition, there are database security tools, data redundancy management tool and they said the bank's had disaster recovery site and they said these all tools are helpful for managing the banks' data by maintaining its availability, confidentiality and consistency.

The answer to the question '*Did you believe the tools and approaches currently used for data management are sufficient for yielding the big data utilization benefits?*' were 'No' by the interview respondents and the reason was the tools currently used in the bank lack the analytics

and interpretation ability. The current tools only provide a flat data, so they will not yield the benefits of big data management.

33.3% of the respondents strongly agree and 52.4% of the respondents agree on big data technology will be compatible with the data captured at the bank and if the bank adopts the technology, most of the respondents 80.9% believed the bank will have the security capabilities.

Table 10 Technological factors at BOA, n= 21

Technology	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Mean Score**
	N (%)	N (%)	N (%)	N (%)	N (%)	
Do you agree Big Data Management/analytics is advantageous for your bank?	12 (57.1%)	9 (42.9%)	0	0	0	1.42
Using Big Data Management/analytics would provide my company with valuable information for decision making	11 (52.4%)	9 (42.9%)	1 (4.8%)	0	0	1.52
Using Big Data Management/analytics would improve the effectiveness of the bank operations and performance	10 (47.6%)	11 (52.4%)	0	0	0	1.52
Do you think Big Data Management/analytics is a complex technology to adopt?	2 (9.5%)	4 (19.0%)	9 (42.9%)	4 (19.0%)	2 (9.5%)	3
The bank quality of internet connections is suitable	5 (23.8%)	9 (42.9%)	3 (14.3%)	3 (14.3%)	1 (4.8%)	2.33
Big Data Management/analytics will be compatible with my company's existing IT infrastructure and IT service management	6 (28.6%)	10 (47.6%)	2 (9.5%)	2 (9.5%)	1 (4.8%)	2.14
The technology will be compatible with the data captured at the bank	7 (33.3%)	11 (52.4%)	2 (9.5%)	0	1 (4.8%)	1.9

Do you think the existing ICT would help in storing Big Data a better way	5 (23.8%)	9 (42.9%)	3 (14.3%)	3 (14.3%)	1 (4.8%)	2.33
The hardware and software required for the building Big Data Management/analytics is readily affordable	2 (9.5%)	7 (33.3%)	3 (14.3%)	1 (4.8%)	0	2.52
The bank has security capabilities to adopt this technology	2 (9.5%)	15 (71.4%)	3 (14.3%)	1 (4.8%)	0	2.14
Total Mean Score						**2.08**

The respondents believed big data technology will be helpful in knowing the banks' customers in a better way.

"*To know our customer's better big data technology will be very helpful other than questionnaire or other methods because we have the real data of the customer so then the technology can help us to better know our customers by using these data*" said one of the respondents.

The respondents answered the bank had no previous big data management framework and the reasons that hinder from building the framework are; lack of knowledge, lack of skill, lack of exposure to the technology and also, they trend is very new and emerging but now the MIS directorate is established and the technical unit of the directorate will be responsible to the future development and they are studying and benchmarking international trends.

As can be seen from the above table, the aggregate mean (mean of the mean) result for the technological dimension is found to be 2.08, which is rated in the "Agree" category. Majority of the respondents agreed with the points raised on the survey.

4.1.5. Environment

The aim of this component is to assess the opportunities and challenges of big data management in BoA in relation the bank itself and the current market environment.

The questions are directed towards the organizational structure, competitive edge, regulations and market opportunities for adopting and exercising big data management.

As shown in table 10, 95.3% of the respondents answered adoption of big data technology will give the bank a competitive advantage, 42.9% of the respondents agree and 33.3% of respondents were neutral to the statement of 'it is easy for the bank to comply with privacy related regulation to implement this technology'.

The respondents asserted that to be highly competitive and a customer's choice in the market place, the bank feel it is a strategic necessity to use big data technology.

In addition, the interview respondents describe the benefits of big data management as follows, for making data driven decision making, for forecasting of profit and loss, for market analysis, for setting investment direction, for service excellence, for proposing new product and for trend analysis.

52.4% of the respondents agreed the bank can adopt the technology given the resources, opportunities and knowledge currently available in the bank.

The mean score on whether the bank is focusing on big data opportunities, but is not yet planning is 2.80, which is nearly neutral, so it is assumed the respondents have no information.

Table 11 Environmental factor for BDM at BoA, 2024, n= 21

Environment	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean Score
	N (%)	N (%)	N (%)	N (%)	N (%)	
Adoption of big data will give the bank a competitive advantage	9 (42.9%)	11 (52.4%)	1 (4.8%)	0	0	1.61
The bank closely follows the competitor's big data initiatives to determine its	1 (4.8%)	7 (33.3%)	8 (38.1%)	4 (19.0%)	1 (4.8%)	2.85

strength and weakness						
The telecommunications infrastructure is reliable and efficient in the bank	0	16 (76.2%)	4 (19.0%)	1 (4.8%)	0	2.28
There are firms in the market who provides support for use of this technology	2 (9.5%)	8 (38.1%)	11 (52.4%)	0	0	2.42
It is not difficult for the bank to protect data privacy for big data initiatives	4 (19.0%)	8 (38.1%)	6 (28.6%)	3 (14.3%)	0	2.38
It is easy for the bank to comply with privacy related regulation to implement this technology	1 (4.8%)	9 (42.9%)	7 (33.3%)	4 (19.0%)	0	2.66
Given the necessary resources, opportunities and knowledge to use big data, the bank will be ready to adopt it	2 (9.5%)	11 (52.4%)	6 (28.6%)	1 (4.8%)	1 (4.8%)	2.42
The bank has begun to focus on big data opportunities, but are not yet planning	0	7 (33.3%)	12 (57.1%)	1 (4.8%)	1 (4.8%)	2.80
		Total Mean Score				**2.42**

As can be seen from the above table, the aggregate mean (mean of the mean) result for the environment dimension is found to be 2.42, which is rated in the "Agree" category. Majority of the respondents agreed with the points raised on the survey.

4.1.6. People

The aim of these questions is to assess the level of education and awareness of the employees on big data management and also to assess the organization change acceptance behavior of the employees.

Most of the respondents 57.1% are neutral on the staff's level of education and training for big data in the bank, in addition 47.6% are also neutral on the banks' program to train the staffs on big data management/analytics.

66.7% were neutral and 19% have disagreed the staffs have knowledge on big data management/analytics.

The respondents believed that the staffs have no sufficient knowledge about big data technology.

One of the interview respondents said "*may be the higher officials have information about big data technology, but not more than concept level and I think that's why they have decided for the road map of data analytics decision"*

The respondents said the bank has programs in place to train the staffs, the bank will prepare capacity building programs like workshops and trainings and they are thinking about what kind of trainings have to be given for the staffs. They also said that they are planning to use other local banks' experience and other international organizations experiences as a benchmark for the capacity building scheme.

Regarding the employees' acceptance of organizational change, the respondents said that "we *can't say there is no resistance but the majority of employees have accepted the changes made in recent years, like the core banking system change".*

As the respondents reported, in the bank currently there is no big data analyst or expert. And also, they said that they don't think there is an expert in the current Ethiopian market.

The mean score 2.42 which is nearly agree indicates that previous organizational change is accepted by majority of the employees.

Table 12 **People aspect towards big data management, n=21**

People	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Mean Score**
	N (%)	N (%)	N (%)	N (%)	N (%)	
The level of training and education for ICT and Big Data in the bank is adequate.	0	1 (4.8%)	12 (57.1%)	5 (23.8%)	3 (14.3%)	3.47
The bank has programs in place to train the users for Big Data Management/analytics	0	1 (4.8%)	10 (47.6%)	8 (38.1%)	2 (9.5%)	3.52
Our employees have basic knowledge about Big Data Management/analytics	0	3 (14.3 %)	14 (66.7%)	4 (19.0%)	0	3.04
Our IT employees have the ability to quickly learn and apply new technologies	3 (14.3%)	13 (61.9 %)	4 (19.0%)	1 (4.8%)	0	2.14

The bank has technical and managerial skills on the use of technological innovation	3 (14.3%)	10 (47.6 %)	5 (23.8%)	2 (9.5%)	1 (4.8%)	2.42
Previous organizational change has been accepted by the majority of employees	4 (19.0%)	6 (28.6 %)	42 (42.9%)	2 (9.5%)	0	2.42
			Total Mean Score			**2.83**

As can be seen from the above table, the aggregate mean (mean of the mean) result for the people dimension is found to be 2.83, which is rated in the "Neutral" category. Majority of respondents hold a moderate stand in their opinions.

4.2. Summary

The major objective of this study is, to assess current Big Data management approaches at BoA, and to develop Big Data management framework and decision making for the bank.

The study tried to investigate the current state of the bank in relation to big data management. The researcher has sought there is a promising progress towards adoption of data management/analytics technologies and as it is seen in the data analysis section the bank has a concerned directorate for this purpose. The department will be responsible for preparing the data, the infrastructure and creating the knowledge for the data management/analytics task.

