AFRICA UNIVERSITY
(A united Methodist-Related Institution)

BUILDING A CONCEPTUAL FRAMEWORK FOR USING BIG DATA ANALYTICS IN SMALL TO MEDIUM ENTERPRISES IN ZIMBABWE

BY

DUMISANI MASUKU

A DISSERTATION/THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF EXECUTIVE MASTER IN BUSINESS ADMINSTRATION IN THE DEPARTMENT OF BUSINESS SCIENCES

2024

Abstract

The impact of the Small to Medium Enterprises to the Zimbabwean Economy is an evergreen debate, the growth of the informal market and its ravaging threat to the established firms and corporates has also been a buzz topic in recent conversations related to the Zimbabwean economy. This study critically analyse the impact of the adoption of Big Data Analytics for Small to Medium Enterprise firms in Zimbabwe. The study is also aimed at laying out a conceptual framework for small businesses to take on Big Data Analytics and the short, medium to long term benefits which come from the adoption of such technologies. The study explores the TOE (Technological, Organisational and Environmental) factors and outlines how these factors affect the adoption of a Big Data Analytics system for Small to Medium Enterprises. The imminent challenges associated with the adoption of a BDA system have been clearly articulated which relate to aspects such as data security and protection and other aspects such as ethical issues and regulatory issues as it relates to personal data protection (Data Protection Act of 2021). A survey was conducted using a questionnaire distributed to business owners and managers of these firms with questions structured to establish the current operational status of the firms as well as to check if there is the appreciation and presence of fundamental processes and infrastructure which relates to Big Data Analytics. Some opportunity points which can be leveraged to ensure and guarantee business growth were identified from the sixty eight companies which were surveyed from different industries in Harare. Results of the study point out to a possible growth potential realisable by on-boarding and successfully running a Big Data Analytics system, also improved customer experience and the ability to make informed decisions which are data driven is another low hanging fruit and a product of a Big Data Analytics system.

Key Words: Big Data Analytics, Small to Medium Enterprises, Machine Learning

Declaration

I declare that this dissertation is my original work except where sources have been cited and acknowledged. The work has never been submitted, nor will it ever be submitted to another university for the award of a degree.

…….

Dumisani Masuku. Signature (**05/12/23**)

……..

Dr. Yogesh Kumar Awasthi Signature (**05/12/23**)

Copyright

Acknowledgments

I want to acknowledge my Supervisor Dr Yogesh Kumar Awasthi for the support and guidance, not forgetting my wife Rufaro for the moral support throughout the course of my study.

Dedication

I dedicate this work to the Small to Medium Enterprise sector and to the future research and study on the subject of Big Data Analytics in Zimbabwe.

List of Acronyms and Abbreviations

BDA	Big Data Analytics
SME	Small to Medium Enterprises.
TOE	Technology, Organisation and Environment
SCM	Supply Chain Management
AI	Artificial Intelligence
GDP	Gross Domestic Product
NDS1	National Development Strategy One
ILO	International Labour Organisation
RAM	Random Access Memory
SSD	Solid State Drive
HDD	Hard Disk Drive
AMQP	Advanced Message Queuing Control
JMS	Java Message Service

Definition of Key Terms

Big Data Refers to sets of data which are beyond the capacity of traditional relational databases

Data Analysis Analytics involves the broad examination of information to extract useful patterns and insights, businesses use analytics for example to understand consumer behaviours.

Descriptive Analytics It is a statistical interpretation used in analysis of historical data so as to identify patterns and relationships.

Diagnostic Analysis The method examines data so as to understand the root causes of certain events and behaviours.

Predictive Analytics This is an arm of advanced data analytics which makes predictions about future events using the historical data, it utilises statistical modelling and data mining techniques and machine learning to identify risks and opportunities.

Prescriptive Analytics It is the use of advanced processes and related tools to analyse data sets to prescribe or recommend the optimal route or course of action moving forward.

Data Acquisition It is the process of translating real-world signals to the digital domain for display, storage and analysis.

Data Curation It is the organisation and integration of data collected from different sources.

Machine Learning It is a branch of artificial intelligence and computer science which focuses on using data and algorithms to enable the imitation of way humans learn gradually improving accuracy.

Operational Efficiency Is the relationship between an organisation's output and input that when healthy helps a firm to cut down on unnecessary costs while driving the topline.

Table of Contents

List of Tables

List of Figures

CHAPTER 1 INTRODUCTION

1.1 Introduction

The study focused on the development of a conceptual framework for the implementation of Big Data Analytics for Small to Medium Enterprises in Zimbabwe. The background of the study, the statement of the problem, the objectives of the study and the research questions are clearly articulated and explained in the chapter so as to provide a foundation through which the research problem can be understood. The purpose, significance and the assumptions of the study were presented and described. The delimitation and the limitation of the study were given and discussed. Key concepts of the research are defined and the structure of the study is laid out.

1.2 Background of the Study

Many companies in Zimbabwe are failing to grow their profitability in the market due to lack of awareness as well as lack of an implementation plan to leverage on data from the customer information resources. Companies are faced with large volumes of data in their day to day operations, and the data comes in many forms, beneath this data are trends and insights which can be leveraged to solve some of the business problems and also inform entrepreneurs to make data driven decisions. Innovation in the marketplace have been seen to promote growth of a firm (North & Smallbone, 2010) Customers provide information through the following resources from the daily business processes, (source:https://monkeylearn.com/blog/customer-feedback-methods/);

- Order forms
- Enquiries, complaints and warranty cards
- Customer satisfaction surveys and feedback cards

- Website and social media polls or messages
- In-person or over-the-counter surveys
- Email correspondence.

These sources provide data which can be analysed and processed to appreciate consumer trends and other hidden opportunities which can be used to grow the business, data in its raw format is of no use for any business owners and unless it is processed intelligently by using different tools the businesses will not benefit from it. Small to Medium Enterprises have been for long an important facet to the Zimbabwean economy, a study by the International Finance Corporation (IFC) shows that Small and Medium Enterprises (SMEs) have been a significant contributor to the Zimbabwean economy and it outlines that SME's constitute more than 90% of all formal businesses in Zimbabwe and employ between 50-60% of the total workforce, it also highlight that the SME's contribute 34% of the GDP annually (Musabayana, Mutambara, & Ngwenya, 2022). According to Statistica, the sector has played a positive role in the development and progress of Zimbabwe contributing in foreign exchange earnings and GDP, in 2016 the Small to Medium Enterprise sector contributed a total of USD$8.58Billion to the country's GDP and employed 5.9million people which was 75% of the total workforce. On the other hand Covid-19 pandemic together with globalisation and the advancement of technology has changed the operating environment and many companies have been pushed to become more agile and innovative in order to respond to the needs of the consumers. According to a new McKinsey Global Survey of executives, most companies have accelerated the digitization of their customer and supply-chain interactions and of their internal processes by three to four years due to the Covid-19 pandemic. Survival is therefore premised on the adaptability of the SME's to the dynamic and ever-changing

environment. Businesses have massive data at their disposal and this data can be in the form of daily recorded transactions with customers and suppliers, feedback from the social interactions of the business and general operational information for example order fill rate, customer complaints and even lost sales reports. Against this background, this study seeks to develop a suitable framework for the implementation of data analytics with a view of improving competitiveness of companies in Zimbabwe with a strong inclination to small to medium enterprises.