The higher officials of the bank had an understanding of big data technology and they have made a decision to open a directorate of MIS for handling of the data analytics task and they believe moving to data analytics scheme will benefit the bank in many ways. The decision made by those higher officials is very important to the future data driven business process because without their decision it will be hard to achieve those benefits of data driven business profit and loss, for market analysis, for setting investment direction, for service excellence, for proposing new product and for trend analysis.

The awareness and knowledge level of the employees to big data technology can be expressed as low, but as it is stated in the discussion the bank has programs in place to train the staffs and create the knowledge.

The staffs organizational change acceptance behavior can be expressed as moderate as 42.9% of the respondents were neutral when answering the question '*Previous organizational change has been accepted by the majority of employees*.' On the other hand, as it is discussed on the interview section changes carried out previously in the bank had been accepted by majority of the employees and this is very helpful the banks' plan of adopting data analytics technology.

As it is discussed in data analysis section the bank needs an additional infrastructure to utilize the benefits of big data management, the data resides in the current tools can be an input for big data technology but these existing tools cannot be used for data analytics.

Currently there is no big data management framework at the bank and the reasons that hinder the bank from building the framework are; lack of knowledge, lack of skill and lack of exposure to the technology.

The researcher comes to understand that the bank is now at an early stage of starting data analytics practices. The commitment of the higher officials, the strategically direction of the bank, the technological readiness, the business environment and the people's knowledge and change acceptance behavior will together have a greater role for the success of the data analytics/data management initiative.

All interviewees have agreed on the main point of building a big data management framework and they believe it is essential as a solution to have a data driven decision making, for forecasting of profit and loss, for market analysis, for setting investment direction, for service excellence, for proposing new product and for trend analysis.

The awareness and knowledge level of the employees to big data technology can be expressed as low, but as it is stated in the discussion the bank has programs in place to train the staffs and create the knowledge.

The staffs organizational change acceptance behavior can be expressed as moderate as 42.9% of the respondents were neutral when answering the question '*Previous organizational change has been accepted by the majority of employees*.' On the other hand, as it is discussed on the interview section changes carried out previously in the bank had been accepted by majority of the employees and this is very helpful the banks' plan of adopting data analytics technology.

As it is discussed in data analysis section the bank needs an additional infrastructure to utilize the benefits of big data management, the data resides in the current tools can be an input for big data technology but these existing tools cannot be used for data analytics.

Currently there is no big data management framework at the bank and the reasons that hinder the bank from building the framework are; lack of knowledge, lack of skill and lack of exposure to the technology.

The researcher comes to understand that the bank is now at an early stage of starting data analytics practices. The commitment of the higher officials, the strategically direction of the bank, the technological readiness, the business environment and the people's knowledge and change acceptance behavior will together have a greater role for the success of the data analytics/data management initiative.

Development and Evaluation

Abyssinia Bank's development and evaluation using big data analytics is focused on improving credit risk management and operational efficiency. The bank uses data mining algorithms to predict non-performing loans, enhance customer service through personalization, and manage risks more effectively, although there are challenges in data collection and implementation.

Development through big data analytics

- ✓ **Credit risk prediction**:

  Abyssinia Bank applies data mining techniques, such as J48 Decision Trees and Naïve Bayes, to historical data to predict customers who are likely to be loan defaulters.

- ✓ **Operational efficiency**:

  The bank uses big data to identify patterns in customer behavior to improve its operational efficiency and make more informed decisions.

- ✓ **E-banking and personalization**:
  Abyssinia Bank uses data from its various e-banking services to understand customer needs and provide personalized services, as seen with its Gize-Pay mobile money product.

Evaluation of big data analytics at Abyssinia Bank

- ✓ **Positive outcomes**:
  - Enhanced ability to identify and mitigate credit risks by predicting loan defaulters.
  - Improved customer satisfaction through personalized services and product suggestions.
  - Stronger competitiveness through better data-driven decision-making.
- ✓ **Challenges**:
  - Data collection and processing can be challenging, with limitations on the amount of data available for analysis, as seen in a study that was limited to four years of data.
  - Implementing big data analytics across all departments can be a challenge.
  - A study found that the bank struggled with negative feedback on its mobile app, indicating a need for better user experience design based on data analysis.

CHAPTER FIVE

5. THE PROPOSED BIG DATA MANAGEMENT FRAMEWORK

In this chapter, the proposed Big Data management framework will be presented. The proposed framework is based on the reuse of existing framework found in the literature and it is an improved Big data management framework to meet the objective and to answer the questions of this research. Furthermore, this chapter addressed the evaluation of the proposed framework in order to ensure its applicability, efficiency, effectiveness and usefulness. The framework is designed based on the findings of the study and related literatures in order to provide an answer to the third research question. In the first section, the proposed framework is presented and the evaluation of the proposed framework is discussed in the second part of the chapter.

5.1. Framework Development

As it is seen in the course of the study, the bank had at least 23 years of data incorporated in its storages and this vast amount data is not utilized in a productive way, the researcher believed the framework proposed will be helpful in managing this incorporated data in a value adding way.

The framework is developed based on best experiences from various literatures and possible modifications. Since the framework development follows a design science approach, some of the components of the framework are taken from existing frameworks and modified to fit in to BOA context. Additional components that are specific to the banking context are also added based on the requirements identified in the data analysis. As per Hevner et al (2004), in IS design-science research the expected output to be created is a purposeful IT artifact that addresses an important organizational issue. (Hevner, et al., March 2004) Therefore, on this research the framework which is proposed to BOA is a purposeful IT artifact that can address Big Data management related issues and helps the bank in efficient utilization of its data and it paves the way for the successful transition of data driven business. This framework is based on the reuse of the existing big data management framework, (Yang, et al, 2019) has developed a Big data governance framework that focuses on high level recommendation for Big Data Governance in Cyber security and the framework specifically address big data governance for network security. The researcher improves the existing (Yang, et al., 2019) proposed framework to (banking business) BoA's context and practice.

According to the questionnaire and the interview result the proposed framework is customized to BoA context and modifications are made on the adopted framework, components on the adopted framework domains are added and also minimized, six components namely stakeholders' selection, big data scope determination, awareness creation, infrastructure development, data cleaning and banking big data standards system are added to the adopted/previous framework and the minimized components are working mechanisms clarification, governance domain determination and metadata management, the researcher believed these components are addressed in the proposed framework components.

5.2. Components of the Framework

In this study, the designed framework is named '*An Integrated Big Data Management Framework for BOA* or '*DI-BDAF-EDM_BOA*' in short. This framework designed based on the findings from the study and E-draw max software is used for designing of the framework.

The framework has three domains and sixteen components, in the following section the domain and the components of the framework will be discussed briefly.

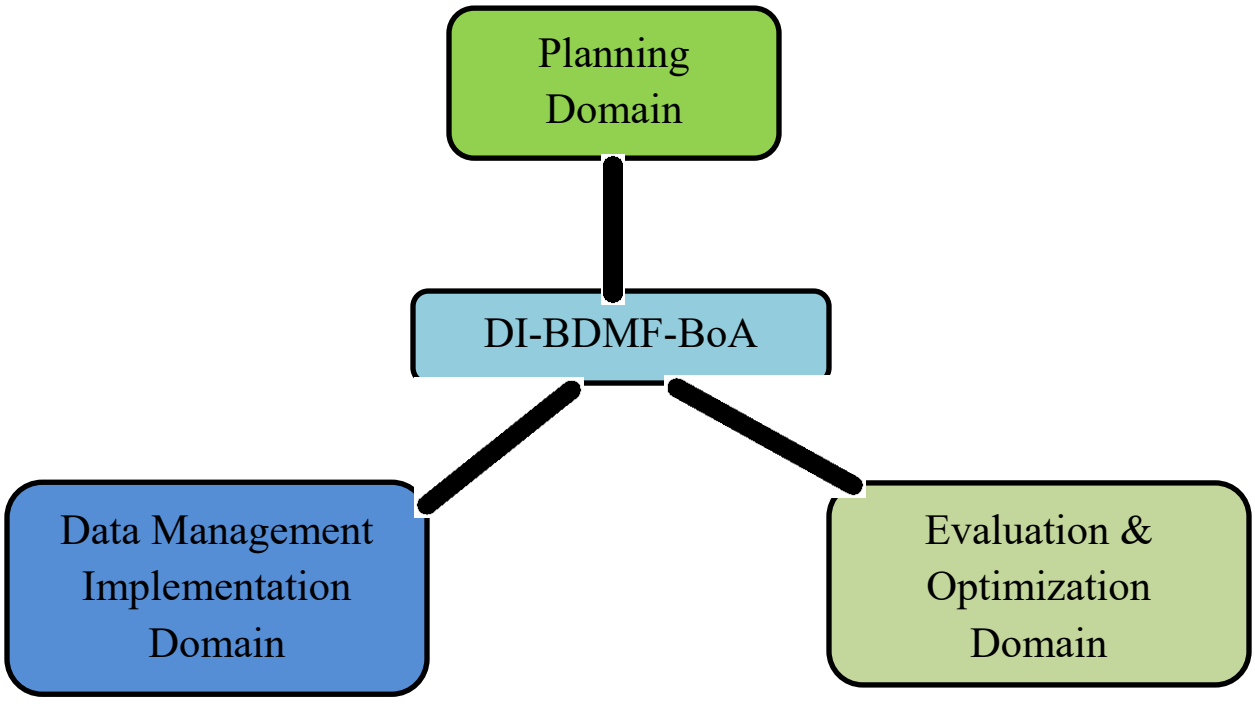


Figure 6 DI-BDAF-EDM_BoA Domains

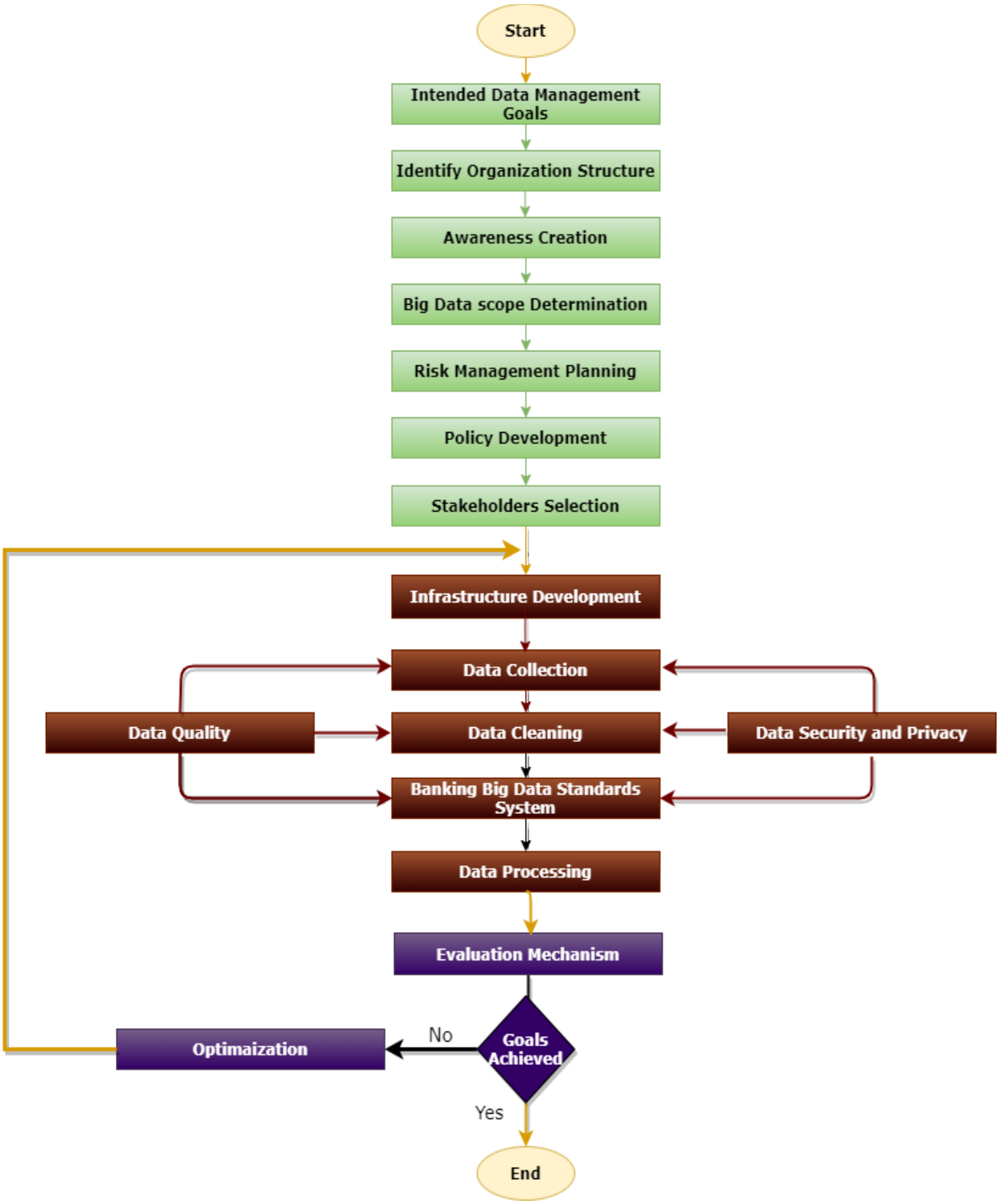


Figure 7 An Integrated Big Data Management Framework for BoA.

5.2.1. Planning Domain

The application of the proposed data management framework for BoA needs to be carefully planned to meet the strategic data-related goals of the bank. The planning encompasses assessing the current environment of the system and the bank's key data related operations.

As discussed earlier, Big data application had so many promises for the banking industry as an example, utilizing the banks data will help in identifying Customer spending patterns, Transaction channel identification, Fraud management & prevention, Risk assessment, compliance & reporting and the like, to meet these goals through big data analytics a proper and well documented data management planning is crucial, through this plan possible data risks during the data management process can be identified and accordingly, corresponding procedures, rules, and principles, such as data storage procedures, data usage principles, ethical procedures, can be developed, to address the identified risks. At the same time, the strategies, tools, and mechanisms in implementing the Big data management project are also determined and developed.

The planning domain is helpful for establishing the standards to minimize the risks and maximize the values of the big data and its utilization.

For the proposed DI-BDAF-EDM_BOA framework the planning domain had seven components (Intended Data Management Goals, identify organizations structure, Awareness creation, Risk Management Planning, Policy Development, Stakeholders selection, and Big data scope determination) and each component will be discussed below.

Intended Data Management Goals:- The proposed big data management framework permits the bank to secure enormous information while making most extreme value from the enormous information.

Big data management goals are defined by assessing the current environment of the system and the organization's key performance indicators (Yang, et al., 2019). For the case of this study initially the goal of the proposed framework is to show the way on how to manage the data stored in the bank for decades and these helps BoA in making better decisions based on their huge data and is to make these vast amount of data ready for big data analytics.

The main goal of big data management according to (Luna, 2013) is to ensure a high level of data quality and accessibility for business intelligence and big data analytics applications. For the future years this objective will be on the banks' key target, thus, the bank will be working towards achieving this objective.

Identify Organizations Structure:- Big Data management should be in accordance with the objectives, missions and vision of the organization (Al-Badia, Ali Tarhinia, & Khana, 2018). Therefore, identifying the organization structure has been taken as a component of the planning domain.

The organization and its structure influence the Big Data governance decisions. The organization's structure requires enhanced study.

Awareness Creation:-- awareness creation among the staffs of the bank is crucial for the success of the goal of data management. Because many projects are failing not for technical reasons rather the natural resistance to change among stakeholders'. And in many projects, this factor is either ignored or counted as a secondary factor. (Agiledss, 2020)

Since the technology of big data is relatively new to the country and to the bank specifically, the bank must invest on big data training and workshops for their employees to get the desired result. As it is discussed in the previous chapter the bank has plans to train its staffs on data analytics technologies. The bank can train its staffs on the following courses; Big Data Hadoop and Spark Developer, data science, MongoDB Developer and Administrator, Business Analytics with Excel, Machine Learning, and other related courses. (Simplilearn)

By creating awareness for the staffs, all the employees will understand the value of big data and they will work towards attaining the intended data management goals.

Big Data Scope Determination:- The scope of the big data management must be defined prior to implementing the big data management task, because the framework or the approach will be failed because of inconsistencies between the scope and the organizations infrastructure. Many problems arise with Big Data due to inadequate technologies to process it effectively. (Al-Badia, Ali Tarhinia, & Khana, 2018)

This scope determination encompasses the following aspects; budget for the project, schedule of the project, staffing procedures and architecture (technical resources).

The scope determination must also assess whether the project affects operations, product development and revenue of the organization.

Risk Management Planning:- Based on the identified goal, potential data risks during the big data management procedure can be recognized; and likewise, corresponding techniques, rules, and standards, for example, data storage methods, information utilization standards can be created, to address the distinguished risks. (Yang, et al., 2019)

Policy Development:- The policy development stage addresses the issues related to data capture, management, consumption, privacy, security, risk, retention, regulatory compliance and data classification (Al-Badia, Ali Tarhinia, & Khana, 2018), at this stage all the procedures, rules, and principles should be framed. As an example, data access rule can be set by defining user access level, or by classifying what type of data will be granted for external and internal users and so on. (Yang, et al., 2019).

Stakeholder's Selection:- Stakeholder's are the actors on the big data management scheme, so it is crucial to critically identify those actors.

The stakeholders include data scientists, data analysts, business steward leads, data stewards, steering committee, etc. (Al-Badia, Ali Tarhinia, & Khana, 2018) Among these stake holders' data analysts and scientists are crucial, a data analyst has reporting-oriented profile, having experience in extracting and analyzing data from traditional data warehouses using SQL. Their tasks are normally either on the side of data storage or in reporting general business results. Data warehousing is by no means simple, it is just different to what a data scientist does. (The Tutorials Point)

Many organizations struggle hard to find competent data scientists in the market. It is however a good idea to select prospective data analysts and teach them the relevant skills to become a data scientist. This is by no means a trivial task and would normally involve the person doing a master degree in a quantitative field, but it is definitely a viable option.