Big Data Analytics (BDA) is now the frontier for value creation, productivity and innovation (Wamba, et al., 217).

1.3 Statement of the problem

Small to Medium Enterprises (SME) represent a critical sector which can drive economic growth in Zimbabwe, to this end it can be seen and proven that many SME's are losing out on opportunities to scale up their operations and fully deploy their resources for growth by using Big Data Analytics (BDA) technologies. Organisations as they interface with customers daily collect valuable information which can be analysed intelligently to point out to underlying trends which can be leveraged to make business decisions. Small to Medium Enterprises have the capacity to contribute immensely to the nation's vision and can be earmarked as growth drivers (Keizer, Dijkstra, & Halman, 2002), the meticulous and yet systematic roll out of a Big Data Analytics framework can be a game changer for the Zimbabwean Small to Medium Enterprises by assisting entrepreneurs and business owners to track their daily business interactions and make data driven decisions. Almost every country that has achieved major economic growth had a concentrated drive to establish SMEs (Manzoor , Wei , & Sahito, 2021). By missing out on patterns behind the data companies miss out on opportunities to give better customer experience, offer tailor made services and

products. (Perdanaa, Leeb, Kohb, & Arisandi, 2021) in their research entitled “Data analytics in small and mid-size enterprises: Enablers and inhibitors for business value and firm performance” to examine the potential for growth realisable due to the adoption and use of Data Analytics (DA) for Small to Medium Enterprises, they investigated the potential enablers and inhibitors of DA business value relevant to SMEs. They found that in-formation quality and system quality were are the strong predictors of DA business value enablers, whereas lack of under-standing and concerns over data security and privacy were the most salient predictors of DA business value inhibitors. The analysis highlights the importance of understanding both enablers and inhibitors in IT business value research. The research offered practical suggestions to the relevant stakeholders on formulating strategies to mitigate potential deterrents of DA business value generation in SMEs, so that they can reap the benefits from DA.

1.4 Research Objectives

The thesis explored and evaluated Big Data Analytics in the context of the SME sector in Zimbabwe, Small to Medium Enterprises can benefit from the philosophies and success drivers behind BDA and the objectives were;

i. To analyze the components of a BDA system.

ii. To evaluate how the implementation of a BDA framework can improve the competitiveness of a small to medium firm

iii. To evaluate the risks encountered when implementing BDA.

iv. To create a suitable implementation framework for BDA for SME’s

1.5 **Research Questions**

i. What are the main processes and infrastructure required for a BDA system?

ii. How does Data Analytics improve a firm's performance?

iii. What are the challenges faced by SMEs in adopting BDA capability in their operation?

iv. How can the BDA framework can be rolled out for an SME?

1.6 **Hypothesis**

H_1: Performance for SME's can be impacted positively by the adoption of Big Data Analytics resulting in growth in sales and ultimately revenue.

1.7 Significance of the Study

The study will add to the growing literature and study on the adoption and use of Big Data Analytics in Zimbabwe, the study has been premised on existing theories on BDA and it is an expansion of the literature relating this new emerging technology in the context of Zimbabwe. The study unearthed the current state of readiness of the Zimbabwean Small to Medium Enterprises from a knowledge view and also from an infrastructure standpoint and consideration has been evaluated of the willingness of the business owners to roll out the system.

1.8 Delimitation of the Study

Due to the large pool of potential participants the study focused on the population of business owners and managers in Harare and in the specific industries which included then Retail Industry, Personal Care, Manufacturing and Mining, Consultancy, Agriculture and Engineering Services. The research focused on the business owners and senior managers of these firms to ensure that the data collected is accurate.

1.9 Limitations of the Study

The study targeted all the different industry categories which fall under the scope of the Small to Medium Enterprises, to ensure contact and close engagement of the participants participants from Harare and the neighbouring towns were targeted for the study, the study focused on the business owners of these firms and in cases where access to the owners was not possible the senior managers were engaged. This was done so as to ensure that the data collected is accurate and it is also has the consent of the known owner. The study was undertaken during a critical phase where there is a growing focus on the informal sector and the Small to Medium Enterprises industry to be particular, the economy also during the period of the study was noted to be volatile under the current multi-currency regime. The study was limited in so far as collection of verified customer information was concerned, for instance, customers could not disclose their variable financial position, and customer information was not shared. There was also a pushback from other respondents and a general fear of exposure of malpractices which affected collection of results.

CHAPTER 2 REVIEW OF RELATED LITERATURE

2.1 Introduction

The purpose of this chapter is to present the theoretical framework for the thesis. The chapter will introduce previous literature on the subject of Big Data Analytics (BDA) after which it will explore the types of BDA and the related criticism and barriers to its adoption in Zimbabwe. BDA will be introduced as the key to unlock performance for Small to Medium Enterprises, the risks associated to the adoption and use of the system will be outlined together with the proposed operational framework to successfully implement it in the context of Zimbabwe.

2.1.1 Definition of an SME

Small to medium-sized enterprises (SMEs), also known as small and medium-sized enterprises (SMEs), are businesses below certain turnover, asset, or employee limits (Caldeira & Ward, 2003). The limits vary from country to country; experts and government agencies have set the thresholds for each country or region. Chapter 24:12 of the Small and Medium Enterprises Act of Zimbabwe defines an SME as a business entity, whether corporate or unincorporated, which, together with any of its branches or subsidiaries, meets the criteria for classification as a micro enterprise, small enterprise, or medium enterprise specified in the second, third and fourth schedules of the Act. The fourth schedule of the same act outlines the classifications of the SME's by industry the differences are in the levels of capitalization, turnover and employment.

2.1.2 Importance of Small to Medium Enterprises

2.1.2.1 Job Creation

Small to Medium Enterprises play a crucial and pivotal role according to a study done by the International Labour Organisation (ILO), they provide two thirds of all formal

jobs in developing countries and as high as 80% in low-income countries. Unemployment has had tremendous impact on the economy and on the Zimbabwean society at large, it has led to drug abuse, crime, violence and promiscuity, the Small to Medium Enterprises sector in providing employment especially for the youth rids the country of these negative aspects and therefore impacts the society positively. The drive of the government of Zimbabwe has been on job creation as has always been reiterated by the president of Zimbabwe, as such Small to Medium Enterprises are a low hanging fruit in creating jobs in the country.

2.1.2.2 Flexibility and Innovation

Small to Medium Enterprises are more flexible and in many instances more innovative than bigger firms especially in the context of the Zimbabwean industry, strategic flexibility contributes to firm growth in Small to Medium Enterprises (Brozovic, Jansson, & Boers, 2023), the study outlines enablers and barriers to strategic flexibility, these include but not limited to strategic orientation, positive impact of slack resources and the negative effect of lack of competent employees (Brozovic, Jansson, & Boers, 2023). As SME's are more flexible than bigger firms this also allows them to be agile and quickly adapt to the market conditions and any changes, on the innovation side SME's are also investing in systems and models to make sure that they serve their customers faster and thereby improving customer experience.