The basic skills a competent data analyst must have are listed below (The Tutorials Point)

- Business understanding
- SQL programming
- Report design and implementation
- DashBoArd development

The role of a data scientist is normally associated with tasks such as predictive modeling, developing segmentation algorithms, recommender systems, A/B testing frameworks and often working with raw unstructured data. (The Tutorials Point)

The nature of their work demands a deep understanding of mathematics, applied statistics and programming. There are a few skills common between a data analyst and a data scientist, for example, the ability to query databases. Both analyze data, but the decision of a data scientist can have a greater impact in an organization. (The Tutorials Point) Here is a set of skills data scientists normally need to have (The Tutorials Point)

- Programming in a statistical package such as: R, Python, SAS, SPSS, or Julia
- Able to clean, extract, and explore data from different sources
- Research, design, and implementation of statistical models
- Deep statistical, mathematical, and computer science knowledge

By selecting the right personnel through a pre-defined selection mechanism, BOA must hire the right persons for the operation of the big data management.

5.2.2. Data Management Implementation Domain

The Data Management implementation domain describes the data management objectives that BOA should focus on whilst conducting data management activities, and on the proposed framework this domain has seven components namely Infrastructure development, data collection, data cleaning, data quality, data security and privacy, banking big data standards system and data processing.

Infrastructure Development:- a successful implementation of big data management relies heavily on the infrastructure an organization deploys.

Infrastructure is the cornerstone of Big Data architecture. Possessing the right tools for storing, processing and analyzing the data is crucial in any Big Data project. (McNulty, 2014)

The infrastructure needs to be a cohesive, programmable infrastructure that can scale as workloads demand. It must ensure the timely application of analytics and the secure transfer, storage and management of data. (Chard, 2017)

Most big data implementations need to be highly available, so the networks, servers, and physical storage must be resilient and redundant. Resiliency and redundancy are interrelated. An infrastructure, or a system, is resilient to failure or changes when sufficient redundant resources are in place ready to jump into action. Resiliency helps to eliminate single points of failure in the infrastructure. (Hurwitz, Alan Nugent, Fern Halper, & Kaufman)

The commonly known big data tools are HADOOP and NoSQL, Hadoop is essentially an open-source framework for processing, storing and analyzing data. Hadoop is a whole ecosystem of different products, largely presided over by the Apache Software Foundation. Some key components include: (McNulty, 2014)

- **HDFS**:-The default storage layer
- **MapReduce**- Executes a wide range of analytic functions by analyzing datasets in parallel before 'reducing' the results. The "Map" job distributes a query to different nodes, and the "Reduce" gathers the results and resolves them into a single value.
- **YARN**- Responsible for cluster management and scheduling user applications

- **Spark**- Used on top of HDFS, and promises speeds up to 100 times faster than the two-step MapReduce function in certain applications. Allows data to loaded in-memory and queried repeatedly, making it particularly apt for machine learning algorithms

NoSQL, which stands for Not Only SQL, is a term used to cover a range of different database technologies, NoSQL databases are adept at processing dynamic, semi-structured data with low latency, making them better tailored to a Big Data environment. (McNulty, 2014)

The different strengths and uses of Hadoop and NoSQL are often described as “operational” and “analytical”. NoSQL is better suited for “operational” tasks; interactive workloads based on selective criteria where data can be processed in near real-time. Hadoop is better suited to high-throughput, in-depth analysis in retrospect, where the majority or all of the data is harnessed. (McNulty, 2014)

For the context of this study BOA must be equipped with the modern data management infrastructure and technology because the traditional data administration tools are not sufficient for handling the big data management/analytics tasks.

Data Collection:- collecting data is the first step in this domain, the quality of the collected data will have a direct impact on the output of the framework, so these data must be collected carefully for a better result of the big data management task.

Mainly there are the following data’s in a bank; customer data and transaction data (of different formats). These data must be collected and stored carefully, then passed to pre-processing.

Data Cleaning: - data cleaning is the method of rectifying or evacuating inaccurate and corrupt data. This procedure is essential and underlined in light of the fact that wrong data can drive a business to wrong decisions, conclusions, and poor analysis. (Bhatt, 2020)

Irrelevant data usually includes duplicate records, missing or incorrect information and poorly formatted data sets. A business can expand this option furthermore by eliminating the data records that are not really necessary for certain business processes. While what gets filtered out depends on the discretion of the business, some basic points like outdated data or details that are not verified can be removed. (Bhatt, 2020)

Data Quality: - Data quality administration is applied to ensure the large data are of fitting esteem. For example, data from various assets may be clashing to one another, and along these lines, it is critical to settling the contention before the data utilized for decision making. (Yang, et al., 2019).

There are seven factors that play a huge role in determining data quality, which are accuracy, availability, completeness, granularity, relevance, reliability and timeliness (Ayswarrya).

Data Security and Privacy:- Data security and privacy, forms a prime concern when gathering, transmitting, storing, and using big data. Big data is regularly assembled from various sources, and usually include private data (Yang, et al., 2019). For example, in a bank context customer's information is a very sensitive data a breach of data security and privacy will have a significant negative result on the existence of the bank, so when doing the big data management tasks, the stake holders must give a due emphasis on the security and privacy of their data.

To ensure the quality of the data the bank can follow this steps and achieve the above characteristics of the data; first start with the source of data by making sure it is reliable, use data quality tools and by following best practices such as educating everyone within the bank on data quality, appointing roles such as data owners, data stewards and data custodians within your organization and establish proper processes to ensure high data quality and investigate quality problems at the source, just like we've mentioned above. (Ayswarrya).

Security is the practice of defending information and information assets through the use of technology, processes and training from: -Unauthorized access, Disclosure, Disruption, Modification, Inspection, Recording, and Destruction. (Priyank Jain, , Manasi Gyanchandani , & Nilay Khare, 2016)

Information privacy is the privilege to have some control over how the personal information is collected and used. Information privacy is the capacity of an individual or group to stop information about themselves from becoming known to people other than those they give the information to. One serious user privacy issue is the identification of personal information during transmission over the Internet. (Priyank Jain, , Manasi Gyanchandani , & Nilay Khare, 2016)

Data privacy is focused on the use and governance of individual data—things like setting up policies in place to ensure that consumers' personal information is being collected, shared and utilized in appropriate ways. Security concentrates more on protecting data from malicious attacks and the misuse of stolen data for profit. (Priyank Jain, , Manasi Gyanchandani , & Nilay Khare, 2016)

Banking Big Data Standards System:- An important challenge for big data is to integrate data from different sources, and standard applications have been proven to promote interoperability between systems (Al-Badia, Ali Tarhinia, & Khana, 2018), as we have seen through the course of the study, the data BOA is storing on their database are of different format, for the big data management task effectiveness, there must a unifying big data standard.

Having a standard is helpful for querying, for integration with other systems, security and governance.

Data standardization results from mapping the source data into a target structural representation. Customer name data provides a good example for a bank—names may be represented in thousands of semi structured forms, and a good standardizer will be able to parse the different components of a customer name (e.g., first name, middle name, last name, initials, titles, generational designations) and then rearrange those components into a canonical representation that other data services will be able to manipulate. (Loshin, 2009)

After the prior stages of these domain the banking big data standards system component will make the preprocessed data to the set standards and pass those standardized data to the next stage.

Data Processing:- The last component on this domain is the data processing, the processing helps to easy extraction of the raw data collected on the previous stages, for this purpose software framework like Hadoop can be used for the purpose of the data processing.

Two of the components data quality and data security and privacy are interrelated to the other components and they run across the domain, because ensuring the data quality and security is the milestone of the DI-BDAF-EDM_BOA..

The last domain on the proposed framework is the evaluation and the optimization domain, and on this domain, there are two components; evaluation mechanism and optimization namely.

These domain aims to evaluate the output of the previous two domains.

5.2.3. Evaluation and Optimization Domain

This component objective is testing the preprocessed data for some kind of decision making or initial analytics. For BOA context the preprocessed data can be fed into an artificial intelligence or some machine algorithm, for example for the prediction of loan return report or customer spending patterns, then if the evaluation yields the intended result we can say that the goal the DI-BDAF-EDM_BOA is met and we can proceed for the next stage which is big data analytics. If not, as it is shown on the framework, we will iterate back to the second domain to optimize the DI-BDAF-EDM_BOA.

5.3. Evaluation of the Proposed Framework

The following section discusses about the method of evaluation done on the proposed framework. According to (Hevner, et al., March 2004) framework is an output as model artifact from IS design science research that needs to be evaluated in order to prove its effectiveness, utility and efficiency. Therefore, the evaluation might be iterative to improve the framework and ensure the quality of the proposed solution so that it can solve real world business problems. An artifact can be evaluated by observational, analytical, experimental, testing, expert validation, and descriptive methods (Hevner, et al., March 2004). For this study purpose, the proposed big data management framework was evaluated by experts who have a direct relation with information technology infrastructure in addition; the proposed framework is compared to the ISO 8000 data governance framework. (Timothy, 2016), (Al-Badia, Ali Tarhinia, & Khana, 2018)

5.3.1. Expert Validation

In this study, expert validation was used to evaluate the proposed framework along with descriptive method. The expert validation was conducted by twelve IT experts through a questionnaire shown in Appendix C to gather individual evaluation of the improved framework regarding the completeness, clarity and correctness of the revised framework.

The content of the questionnaire is derived from the evaluation criteria recommended by (Hevner, et al., March 2004)which consists of fit to the organization, comprehensiveness, reliability, clarity, correctness, and usability quality attributes and the evaluation checklist is adopted from (TIGIST, 2018), because the researcher finds the checklist appropriate for this study.

In order to evaluate the consistency of the study, Cronbach ‘s alpha reliability test is applied and the result is depicted in table 12. The value of the coefficient alpha (0.741) indicates that the survey is reliable since it is greater than 0.7.

Table 13 Reliability Statistics for the Proposed BoA

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.741	.753	10

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.467	4.167	4.750	.583	1.140	.034	10

Table 14 Mean and standard deviation of the Framework Evaluation Survey

Descriptive Statistics

Items	N	Minimum	Maximum	Mean	Std. Deviation
The proposed framework is comprehensive in terms of coverage	12	4.00	5.00	4.5833	.51493
The organization and presentation of the framework is suitable for	12	4.00	5.00	4.4167	.51493

BoA					
The objective of the framework is clear	12	4.00	5.00	4.7500	.45227
The content of the proposed framework is complete	12	4.00	5.00	4.4167	.51493
The content of the proposed framework is relevant	12	4.00	5.00	4.5833	.51493
The content of the proposed framework is clear	12	4.00	5.00	4.5000	.52223
The content of the proposed framework is Scalable	12	3.00	5.00	4.2500	.86603
The proposed framework is easy to be applicable.	12	4.00	5.00	4.1667	.38925
The applicability of the proposed framework can improve Resource utilization, scalability, data management and yield data management benefits	12	4.00	5.00	4.6667	.49237
The implementation of the proposed framework fits with the organization problems	12	4.00	5.00	4.3333	.49237
Overall rating of the proposed framework	12	3.91	4.91	4.3106	.31122
Valid N (list wise)	**12**				

From the evaluation result we can see on table 11 the proposed framework is easy to be applicable scored the lowest mean value (4.1667) among all quality attributes used to measure the framework.

This indicates the need for further improvement of the framework to show the ways of implementing the framework, next to the above statement the lowest mean score is recorded on the following statement, 'the content of the proposed framework is scalable' has scored (4.2500) but the result revealed that majority of the evaluators agreed to the statement.

Similarly, the organization and presentation of the framework is suitable for BoA is agreed by majority of the experts with mean result of 4.4167.

The evaluation result showed that the objective of the framework is clear and the statement has the highest means value (4.7500) which revealed that majority of the experts strongly agreed to it.

The content of the proposed framework is relevant and clear as the evaluation result shows a mean of 4.5833 and 4.5 respectively, so we can conclude that the evaluators strongly agreed on clarity and relevance of the proposed framework. The applicability of the proposed framework can improve resource utilization, scalability, data management and yield data management benefits (4.6667) and the implementation of the proposed framework fits with the organization problems (4.3333) revealed the validity of the framework to be implemented at BOA.

The descriptive analysis (mean and standard deviation) of the survey result is computed and as it can be seen in table 12 the mean result of the evaluation variables is found to be greater than 3 which indicated that the respondents agreed on the clarity, completeness, usefulness, correctness of the proposed framework.

The overall rating of the proposed framework is 4.3106 which represent the category of 'Very Good'. This indicates that the IT experts participated on the evaluation survey confirmed completeness, correctness and clarity, applicability of the proposed framework. Based on the above analysis evaluation result proves how the proposed framework can improve resource utilization, scalability, data management and yield data management benefits at BoA. Moreover, based on the survey result, the easy applicability and the scalability of the framework can be considered as an improvement area to further develop the framework so that it can serve its intended goal.

5.3.2. Comparison with ISO 8000 Data Governance Framework

The ISO 8000 data governance framework has the following components; planning, process identification, data identification, stakeholder's identification, technology identification, conformance to requirements and implementation and evaluation. (Timothy, 2016), (Al-Badia, Ali Tarhinia, & Khana, 2018)

The objective of the comparison was to assess the proposed framework is developed in respect to the ISO 8000 components and to check for possible improvement opportunities.

The comparison shows the proposed framework met 87% of the ISO 8000 data governance framework standard.

The blue color shows the components on the ISO 8000 framework has been fully represented by the framework.

Table 15 Comparison of BoA with ISO 8000 data governance framework standard

ISO 8000 data governance framework	planning	process identification	data identification	Stakeholders identification	technology identification	conformance to requirements	implementation and evaluation	Summary (%)
DI-BDAF-BoA								87%

The yellow is a partial representation because the framework is evaluated but not implemented.

This research is believed to produce results that can improve the Big data management process at BoA. In this chapter the proposed framework of big data management is presented. This framework is believed to help BOA in many ways including paving the way for data driven business and utilizing the data stored in the bank's storages in a productive way.

CHAPTER SIX

6. CONCLUSION AND RECOMMENDATIONS

This chapter presents conclusions drawn from the study, some recommendations based on the evidences presented during the course of the study and also suggestions for future research.

6.1. Conclusions

In today's world change is happening everywhere, its speed and complexity are increasing, the technological advancement and the new ways of doing business are changing at a fast pace and the future success of any organizations depends on how successfully to respond to these changes and being able to cope up with these changes. So, to keep pace with the market organization must be creative and proactive for the ongoing changes.

One of the current advancements in today's world is the technology of big data analytics. The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. As we have seen through the course of the study big data utilization will impact the business of a bank in so many beneficiary ways.

In the near future using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world. Leaders of the banking business need to recognize the potential opportunity as well as the strategic threats that big data represent and should assess and then close any gap between their current IT capabilities and their data strategy and what is necessary to capture big data opportunities relevant to their organization.

The implementation of big data technologies is becoming increasingly essential for the banking sector. With the highly competitive market that we live in today, and growth and alignment of the emerging information technology with business, adding more value to the business processes such as big data management, can be a value adding for the organization and it helps them to be on the competition. Hence, investigating and addressing big data management issues is essential. Big data management help organizations on how to utilize organization's data in a productive way.

The Big Data management framework will have a great importance in the bank's decisions and policymaking. This study tries to assess the concept of big data management and previous big data management frameworks. For developing countries like Oromo Special Zone, Ahmhara Ethiopia the researcher assumes the concept is very new and still very little is known about it. There are few studies on the concept of big data. In this study, a framework for Big Data management has been proposed specially for BoA.

In this study, BoA is considered as a data driven business and the study tries to address the management of big data for the bank. This study tries to investigate the issues on the existing big data management process through quantitative and qualitative survey and design a framework for addressing the identified issues.

The research sought to determine multiple frameworks from different sectors and perspectives by conducting extensive literature review. In order to answer the research questions and achieve the objectives, the study tries best to assess the available literature in the topic and provided a detailed investigation and understanding of the big data management process in different businesses.

From the study the researcher has acknowledge that many scholars are striving to make a great contribution for the big data discipline, this effort results in enormous knowledge and expertise on the topic, but when we come to the case of our country and the organizations, the researcher assumes that, very little is known about the topic and the researcher concluded that still many studies have to be conducted on the coming years.

On the basis of the research findings, it is clearly seen than banks will be very beneficiary by exercising big data management practices. To explain, much of the literature on big data management reveals companies are harvesting the fruits of big data management and analytics.

The study produces a framework for BoA based on the data collected and the researcher assumes this framework is helpful for the bank. The process and the controls performed on each domain are clearly described, and the researcher tries to customize it for the banks' structure as much as possible. It is expected that the findings and the framework will provide a better understanding of big data management and this knowledge would provide a positive contribution to the understanding of how big data management is understood in BoA in Oromo special Zone.

6.2. Recommendations

The output of this study is assumed to provide guidance to managers and IT professionals planning to implement big data technologies to their organizations. This study provides some useful insight for the banks' IT managers who often need to take decisions with regard to big data technologies. The proposed framework also provides general guidelines and controls for big data management initiatives to structure an effective big data management process.

BoA and the other banks in Oromo Special Zone should be ready for big data technologies. Accordingly, the execution of the proposed framework is recommended in order to assess their readiness and to know their position for the future big data era.

Based on the study findings and the above conclusions the following recommendations are forwarded.

- ✓ The researcher suggests the bank should consider using the DI-BDAF_BoA for the management of its data.
- ✓ The researcher recommends improving the proposed framework iteratively by incorporating the ongoing business and technological advancements.
- ✓ The researcher recommends applying the framework in other different data driven sectors like health care, telecommunications, transportation energy and pharmaceuticals by customizing to their business context, and this would help in creating an understanding of the topic and may yield additional insights.

6.2.1. Suggestions for Future Research

The study conducts a framework comparison and expert feedbacks to evaluate the proposed framework, the researcher recommended to others to extend the framework with different evaluation methods, like user satisfaction assessment, for better output, in addition future researchers can do their study on exploring incompatibility between big data technologies and the underlining infrastructure of an organization.