2.1.2.3 Competition in Markets

Small to Medium Enterprises play a crucial role in increasing competition in the market for larger firms. By being more flexible, these smaller companies can adapt quickly and therefore respond to the urgent market changes thereby offering better customer services. Small to Medium Enterprises are not burdened by overheads and other obligations therefore for some sectors they become more competitive than the

established firms. Without SMEs large corporates will hold monopoly in all the activity areas, therefore by providing competition SMEs will create a competitive and healthy economy. Small Medium Enterprises also help big organisations in some specific areas of operation so that that they are better able to supply the market therefore if the SMEs are were to be removed the big firms would be forced to be involved in non-core activities which will be inefficient for the enterprise, for example companies like Irvines depend on the SMEs in the their business processes to augment supply of birds to the company. The recent reports by one of Zimbabwe's biggest retailers OK Zimbabwe shows that the firm has been battling headwinds from the Informal market where the company has lost its market share as the average Zimbabwean prefers buying from the informal retailers as they offer their products cheaper. This competition will result in less risks of arbitrage by the big firms and rationalization of margins which will result in more affordable pricing.

2.1.2.4 Economic Growth

The Small to Medium Enterprises sector is the pillar of many economies in the world including Zimbabwe as they drive economic growth through creation of jobs, design of new products and services, some of these products can be exported to regional and international markets. Small to Medium Enterprise also come with innovative solutions which lead to improved yields and efficiencies, they play a crucial role in both developed and developing economies,

Solving Productivity Challenges*:*

Sluggish productivity growth poses a significant threat to overall economic progress. It affects citizens' well-being by affecting income growth, increasing inequality, and creating challenges with loan repayment. In recent years, productivity growth has stalled in many places. For instance, between 2010 and 2014, the average productivity

growth in seven OECD countries dropped to 0.5 percent per year from 2.4 percent per year between 2000 and 2004. SMEs contribute to this productivity problem. The productivity gap between large companies and SMEs can vary significantly within the same sector or among countries. Increasing the productivity of SMEs. Adopting established processes by incorporating technologies and established processes from larger organizations, SMEs can grow more quickly. It is necessary to close this adoption gap. Start-ups are a vital subset of SMEs that are the drivers of innovation. They are valuable providers of fresh ideas because of their capacity to reconsider accepted procedures and push the boundaries of the industry. By using specialized strategies, governments everywhere can aid in closing this gap. Establishing a business-friendly environment that fosters the growth of SMEs. Enabling talent, resources, and competencies to be more easily accessible. Supporting innovation and the uptake of new technologies. Managing SMEs' funding needs is a component of financial inclusion. SMEs have enormous potential for innovation and productivity increases in addition to being essential for economic success. Governments and interested parties need to work together to realize their full potential.

2.1.3 Challenges faced by SME's in Zimbabwe

2.1.3.1 The Operating Environment

A number of legal, economic and social developments are usually a direct result of the political decisions made by the governments in place, these decisions and policies include the government stance on indigenisation, policies on privatization of state-owned industries and the measures rolled out to curtail inflation among other issues, SME's are affected to a very great extent by the political and legal environment, these factors affect companies local and international (Dumbu & Chadamoyo, 2012). The political environment affect investor confidence, and when there is general stability

and consistency of policies investors tend to favour such environments. There is a need to increase investor confidence by introducing laws that do not frighten investors. Government support plays an important role in the development and growth of small and medium-sized enterprises (Alkahtani, Nordin , & Ullah, 2020) this includes creating a favourable business environment within a stable political climate.

To kindle economic growth through entrepreneurship, governments need to focus on supporting entrepreneurs and this is attainable through the right mix of policies. Consistent policies guarantee stability and will allow SMEs to plan and forecast without any fear of future unforeseen eventualities. Government measures are crucial to protect small and medium-sized enterprises from internal competition (from large corporations) and external competition. Regulatory restraints such as licensing and regulations also pose a serious threat to small businesses.

2.1.3.2 Finance and Operational Constraints

Small to medium-sized enterprises require financing in order to run their daily operations, to restock, to open more routes to the market, and to on-board new technologies and survive in the market. Cash flow problems are an obstacle to most businesses in Zimbabwe, and liquidity glitches and inability to service debts are the major causes of financial difficulties for businesses. A number of SMEs rely on internal funding sources during the earliest years of their lifecycle but as they grow in size and scale up their activities they require access to external funding, the lack of facilities for Small to Medium Enterprises hamper growth and expansion. The runaway inflation in Zimbabwe coupled with the chronic shortages of foreign currency in the market has resulted in the stifling of performance of SMEs. A deliberate effort to develop the economy of Zimbabwe will have a positive impact on the SMEs. On the

other hand companies can improve their internal organisation so as to better attract capital and should have solid business models which are bankable.

2.1.4 Big Data Analytics

Data analytics is the science of integrating heterogeneous data from diverse sources, drawing inferences, and making predictions to enable innovation, gain competitive business advantage, and help strategic decision-making (Gudivada, 2017). Big Data can also be viewed as high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making (Gandom & Haider, 2015). The two definitions bring out the connection or the link between Big Data Analytics (BDA) and Business Intelligence in a firm, where businesses can competently manage data and incorporate it into their decision-making processes; this will result in growth in sales and profitability consequently. BDA can also defined as the process of exploring large and diverse data sets to uncover hidden patterns, unknown relationships, market trends, customer preferences, and other useful business using advanced analytics techniques such as machine learning, data mining, natural language processing, and predictive analytics (Hu, Wen, Chua, & Li, 214). Big data analytics is used in a number of industries for instance the healthcare, finance, retail, and manufacturing to ensure process and system improvements. The three Vs of big data analytics explain how big data varies from traditional data management, these three V's are namely volume, velocity and variety

- Volume: Refers to the quantity of data generated. Big data is characterized by enormous volumes of data which is generated from several processes and it is very difficult to manage using traditional data management methods.

- Speed: Refers to the speed at which data is created and processed. Big data is frequently generated in real time therefore businesses need to be able to swiftly process this data to get insights
- Diversity: Refers to the different types of data generated. Big data comes in several forms namely structured data, semi-structured data, and unstructured data

2.2 Theoretical Framework

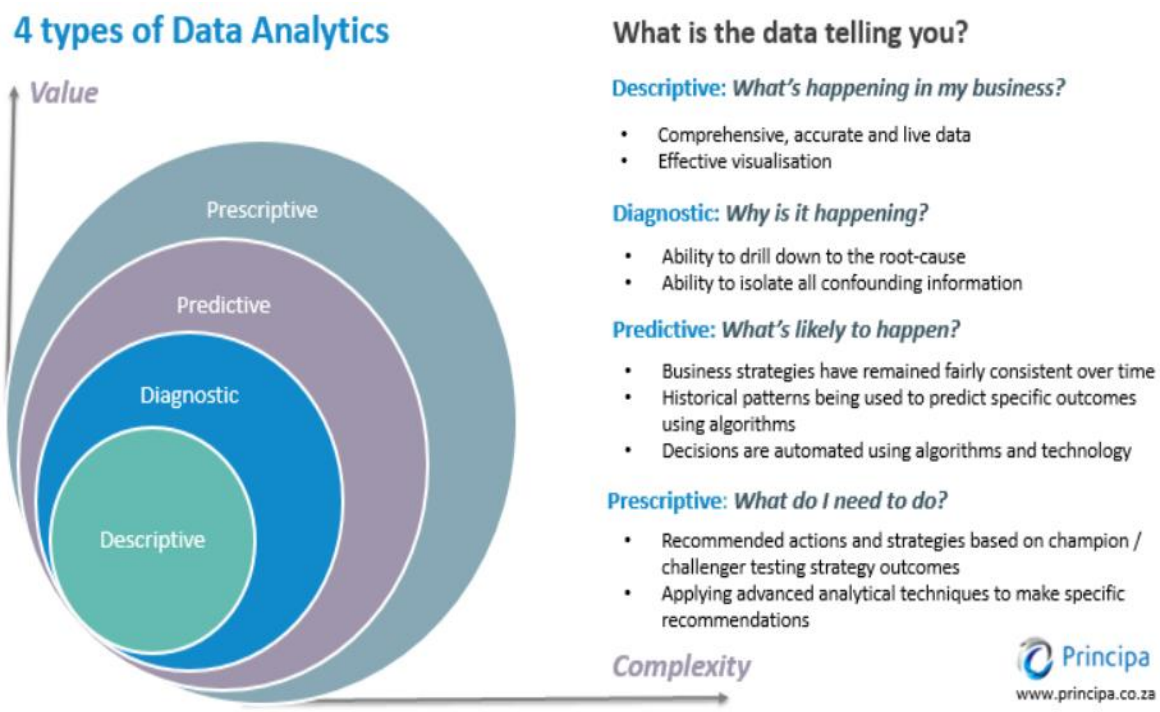
2.2.1 Data Analytics Types and related theories