References

Abyssinia Bank About Abyssinia. https://www.thereporterethiopia.com/42605/

Adam, K., A. Majid, M., & Jasni , Z. M. (2014). *Big Data Management and Analysis.*

agiledss. (2020). Raising awareness of Big Data. *https://www.agiledss.com/en/success-stories/raising-awareness-big-data*.

AK. (2018). Top 5 Advantages Of Big Data In Banking You Should Know. *https://acadgild.com/blog/big-data-in-banking*.

Al-Badia, A., Ali Tarhinia, & Khana, A. I. (2018). Exploring Big Data Governance Frameworks. *Procedia Computer Science 141*, 271-277.

Al-Shiakhli, S. (2019). Big Data Analytics: A Literature Review Perspective.

Ayswarrya, G. (n.d.). What is Data Quality and How to Ensure It in Your Organization. *https://humansofdata.atlan.com/2020/02/what-is-data-quality/*.

Bank of Abyssinia. https://www.bankofabyssinia.com/working-capital-financing/

Barnes, S. (2019). BANKING AND BIG DATA: THE PERFECT MATCH? *https://internationalbanker.com/banking/banking-and-big-data-the-perfect-match/*.

Bhatt, V. (2020). The Significance of Data Cleansing in Big Data. *https://www.aithority.com/guest-authors/the-significance-of-data-cleansing-in-big-data/*.

Big Data Framework. (2018). An overview of the Big Data Framework. *https://www.bigdataframework.org/an-overview-of-the-big-data-framework/*.

Blasiak, K. (2014). *Big Data; A Management Revolution.* Helsinki.

C.L. Philip Chen, & Chun-Yang Zhang. (2014). *Data-intensive applications, challenges, techniques.* Macau, China.

Chandani, A., Mita, M., Neeraja, B., & Om, P. (2015). *BANKING ON BIG DATA: A CASE STUDY.* Chennai, India.

Chard, I. (2017). 5 Reasons why infrastructure matters for Big Data success. *https://gblogs.cisco.com/uki/5-reasons-why-infrastructure-matters-for-big-data-success/*.

Cothari, C. (2004). Research Methodolgy: Methods and techniques. *New Age International (P) Limited, Publishers.*

Dean C. Vitale, Achilles A. Armenakis, & Hubert S. Feild. (2008). Integrating Qualitative and Quantitative Methods for Organizational Diagnosis: Possible Priming Effects? *Journal of Mixed Methods Research*, 87-105.

Egetoft, K. (2019). Data Management Challenges For Financial Services. *https://www.digitalistmag.com/customer-experience/2019/01/15/data-management-challenges-for-financial-services-06195118*.

Fahim, G. (2017). Awareness of Target Groups about Big Data Analytics and FutureDemand Estimation of Big Data Analytics. *https://www.academia.edu/*, 37.

Fedak, V. (2018). Big Data analytics in the banking sector. *https://medium.com/datadriveninvestor/big-data-analytics-in-the-banking-sector-b7cb98d27ed2*.

Gagnon, R. (2019). Your Go-to Guide to Big Data Analytics in Banking. *https://us.hitachi-solutions.com/blog/big-data-banking/*.

Gascoigne, C. (2019). Why aren't banks making the most of data? *https://www.raconteur.net/finance/data-management-banking*.

Hevner, A. R., Salvatore T. March, Jinsoo Park, & Sudha Ram . (March 2004). Design Science in information system research. *MIS Quarterly, Vol. 28*(No. 1), 75-105.

Hurwitz, J., Alan Nugent, Fern Halper, & Kaufman, M. (n.d.). Basics of Big Data Infrastructure. *https://www.dummies.com/programming/big-data/engineering/basics-of-big-data-infrastructure/*.

Ilyukha, V. (2020). 10 Best Big Data Tools for 2020. *https://jelvix.com/blog/top-5-big-data-frameworks*.

Laurila, M. (2017). *BIG DATA IN FINNISH FINANCIAL SERVICES.* Finland.

Legner, P. C. (2017). Master Thesis Topics. *https://wp.unil.ch/bisa/masters-thesis-topics/*.

Lineth Rodríguez, Mihalis Giannakis, , & Catherine da Cunha. (2018). Investigating the Enablers of Big Data Analytics on Sustainable Supply Chain. *IPSERA*.

Loshin, D. (2009). Data Standardization. *https://www.sciencedirect.com/topics/computer-science/data-standardization*.

Luna, M. (2013). big data management. *https://searchdatamanagement.techtarget.com/definition/big-data-management*.

Marta, G. (2019). *Assessment of Ethiopian banking sectors Readiness for the Implementation of big data analytics(unpublished).*

Mauricio. (2016). The role of big data in the banking industry. *https://bigdata-madesimple.com/role-big-data-banking-industry/*.

MBALUKA, W. (2013). BIG DATA MANAGEMENT AND BUSINESS VALUE IN THE COMMERCIAL BANKING SECTOR IN KENYA.

McNulty, E. (2014). Understanding Big Data: Infrastructure. *https://dataconomy.com/2014/06/understanding-big-data-infrastructure/*.

Mizuno, Y. (2015). Enhanced Data Management: A Key Competitive Advantage for Japanese Banks. *https://www.moodysanalytics.com/risk-perspectives-magazine/risk-data- management/approaches-to-implementation/enhancing-data-management-advantage-japanese-banks.*

Nigel Mathers, Nick Fox, , & Amanda Hunn . (2002). Using Interviews in a Research Project. *TRENT FOCUS GROUP*.

OCHIENG, G. F. (2015). *THE ADOPTION OF BIG DATA ANALYTICS BY SUPERMARKETS IN KISUMU COUNTY.* Nairobi.

Oracle. (2015). *Big Data in Financial Services and Banking.*

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

Peffers, K., Tuure Tuunanen, Charles E. Gengler, Matti Rossi, Wendy Hui, Ville Virtanen, & Johanna Bragge. (2006). THE DESIGN SCIENCE RESEARCH PROCESS: A MODEL FOR PRODUCING AND PRESENTING INFORMATION SYSTEMS RESEARCH. *DESRIST. Claremont, CA*.

Peffers, K., Tuure Tuunanen, Marcus A. Rothenberger, & Samir . (2007-8). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems,, 24*(3), 45-78.

Priyank Jain, , Manasi Gyanchandani , & Nilay Khare. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*.

Russom, P. (2013). *MANAGING BIG DATA.*

SAS. (2020). Data Management for Banking. *https://www.sas.com/en_sg/industry/banking/technology/data-management.html*.

simplilearn. (n.d.). Big Data Analytics Courses and Certification Training. *https://www.simplilearn.com/big-data-and-analytics/*.

Srivastava, U., & Gopalkrishnan, S. (2015). *Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks.*

Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. pp. 18-27.

Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management (IJARM)*, 28-36.

The financial brand. (2020, Apr). Best Practices Guide to Achieve Better Data Management in Banking. *https://thefinancialbrand.com/95562/bank-credit-union-data-management-best-practices/*.

The Tutorials Point. (n.d.). Big Data Analytics. *https://www.tutorialspoint.com/big_data_analytics/data_scientist.htm*.

TIGIST, Y. (2018). *A FRAMEWORK OF VIRTUALIZED INFRASTRUCTURE AS A SERVICE FOR ETHIOPIAN BANKING INDUSTRY.* Addis Ababa: AAU.

Timothy, K. (2016). ISO 8000: An ISO framework for data governance. *Wolverhampton Branch Meeting, University of Wolverhampton*.

Vossen, G. (2014). Big data as the new enabler in business and other intelligence. *Vietnam J Comput Sci* , 3-14.

Weathington, J. (2014). Launch a big data strategy awareness campaign. *https://www.techrepublic.com/article/launch-a-big-data-strategy-awareness-campaign/*.

Yang, L., Li, J., Elisa, N., Prickett, T., & Chao, F. (2019). Towards Big data Governance in Cybersecurity. *https://doi.org/10.1007/s41688-019-0034-9*.

APPENDIX A: Questionnaire

Dear Sir/Madam I am a graduate student in school of *Ethiopian Institution Technology Mekelle University Department of, EiT-M*. Currently I am conducting a research entitled "*Design and Implementation of Big Data Analytics Framework to Enhance Decision-Making at Abyssinia Bank: A Case Study in the Oromia Special Zone, Amhara Region*.". The research is undertaken as an academic requirement of partial fulfillment of the requirements for the Degree of Master of science in Information Technology.

The questionnaires are framed and directed towards Strategy, Technology, People and Environment point of view. To assist me, you will, on the following pages, be asked to indicate whether or not you agree with a set of statements concerning the bank's relationship with big data technology. I do sincerely hope that you take the time to answer all questions, and do so honestly.

Big data is a term most people have heard of, yet very few are truly familiar with. Therefore, I begin the survey by offering the definition of big data. This is to ensure that we share a common understanding of the term.

Finally, I confirm you that your response will be kept confidential and only used for academic purpose. Thank you in advance for your kind cooperation and dedicating your time.

If you have any review, please feel free and contact me at *bayew2003@gmail.com*

Yours Sincerely

Bayew Girma

Big Data Definition

What is big data?

Big data is data that is characterized by five V's: Volume, Variety, Velocity, Veracity and Value.

- ✓ **Volume** - represents the amount of data collected by a company, with data sets ranging from terabytes to zeta bytes in size.
- ✓ **Variety** - different types of data are available from a range of data sources, both external and internal to the firm, from sources such as customer databases, social media, ATM transactions and the like.
- ✓ **Velocity** – refers to the speed or frequency at which data are gathered and processed. It could be in real time or near time sensors, sales transactions and social media posts.
- ✓ **Veracity** –the degree in which the data is trusted and protected from unauthorized access to make a decision.
- ✓ **Value** – refers to economically useful benefits that an organization obtained from big data. The usefulness of the data is measured by how predictive the data is and how useful the information is that the organization can derive from this data.

Big data provides no value by itself. The value of big data comes from the ability to analyze data that was not previously available, or were too expensive to store or process, to provide new insights and improve the basis of decision making. As such data is either difficult or impossible to manage using traditional database or analytics tools, companies have begun exploring new technologies.

Big data management is about two things—big data and data management—plus how the two work together to achieve business and technology goals. Big data is mainly about the quantity of data, including massive data sets measured in tens of terabytes, or at times in hundreds of terabytes or peta bytes. Whereas big data management is the application of data management disciplines, tools, and platforms to the management of big data.

Part I. Respondent Information

Please provide your answer by putting (X) sign

2. Gender:

Male ☐

Female ☐

3. What is the highest level of formal education you have completed? (Please check only one.)

Attended High School ☐ Graduated High School ☐

Graduated College/University ☐ Post-Graduate Degree

4. Your role at your company ☐

5. Work experience

Less than 5 year ☐ From 5-10 years ☐

From 11-15 years ☐ More than 15 years ☐

6. Have you ever heard of Big Data?

Yes ☐

No ☐

Part II. Instruction: Please select your answers by putting (X) sign on the scale ranging strongly Agree through strongly Disagree in the appropriate space provided.

Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Strategy					
The bank has a clear mission, vision about implementing Big Data technology					
The bank policies are in place to promote and manage use of Big Data Management in the organization					
The bank has a direction, commitment and plan towards Big Data Management					
The bank has future development plan for basic ICT and information infrastructure to implement Big Data Management					
The bank implements strategic plan related to ICT provisioning/supply and support					
The bank has future development plan and respond towards the dynamic ICT advancement					
The bank's ICT HR strategy related to hiring experienced and professional experts' strength the efficiency of the bank					
The bank feel it is a strategic necessity to practice big data management to compete in the market place					
Technology					
Do you agree Big Data Management/analytics is advantageous for your bank?					
Using Big Data Management/analytics					

would provide my company with valuable information for decision making					
Using Big Data Management/analytics would improve the effectiveness of the bank operations and performance					
Do you think Big Data Management/analytics is a complex technology to adopt?					
The bank quality of internet connections is suitable					
Big Data Management/analytics will be compatible with my company's existing IT infrastructure and IT service management					
The technology will be compatible with the data captured at the bank					
Do you think the existing ICT would help in storing Big Data a better way					
The hardware and software required for the building Big Data Management/analytics is readily affordable					
The bank has security capabilities to adopt this technology					
People					
The level of training and education for ICT and Big Data in the bank is adequate.					
The bank has programs in place to train the users for Big Data Management/analytics					
Our employees have basic knowledge about Big Data Management/analytics					
Our IT employees have the ability to quickly					

learn and apply new technologies					
The bank has technical and managerial skills on the use of technological innovation					
Previous organizational change has been accepted by the majority of employees					
Environment					
Adoption of big data will give the bank a competitive advantage					
The bank closely follows the competitor's big data initiatives to determine its strength and weakness					
The telecommunications infrastructure is reliable and efficient in the bank					
There are firms in the market who provides support for use of this technology					
It is not difficult for the bank to protect data privacy for big data initiatives					
It is easy for the bank to comply with privacy related regulation to implement this technology					
Given the necessary resources, opportunities and knowledge to use big data, the bank will be ready to adopt it					
The bank has begun to focus on big data opportunities, but are not yet planning					

APPENDIX B: Interview Guide

1) Did the bank have a clear vision about implementing big data technologies?
2) Did the bank feel it is a strategic necessity to use big data technology to compete in the market place?
3) What do you think are the benefits of big data management for the bank?
4) Do you believe the bank have security capabilities to adopt this technology?
5) Did the employees have basic knowledge about big data technology?
6) Did the bank have programs in place to train the staffs on big data technology?
7) Had previous organizational change/ system change been accepted by majority of the employees?
8) Do you have a big data analyst or expert at the bank?
9) Do you believe adoption of big data technology gives the bank a competitive advantage? If yes, how?
10) Will big data technology be compatible with the bank's existing operating practices and infrastructure?
11) What tools and approaches are used for the management of the banks' data?
12) Did you believe the tools and approaches currently used for data management are sufficient for yielding the big data utilization benefits?
13) If you have developed a big data management framework (BDMF) in your bank, what kind of standard or framework the bank is employed in the process of BDMF development? If you have not developed BDMF, what are the reasons?

APPENDIX C: Proposed Framework Evaluation Survey

Mekelle University

Ethiopian Institution Technology

Department of, EiT-M

Dear Sir or Madam:

In partial fulfillment of the requirements for the Degree of Master of Science in Information Science, I am undertaking a research on — *Design and Implementation of Big Data Analytics Framework to Enhance Decision-Making at Abyssinia Bank: A Case Study in the Oromia Special Zone, Amhara Region*. Based on the study findings, I have amended the proposed framework and accordingly prepared this questionnaire. The objective of the questionnaire is to evaluate the proposed framework with respect to its comprehensiveness, clarity, completeness, correctness, and applicability.

This research is believed to produce results that can improve BoA data management tasks. Thank you for your dedication to provide your genuine feedback regarding the proposed framework.

Thank you again!