Data analytics is the practice of evaluating or examining data sets to postulate the underlying facts they contain, this is done with the help of specialized systems and software. There are four main types of data analytics, namely;

- Descriptive Analytics.
- Diagnostic Analytics.
- Predictive Analytics.
- Prescriptive Analytics.

BDA encompasses extensive analysis of data to meet different needs, the different types of data require therefore require different approaches. Analytics technologies are often classified into four categories (Lismont, Vanthienen, Baesens, & Lemahieu, 2017). Figure 1 shows the different types of data analytics and what they entail.

Figure 1: Types of Data Analytics



Note: The four types of Data Analytics. From: "The science and types of Data Analytics,"https://www.datasciencecentral.com/the-4-types-of-data-analytics/

2.2.1.1 Descriptive analytics

Descriptive analytics is a method of data analysis which breaks down large sums of data into smaller usable pieces of information so as to help businesses to understand what occurred during a particular operation, process, or series of transactions. It is used to detect trends and underlying relationships in data so as to answer the big question which is, "What happened?". It is the simplest form of data analysis as it only defines trends and relationships but does not go a step ahead to provide deep explanations or clarifications. Because of its simplicity, it is likely to be used routinely in an organization. Statistical software such as Microsoft Excel and data visualization tools for example Google Charts and Tableau will help analyze the data and identify trends and relationships between different variables and consequently present this

information visually. Descriptive analysis is mainly valuable when using trends as the starting point for analysis of data so as to converse any changes over time and drive decision-making. The information which is provided will be computed from the past by calculating the descriptive quantities such as the average, mode and the standard deviation. The results extracted will then be presented in graphical format such as charts and lines. This method is the most popular type in analytics used by most organizations. This form of analysis is used to answer questions such as “What is the trend?”, “What are the drivers?”, and “Why is this happening?” (Fitz-Enz, 2010).

2.2.1.2 Predictive analytics

In predictive analytics the main objective will be to come up with models which can be used to predict future events based on the assumption that past events are bound to happen in the future (Zhao, 2020). The regularly used method is regression modeling where the objective will be to create a model where parameters are calculated in a manner which minimizes residuals, the other methods used include decision trees, support vector machine and neural networks. The most relevant question predictive analytics can answer is “What would happen if?” (Epstein, 2013). Predictive analytics differs from prescriptive analytics, which determines the best course of action based on predictions made by predictive analytics (Cavanillas, Curry, & Wa, 2016). Predictive analytics differs from descriptive analytics as it focuses on examining past data to appreciate what happened and the reasons why certain events happened, prescriptive analytics is used to gain understanding of historical trends and patterns, while predictive analytics is used to make predictions about future events.

2.2.1.3 Diagnostic Analysis

Diagnostic Analysis assists to answer the important question, which is why did it happen (Soltanpoor & Sellis , 2016)? This is achieved by a thorough scrutiny of the

historic data and merging multiple sources in search of patterns, trends and correlations. It can be done manually using algorithms or by using statistical software for example Microsoft Excel. This is used to ascertain behaviours and patterns with the goal of finding out why certain results occur. The diagnostic analysis route can be well understood as the next step which comes after identification of trends using descriptive analysis. Diagnostic Analysis has its main focus on cause and effect and it is sometimes named root cause analysis. The objective is to find and explain irregularities and outliers. Even though the method of Diagnostic Analysis has many benefits it should be noted that it also comes with its own downside. For instance, it is dependent on data quality, if the data is imperfect, incorrect, or skewed it will entail that the conclusions drawn from the data will also be incorrect. The main objective of diagnostic analytics is to;

- Identification of incongruities: the areas which need examination, these incongruities can be pulled out for querying and highlighting significant questions which are impossible to be answered by merely looking at the data, for example, why have the sales gone down for a particular period.
- Determination of causal relations: when the anomalies have been singled out, advanced statistical techniques will be used to see whether there any relationships.

Latest advances in diagnostic analytics use machine-learning techniques to supplement the analysis. Computers process vast amounts of data to identify patterns, pick anomalies, and show outliers or uncharacteristic trends (Ridzuana & Mohd Nazmee Wan Zainon, 2022)

2.2.1.4 Prescriptive analytics

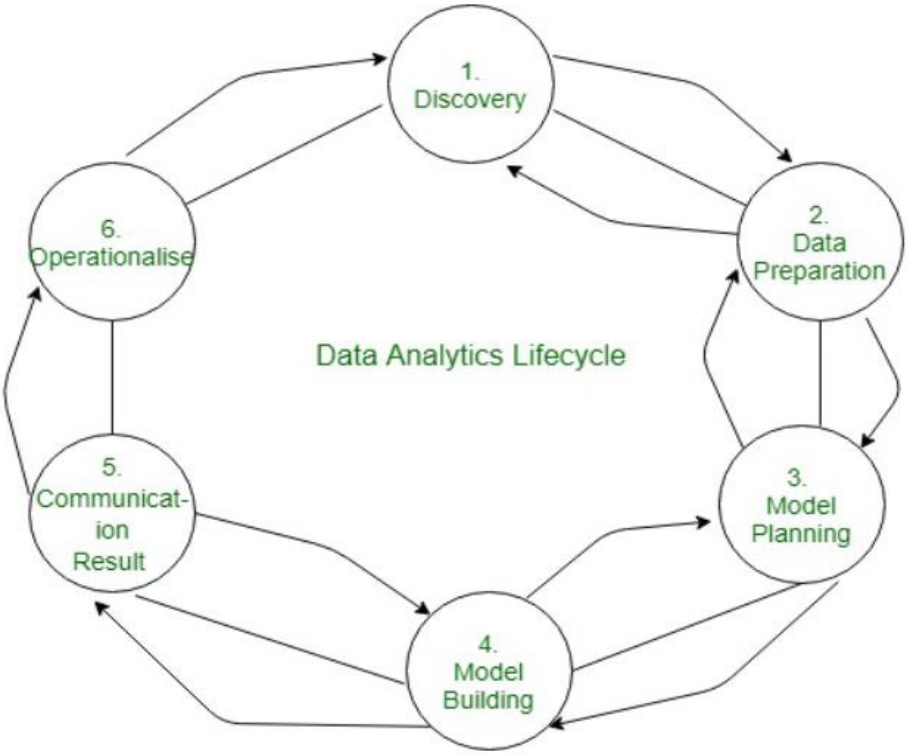
This is an emerging technology which goes further than the formerly discussed methods of Big Data Analytics as it goes beyond to recommend action routes and computes the most probable result for each decision path. This technique will not only predict possible future outcomes but it will also outline various probable futures based on the actions of the decision-maker. Prescriptive analytics need a predictive model with two components, these are actionable data and a feedback system which tracks the outcome produced by the action taken (Lepeniotia, Bousdekis, Apostolou, & Mentzas, 2020). The method seeks to answer the question, "What should we do to make this happen?" The process will involve utilizing tools to analyze data in its raw form and to assist companies to make better-informed decisions. Prescriptive analysis takes into account the information and facts about probable scenarios, the resources which are available, the past and current performance, and will propose an optimal course of action and strategy. The method can is critical in making decisions for any period from immediate to long term. Prescriptive analysis is the direct opposite of descriptive analysis method, where the latter examines outcomes and decisions after an action is done. If the method is used accurately and efficiently it will allow businesses to make informed decisions based on the predictions which will be weighted with strong facts there avoiding the mistake of making intuitive conclusions which can be wrong and in times detrimental.

The new age of data and digital networks has seen the generation of enormous amounts of data, because of the availability of high computing power this data is not out to waste but can be processed and analyzed to make data driven decisions.

2.2.1.5 The Data Analytics Life Cycle

The data analysis lifecycle involves six phases as shown in Figure 2 below.

Figure 2: Data Analytics Life Cycle



Note: Data Analytics life cycle. From: "Life Cycle Phases of Data Analytics), https://www.geeksforgeeks.org/life-cycle-phases-of-data-analytics/

a) Discovery: The data science team probes the problem to gain full understanding; they will also develop an initial hypothesis which can be later tested using the data collected.
b) Data Preparation: The data will then be inspected and pre-processed before modelling and analysis, the tasks can be performed several times and not in a defined order.
c) Model planning: The team will examine the data and explore the relationships between the variables, choose the strategic variables and then scribe the most suitable model.

d) Model building: This stage involves the teams developing the data sets for testing and the training and production processes, the team will also checks the existing tools to see if they are sufficient to run the model, also consideration is made of the environment to check if a robust environment will be required to run the model.

e) Communication Results: After the model is run, the data science team will compare the modelling results to established results to establish the criteria for success and failure. The team communicates the findings and the results to the team members and all the relevant stakeholders, the team will identify the key results and measure the business value.

f) Operationalize: The team will set up a pilot so as to fully communicate and prove the benefits of the project in a controlled manner before scaling the work across all the departments within the organisation. The process will allow teams to understand the model performance, do the necessary checks and see if there any limitations in a small environment and make the necessary adjustment before full deployment.

2.2.2 Big Data Components and Value Chain (last)

A BDA system is very complex and therefore it requires numerous components so as to function efficiently, in the context of a Small to Medium Enterprise every component is critical to ensure that the system can deliver value. The three-dimensional perspective of Big Data is viewed as Volume, Velocity and Variety. The three V's of Big Data Analytics represent the challenges and the problem which BDA seeks to solve.

2.2.2.1 Volume of data

Treatment of big volumes of data for industries for example the Global supply chains and financial systems is a difficult task as this involves handling high volumes of data generated every second. Volume refers to the size of data sets that a business collects for processing and further analysis. The sets of data are now in larger sizes in the regions of terabytes and petabytes. BDA is therefore the process of analysing the large volumes of data to develop trends and gain insights. The data storage system in a Big Data Analytics system platform will arrange the collected information in a fitting format to allow for further analysis and mining of value. The data storage subsystem provide two sets of landscapes:

- The storage substructure which must house information reliably.
- The data storage system which provide right of entry and an interface to query and analyse the immense volumes of the data.

Data storage can be divided into hardware infrastructure and data management.

2.2.2.1.1 Storage Infrastructure

The hardware infrastructure is responsible for the storage of the collected data, storage units are categorized based on the specific technology, the widespread storage technologies are the following.

• **Random Access Memory (RAM):** This is a form of data storage related with volatile types of memory; these types are bound to lose information when powered off. Emerging technologies have seen the development of static RAM (SRAM), phase-change memory (PRAM) and dynamic RAM (DRAM). DRAM is the general form of computer memory.

• **Magnetic Disks and Disk Arrays:** Magnetic discs are the main component in contemporary storage systems. A hard disk drive consists of one or more rigid and rapidly rotating discs with magnetic heads that are located on a moving actuator arm for reading, interpreting and writing data to the surfaces. Even when the HDD is turned off and costs significantly less per capacity, it retains data even if it is read or write operations are slower. Disk arrays, due to the high cost of a single large capacity disk, are configured with several disks that can increase capacities, speed and availability at lower costs.

• **Storage Class Memory:** Refers to non-mechanical storage media such as flash memory, flash memory is used to build solid-state drives (SSDs) and they differ from HDDs, SSDs have no mechanical components as such they operate with no noise, and again they have low access times and less latency than Hard Disk Drives. SSDs are more expensive per unit of storage than Hard Disk Drives. These storage structures have different performance metrics; the performance metrics can be leveraged to size a scalable high-performance big data storage system.

2.2.2.1.2 Database Technologies

Different database systems can be utilised for different sizes of datasets and applications, traditional database systems are incapable of addressing the critical issue of volumes and scale challenges required by big data.

2.2.2.2 Speed of data

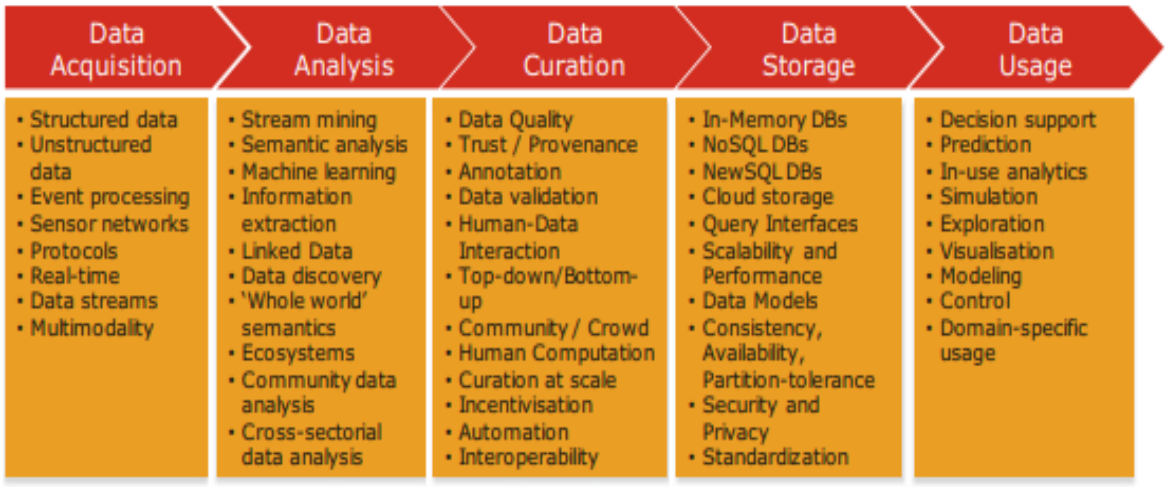
Treatment of streams of high frequency real-time data for example information from IoT, sensors and electronic data is a very strenuous practice. The speed of data in Big Data Analytics is also known as Big Data Velocity, it is the rate at which big sets of data are assimilated and processed. Due to its constant high pace of production and

frequent real-time availability, big data necessitates special processing methods; information is absorbed at this speed.