Bayew Girma

APPENDIX D: Python Script Code

python

```
import pandas as pd
import numpy as np
from datetime import datetime, timedelta
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
import warnings
warnings.filterwarnings('ignore')

# Set style for visualizations
plt.style.use('seaborn-v0_8')
sns.set_palette("husl")

class BankDataAnalyticsFramework:
    """
    A framework for big data analytics in banking decision-making
    """

    def __init__(self):
        self.customer_data = None
        self.transaction_data = None
        self.loan_data = None
        self.analysis_results = {}

    def generate_sample_data(self, num_customers=10000, num_transactions=500000):
        """
        Generate sample banking data for demonstration purposes
        In a real scenario, this would be replaced with actual data ingestion from various sources
        """
        print("Generating sample banking data...")

        # Generate customer data
        np.random.seed(42)
        customer_ids = [f'CUST_{i:06d}' for i in range(1, num_customers+1)]
```

```
        ages = np.random.normal(45, 15, num_customers).astype(int)
        ages = np.clip(ages, 18, 100)

        self.customer_data = pd.DataFrame({
            'customer_id': customer_ids,
            'age': ages,
            'income': np.random.lognormal(10, 0.8, num_customers).astype(int),
            'credit_score': np.random.normal(650, 100, num_customers).astype(int),
            'account_age_days': np.random.exponential(2000, num_customers).astype(int),
            'branch_id': np.random.choice([f'BRANCH_{i}' for i in range(1, 21)], num_customers),
            'region': np.random.choice(['Oromia Special Zone', 'Amhara Region', 'Addis Ababa', 'Oth
er'],
                                       num_customers, p=[0.4, 0.3, 0.2, 0.1])
        })

        # Ensure credit scores are within valid range
        self.customer_data['credit_score'] = np.clip(self.customer_data['credit_score'], 300, 850)

        # Generate transaction data
        transaction_dates = [datetime.now() - timedelta(days=np.random.randint(0, 365))
                             for _ in range(num_transactions)]

        self.transaction_data = pd.DataFrame({
            'transaction_id': [f'TRX_{i:08d}' for i in range(1, num_transactions+1)],
            'customer_id': np.random.choice(customer_ids, num_transactions),
            'transaction_date': transaction_dates,
            'amount': np.random.exponential(150, num_transactions),
            'type': np.random.choice(['DEPOSIT', 'WITHDRAWAL', 'TRANSFER', 'PAYMENT'],
                                     num_transactions, p=[0.3, 0.4, 0.2, 0.1]),
            'channel': np.random.choice(['BRANCH', 'ATM', 'MOBILE', 'ONLINE'],
                                        num_transactions, p=[0.2, 0.3, 0.3, 0.2])
        })

        # Generate loan data
        loan_customers = np.random.choice(customer_ids, int(num_customers * 0.4), replace=Fals
e)

        self.loan_data = pd.DataFrame({
            'loan_id': [f'LOAN_{i:06d}' for i in range(1, len(loan_customers)+1)],
```

```
            'customer_id': loan_customers,
            'loan_amount': np.random.lognormal(9, 1.2, len(loan_customers)).astype(int),
            'interest_rate': np.random.normal(8, 2, len(loan_customers)),
            'term_months': np.random.choice([12, 24, 36, 48, 60], len(loan_customers)),
            'status': np.random.choice(['ACTIVE', 'CLOSED', 'DELINQUENT', 'DEFAULT'],
                                       len(loan_customers), p=[0.6, 0.3, 0.07, 0.03])
        })

        print(f"Generated data: {num_customers} customers, {num_transactions} transactions, {len
(loan_customers)} loans")

    def preprocess_data(self):
        """Preprocess and clean the data"""
        print("Preprocessing data...")

        # Calculate additional features
        current_date = datetime.now()

        # Customer activity features
        transaction_counts = self.transaction_data.groupby('customer_id').size()
        transaction_avg_amount = self.transaction_data.groupby('customer_id')['amount'].mean()

        last_transaction = self.transaction_data.groupby('customer_id')['transaction_date'].max()
        days_since_last_transaction = (current_date - last_transaction).dt.days

        # Merge features into customer data
        self.customer_data = self.customer_data.merge(
            transaction_counts.rename('transaction_count'), on='customer_id', how='left'
        )
        self.customer_data = self.customer_data.merge(
            transaction_avg_amount.rename('avg_transaction_amount'), on='customer_id', how='left'
        )
        self.customer_data = self.customer_data.merge(
            days_since_last_transaction.rename('days_since_last_transaction'), on='customer_id', ho
w='left'
        )

        # Fill NaN values for customers with no transactions
        self.customer_data['transaction_count'] = self.customer_data['transaction_count'].fillna(0)
```

```
        self.customer_data['avg_transaction_amount'] = self.customer_data['avg_transaction_amount'].fillna(0)
        self.customer_data['days_since_last_transaction'] = self.customer_data['days_since_last_transaction'].fillna(365)

        # Create customer segments based on activity and value
        self.customer_data['customer_segment'] = pd.cut(
            self.customer_data['transaction_count'] * self.customer_data['avg_transaction_amount'],
            bins=[-1, 100, 1000, 10000, float('inf')],
            labels=['LOW', 'MEDIUM', 'HIGH', 'VIP']
        )

    def analyze_customer_behavior(self):
        """Analyze customer behavior patterns"""
        print("Analyzing customer behavior...")

        # Regional analysis
        regional_analysis = self.customer_data.groupby('region').agg({
            'income': 'mean',
            'credit_score': 'mean',
            'transaction_count': 'mean',
            'customer_id': 'count'
        }).rename(columns={'customer_id': 'customer_count'})

        # Channel preference analysis
        channel_analysis = self.transaction_data.groupby(['channel', 'type']).size().unstack(fill_value=0)

        # Customer segmentation analysis
        segmentation_analysis = self.customer_data.groupby('customer_segment').agg({
            'income': 'mean',
            'credit_score': 'mean',
            'transaction_count': 'mean',
            'customer_id': 'count'
        }).rename(columns={'customer_id': 'customer_count'})

        self.analysis_results['regional_analysis'] = regional_analysis
        self.analysis_results['channel_analysis'] = channel_analysis
        self.analysis_results['segmentation_analysis'] = segmentation_analysis
```

```
    return regional_analysis, channel_analysis, segmentation_analysis

def predict_loan_risk(self):
    """Build a model to predict loan default risk"""
    print("Building loan risk prediction model...")

    # Prepare data for modeling
    loan_customers = self.loan_data.merge(self.customer_data, on='customer_id', how='left')

    # Create target variable: 1 for default, 0 otherwise
    loan_customers['default_risk'] = (loan_customers['status'] == 'DEFAULT').astype(int)

    # Select features for model
    features = ['age', 'income', 'credit_score', 'account_age_days',
            'transaction_count', 'avg_transaction_amount', 'loan_amount', 'interest_rate']

    X = loan_customers[features]
    y = loan_customers['default_risk']

    # Handle missing values
    X = X.fillna(X.mean())

    # Split data into train and test sets
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

    # Train a Random Forest model
    model = RandomForestClassifier(n_estimators=100, random_state=42)
    model.fit(X_train, y_train)

    # Make predictions
    y_pred = model.predict(X_test)

    # Evaluate model
    report = classification_report(y_test, y_pred, output_dict=True)
    cm = confusion_matrix(y_test, y_pred)

    # Feature importance
    feature_importance = pd.DataFrame({
```

```
        'feature': features,
        'importance': model.feature_importances_
    }).sort_values('importance', ascending=False)

    self.analysis_results['risk_model_report'] = report
    self.analysis_results['risk_model_cm'] = cm
    self.analysis_results['risk_model_feature_importance'] = feature_importance

    return report, cm, feature_importance

def generate_visualizations(self):
    """Generate visualizations for insights"""
    print("Generating visualizations...")

    # Set up the figure
    fig, axes = plt.subplots(2, 2, figsize=(15, 12))
    fig.suptitle('Abyssinia Bank - Analytics Insights', fontsize=16, fontweight='bold')

    # Regional customer distribution
    region_counts = self.customer_data['region'].value_counts()
    axes[0, 0].pie(region_counts.values, labels=region_counts.index, autopct='%1.1f%%')
    axes[0, 0].set_title('Customer Distribution by Region')

    # Credit score distribution by region
    sns.boxplot(data=self.customer_data, x='region', y='credit_score', ax=axes[0, 1])
    axes[0, 1].set_title('Credit Score Distribution by Region')
    axes[0, 1].tick_params(axis='x', rotation=45)

    # Transaction channels
    channel_counts = self.transaction_data['channel'].value_counts()
    axes[1, 0].bar(channel_counts.index, channel_counts.values)
    axes[1, 0].set_title('Transaction Channel Usage')
    axes[1, 0].tick_params(axis='x', rotation=45)

    # Customer segments
    segment_counts = self.customer_data['customer_segment'].value_counts()
    axes[1, 1].bar(segment_counts.index, segment_counts.values)
    axes[1, 1].set_title('Customer Segmentation')
```

```
    plt.tight_layout()
    plt.savefig('bank_analytics_insights.png', dpi=300, bbox_inches='tight')
    plt.show()

    # Feature importance visualization
    if 'risk_model_feature_importance' in self.analysis_results:
        fig, ax = plt.subplots(figsize=(10, 6))
        importance_data = self.analysis_results['risk_model_feature_importance']
        ax.barh(importance_data['feature'], importance_data['importance'])
        ax.set_title('Loan Default Risk - Feature Importance')
        ax.set_xlabel('Importance')
        plt.tight_layout()
        plt.savefig('risk_feature_importance.png', dpi=300, bbox_inches='tight')
        plt.show()

def generate_recommendations(self):
    """Generate business recommendations based on analysis"""
    print("\n" + "="*50)
    print("DATA-DRIVEN RECOMMENDATIONS FOR ABYSSINIA BANK")
    print("="*50)

    # Regional recommendations
    oromia_data = self.analysis_results['regional_analysis'].loc['Oromia Special Zone']
    amhara_data = self.analysis_results['regional_analysis'].loc['Amhara Region']

    print(f"\n1. Regional Strategy:")
    print(f"   - Oromia Special Zone has {int(oromia_data['customer_count'])} customers with "
          f"average credit score of {oromia_data['credit_score']:.1f}")
    print(f"   - Amhara Region has {int(amhara_data['customer_count'])} customers with "
          f"average credit score of {amhara_data['credit_score']:.1f}")

    # Channel recommendations
    mobile_transactions = self.analysis_results['channel_analysis'].loc['MOBILE'].sum()
    branch_transactions = self.analysis_results['channel_analysis'].loc['BRANCH'].sum()

    print(f"\n2. Digital Transformation:")
    print(f"   - Mobile channel accounts for {mobile_transactions} transactions")
    print(f"   - Branch transactions: {branch_transactions}")
    print(f"   - Recommendation: Invest in mobile banking features to reduce branch costs")
```

```
        # Risk management recommendations
        if 'risk_model_feature_importance' in self.analysis_results:
            top_features = self.analysis_results['risk_model_feature_importance'].head(3)
            print(f"\n3. Risk Management:")
            print(f"   - Top factors predicting loan default: {', '.join(top_features['feature'].tolist())}")
            print(f"   - Recommendation: Enhance credit scoring model with these predictive factors"
)

        # Customer segmentation recommendations
        vip_customers = self.analysis_results['segmentation_analysis'].loc['VIP']
        print(f"\n4. Customer Segmentation:")
        print(f"   - VIP customers: {int(vip_customers['customer_count'])} high-value customers")
        print(f"   - Recommendation: Develop personalized services for VIP customers to improve r
etention")

        print("\n" + "="*50)

    def run_full_analysis(self):
        """Execute the complete analytics pipeline"""
        print("Starting Abyssinia Bank Analytics Framework...")
        self.generate_sample_data()
        self.preprocess_data()
        self.analyze_customer_behavior()
        self.predict_loan_risk()
        self.generate_visualizations()
        self.generate_recommendations()
        print("Analysis complete! Results saved to visualizations.")


# Execute the analysis framework
if __name__ == "__main__":
    analytics_framework = BankDataAnalyticsFramework()
    analytics_framework.run_full_analysis()
```