2.2.2.3 Big Data Value Chain

Spreadsheets and schemas are two examples of the various syntactic forms in which the data managed by a BDA system may be found. According to the European Commission, the data value chain will be at the heart of the knowledge economy of the future, bringing the benefits of digital innovations to more established industries like manufacturing, transportation, financial services, healthcare, and retail. This is the data value chain in Figure 3.

Figure 3: Big Data Value Chain



Note: The Big Data Value Chain. From: (Cavanillas, Curry, & Wa, New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, 2016)

2.2.2.3.1 Data Acquisition

Making collected raw data suitable for employment in decision-making and domain-specific applications is the focus of big data acquisition. In order to identify pertinent

data, synthesize it, and extract highly valuable hidden information from a commercial perspective, data analysis entails examining, altering, and modelling data. Related areas include data mining, business intelligence, and machine learning. (Cavanillas, Curry, & Wahlster, New Horizons for a Data Driven Economy, 2016). Oracle in their article Oracle Enterprise Architecture white paper suggested a three-step approach for data processing.

- A scalable storage system, such a NoSQL database or the Hadoop Distributed File System, is utilized to access and store material from many data sources. A big data analytics technique is then used to further analyse the stored data once it has been processed, arranged, and saved in a BDA program that can handle SQL. The primary function of data acquisition is to collect data from various sources and store it in a big data-capable, scalable storage system. To do this, three components are essential.
- Obtaining the protocols necessary to enable the collection of unstructured, semi-structured, and structured data streams of any kind.
- Frameworks where various protocols will be used to gather data from sources.
- The technologies that will enable the data that the frameworks would have retrieved to be stored.

The amount of data being collected from various sources these days is increasing at an exponential rate. To help businesses properly respond to inquiries and have a picture of the market and trends, all of this data has to be combined. The two primary data acquisition technologies are;

i. Protocols

Businesses that use Big Data Analytics in-house have created enterprise-specific protocols. The following open protocols are frequently employed in data acquisition:

a) AMQP (*Advanced Message Queuing Protocol*)

In order to bridge the gap and meet the need for an open protocol that could meet the demands of large companies regarding data gathering, AMQP was created. AMQP employs data types like numbers, symbols, strings, etc. and operates on four separate layers: a transport layer, a message layer, a transaction layer, and a security layer. The message processing is related to the transport layer. These messages are extracted from senders and can either be ingested by nodes or sent by relays in an AMQP network, which is made up of nodes connected by links. The transaction layer will enable coordinated outcomes of autonomous transfers and the encryption of the content of AMQP messages.

b) Java Message Services (JMS)

Enterprise messaging system messages and notifications can be created, sent, and received by Java-related programs thanks to Java Message Services. There are two message alleys available in JMS: publisher-subscriber and point-to-point. JMS and Advanced Message Queuing Protocol work together.

ii. Software Tools

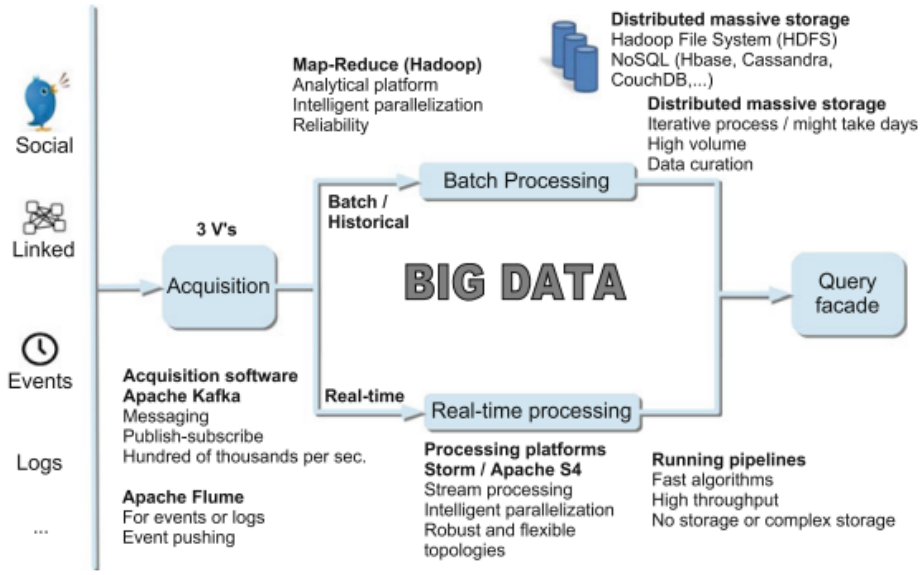
Several software are available for data acquisition, the most known are the following,

- Storm
- S4

- Kafka
- Flume
- Hadoop

Each tool's use will require a profound comprehension of the internal workings and extensive software usage skills. The schematic design Figure 4 provides a broad overview of the entire big data workflow, emphasizing the data acquisition.

Figure 4: Big Data Value Map



Note: The Big Data Value Map. From (Cavanillas, Curry, & Wa, New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, 2016)

2.2.2.3.2 Data Analysis

There are numerous alternative data formats; comparing them can be accomplished, in part, by analysing their degree of structure. More organized datasets will facilitate machine processing since semantic representations will allow for machine reasoning. "Big data analytics" is a subset of big data that focuses on organizing data to enable

accurate decision-making. Data analysis is an essential component of the Big Data value chain as a whole. Most of the retrieved data would not be in a usable state and would be worthless without big data analysis. Before raw data may be ready for data curation, storage, and use, it must first be transformed because it is frequently unstructured and contains a variety of formats. As shown in Figure 1, linked data and data discovery, machine learning and information extraction, stream mining and semantic analysis, ecosystems and community data analysis, cross-sectorial data analysis, and data quality are among the key elements of big data analysis.

2.2.2.3.2.1 Stream Mining

The process of extracting knowledge from continuously flowing, fast-moving data streams that enter a system in a stream is known as data stream mining. This is a subset of the fundamental ideas in data mining, knowledge extraction, and machine learning. With the help of stream mining, massive volumes of data may be instantly analysed with little resources. The most well known open source framework for intense data stream mining is called MOA, and it includes a variety of machine learning algorithms, including those for recommendation systems, classification, regression, clustering, and outlier detection. Because data streams are inherently temporally oriented, the distribution of samples received may vary over time. Algorithms for stream mining are lightweight and have quick computation times.

2.2.2.3.2.2Semantic-Analysis

Semantic analysis is a natural language processing technique that examines word and phrase meaning in order to determine a sentence's or paragraphs intended purpose. This is the process of taking a text's meaning and turning it into a workable action plan. Knowledge management system organizational procedures may be streamlined, consumers can be served, and operational changes can be implemented with the use of semantic analysis. Advanced reasoning techniques can be used to process data automatically thanks to Semantic Web technologies. The technologies for knowledge-level data analysis, data visualization, and heterogeneous data integration have been effectively implemented in numerous contexts.

2.2.2.3.2.3 Linked Data

A collection of design guidelines known as "Linked Data" is used to share relevant, machine-readable data. This is among the cornerstones of the Web of Data 1, commonly referred to as the Semantic Web. Best practices for building machine- and human-understandable linkages between datasets are provided by linked data. Linked Open Data (LOD) is what happens when open data—data that is freely shared and used—is combined with other data. Similar to GraphDB, RDF database Ontotext is an example of LOD; it can handle enormous amounts of datasets from various sources and connects them with publicly available data, facilitating efficient data-driven research and knowledge discovery.

Large data volumes that vary greatly in terms of encryption methods, structure, and underlying semantics must be handled through the process of big data analysis, which is an essential link in the big data value chain. Large-scale inference is one of the older methods used in big data analysis that has been redesigned to address the issues raised by the three Vs. (E.g. stream inference). It is obvious that having scalable data analysis,

inference methods, and a toolkit full of easily accessible resources is crucial for data analysis.

For social and economic gain, business models and procedures are being altered. Big data analysis is a crucial link in the big data value chain; without it, the chain would not function as a whole.

2.2.3 Data Curation

The methodological and technological underpinnings of data management that address concerns with data quality and optimize data usefulness are outlined in data curation. The active and ongoing maintenance of data throughout its lifecycle is known as data curation. Curation actions ensure long-term reuse, provide value, preserve quality, and facilitate the search and retrieval of data.

With the number of data sources and data generating platforms increasing, data curation is becoming a crucial component of data management. Content production, selection, classification, transformation, validation, and storage are just a few of the activities that make up the data curation process. The choice and application of a data curation procedure is a complex matter that depends on how standards, incentives, economics, and technology interact. This section examines the data dynamics in which data curation is employed, while also taking future requirements and developing trends in data curation into account.

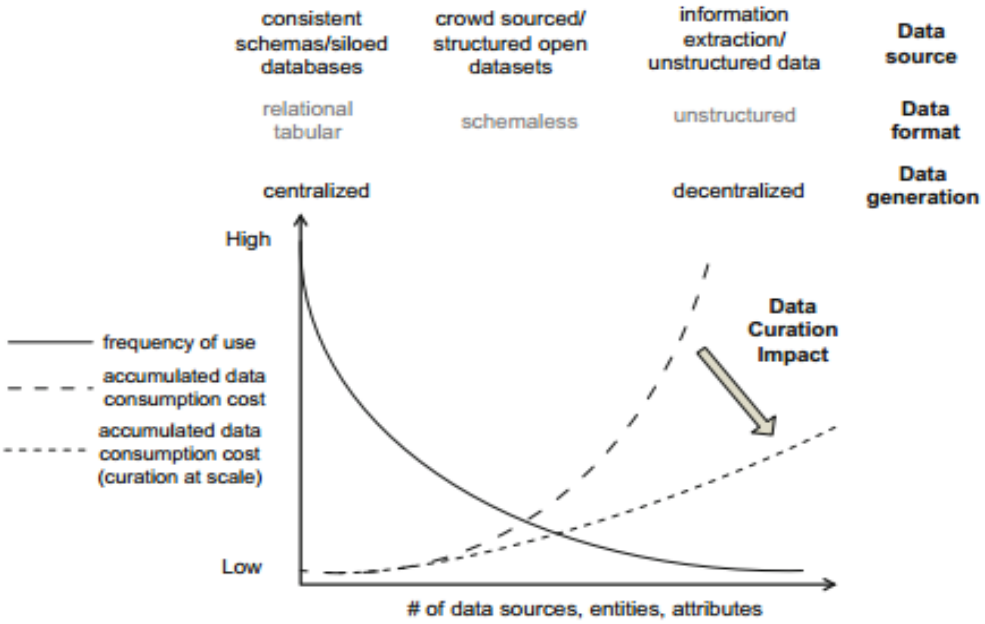
It is more expensive to solve data heterogeneity and quality issues when using data that was generated under different constraints; also, the data may not be full or may need to be transformed before it can be used.

Data re-use is the use of the feed that is received from various resources, ideally, it may be extremely difficult to collect the data but useful knowledge from can be extracted from it (Cavanillas, Curry, & Wa, 2016). Consumer data can be subjected to

data curation, which involves choosing and modifying data sources according to the needs of the data consumer. Activities related to data curation rely on the difficulties presented by big data, particularly the diversity of data. Even with tiny volumes of data, managing data variety can be expensive. The difficulties lie not so much in the sheer volume of data—petabytes these days—but rather in the diversity of the data itself, which necessitates a large investment of resources in order to make sense of it. In the context of BDA, data diversity refers to data management's propensity to handle data from numerous sources.

Data curation, once limited to tiny research houses and used to solve issues with data quality, is becoming a more popular data management activity.

Figure 5: Data Curation



Note: Big Data Curation. From: (Cavanillas, Curry, & Wa, New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, 2016)

2.2.3.1 Data Curation Platforms

2.2.3.1.1 Data Tamer

With Data Tamer, automatic data integration will take the role of existing platforms like ETL (Extract, Transform, and Load) processes. The system maps schemas and reproduces entities on-the-fly using sets of algorithms. Integration modifications will then be reviewed by crowds and human specialists.

2.2.3.1.2 ZenCrowd

Designed to bridge the gap between automated and manual linkages by optimizing the outcomes of automatic links with people, ZenCrowd was created to address the challenge of tying up and linking named things that arrive in text form to a knowledge base.

2.2.3.1.3 CrowdDB

This technology answers SQL-based queries that search engines and database management systems are unable to handle. CrowdDB allows for human-performed nebulous functions like equivalency comparison and item rating and profiling based on relevance.

2.2.3.1.4 Qurk

Qurk and CrowdDB are comparable in that they both seek to lower costs and latency through the use of techniques including batch processing, filtering, and output matching.

2.2.3.1.5 Wikipedia Bots

The technology uses what are known as "bots"—programmed algorithms—to assess the text article quality. The bots identify and highlight articles that need more in-depth evaluations by professionals.

2.2.4 Data Storage

It involves managing and storing enormous volumes of data in order to satisfy the demands of applications that need access to the data. A perfect big data storage system should be able to store an infinite quantity of data and be able to effectively manage both high-speed random data flow with many data models. Big data storage technologies are those that take care of the volume, velocity, and variety difficulties that come with big data. The standard storage frameworks, which are ineffective at storing the volumes of data that the BDA system can handle, are improved by these storage facilities. Hadoop is often regarded as the most economical and effective data storage solution. Big data storage systems use distributed shared architecture to expand into new nodes that offer processing power and storage, which allows a system to address growing storage requirements. This eliminates the challenges associated with storing large amounts of data. A storage cluster can seamlessly incorporate new machines, with the storage system managing the data distribution among nodes. In order to address the issues of velocity and variety of data, storage solutions must be devised. Speed is important because it affects query latency, or the amount of time it takes to get a response to a particular request. Big data storage is becoming a norm and corporate requirement, therefore scalable storage technology at the enterprise level that can handle infinite amounts of any business that wants to benefit from big data analytics must invest in data. In contrast to conventional database management systems that employ expensive caching techniques and row-based storage, the latest big data storage solutions offer better scalability—a crucial feature for enterprises—and lower operational complexity and costs. The application of big data storage is similar to that of standard relational database management systems for structured or semi-structured data warehouses and online transaction processing solutions (OLTP);

its main advantage is its capacity to handle large volumes of unstructured and semi-structured data. This section assesses the current state of data storage technologies capable of handling large amounts of data and identifies trends related to data storage (Cavanillas, Curry, & Wa, 2016).

Types of storage systems include:

2.2.4.1 Distributed file systems

On commodity hardware, file systems such as the Hadoop file system (HDFS) can dependable store substantial volumes of unstructured data. Even if there are other file systems with higher performance, HDFS is a crucial component of the Hadoop framework. Large data files are the target audience for this system, which can handle mass processing and faster data collection.

2.2.4.2 NoSQL Databases

The NoSQL databases uses data models from outside the relational world which do not necessarily match with the transactional properties of Atomicity, Consistency, Isolation, and Durability (ACID) (Cattell, 2010).

2.2.4.3 NewSQL Database

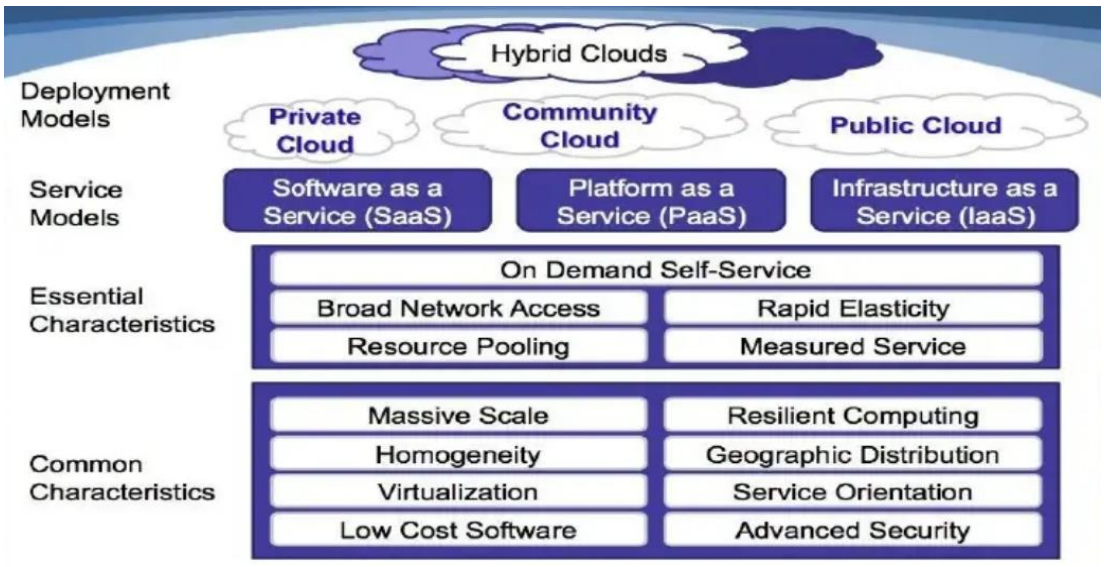
NewSQL is a modern form of relational database and it aims to achieve scalability as compared to the NoSQL databases while also maintaining the transaction guarantees of traditional database systems (Moniruzzaman, 2014).

2.2.4.4 Cloud Storage

Big data is becoming more and more influential as cloud computing grows in popularity. While some companies, like Google and Microsoft, are developing and constructing their own cloud systems, other companies, like IBM, Dell, and Cisco, are utilizing OpenStack, an open source cloud system.

Cloud storage is a distributed file system with complicated architecture (Calheiros, Ranjan, Beloglaz, De Rose, & Buyya, 28), it is installed over the cloud computing infrastructure, which is built on physically unstable, virtualized, and reasonably priced hardware. Cloud storage supports large server scale, and has efficient heavy data storage (Armbrus, et al., 2010). Cloud storage is the lower layer of cloud computing system which supports the service of the other layers above it, in addition, it is an effective way to save and manage heavy data (Al-Fuqaha, Guizani, Mohammadi, Aledhari, & Ayyash, 2015).

Figure 6: Cloud Storage



Note: Cloud Storage. From: (Armbrus, et al., 2010)

2.2.4.4.1 Data Usage

- Big data is going to have a big impact on how effective data-driven decision making becomes in the future. The decision/recommendation time perspective, which can be short- or long-term, and the databases that are appropriate for various scenarios from previous historical data to current data are the primary elements that will affect this.

- The emergence of new markets is greatly impacted by data-driven applications. The requirement for fresh, crucial insights into the application of big data could stand in the way of that progress. Predictive analytics applications with notable examples include:

Proactive maintenance based on sensor and contextual data to predict deviations from standard maintenance intervals, if the data shows that the system is stable, maintenance intervals can be extended, resulting in lower maintenance costs (Cavanillas, Curry, & Wa, 2016). When data highlights problems prior to the completion of planned maintenance, breakdowns, repair expenses, and downtime can be avoided, potentially saving even more money. In addition to sensor data, information bases will also contain other contextual and environmental data, such as machine utilization statistics (e.g., heavy load). Leading manufacturers are altering their business models and investing in new infrastructure themselves (to achieve economies of scale) and increasing their customers' we rent equipment because predictive analytics depends on new sensor and data processing infrastructure. Due to the evolution of outdated production methods, the existence of numerous sensors and other data streams, and other factors, Industry 4.0 is becoming more and more popular in the manufacturing sector when it comes to the usage of cyber-physical systems, the urgent necessity to leverage usage scenarios and communication networks to connect all data. It results in possible advantages. Industry 4.0, which brings IT into manufacturing, creates a number of difficulties for IT support.

The challenges include planning and simulation, monitoring and control, interactive use of machinery, logistics and enterprise resource planning (ERP), predictive analytics, and ultimately prescriptive analytics that can control the decision-making process (Cavanillas, Curry, & Wa, 2016). Data analysis is used to accomplish this

automatically, an extension of the Industry 4.0 scenario previously outlined, smart data and service integration centres on services that address the current goal. Technical and organizational factors must be taken into account in order to employ smart services to handle big data utilization challenges. Regulations, new legal obstacles (including ownership issues for derived data), privacy and data protection issues, and regulatory issues must all be handled. To enable service contact, multiple technical requirements must be met. at the hardware level, encompassing networks, individual computers, and facilities. Conceptual level encompassing intelligent systems, smart gadgets, and decision-making. From IaaS to PaaS, SaaS, and new big data services, infrastructure and even business processes and knowledge as a service. Interactive Exploration When dealing with large amounts of disparate data, underlying models of functional relationships are often missing (Cavanillas, Curry, & Wa, 2016).This implies that there is a greater need for dataset exploration and analysis among data analysts. This is handled by data visualization and new dynamic visual analysis methodologies, but it also calls for a new user interface with additional features for data exploration. Businesses may evaluate multiple studies, different parameter settings, and competing models with the support of the unified data consumption environment. Assistance with decision making, these methods are used by decision support systems that are now in use insofar as they are based on static reports, but they are used enough dynamically to fully utilize exploratory search. Nonetheless, these clusters comprise the subsequent corporate goals, arranged according to intricacy:

• **Search:** Only a multitude of reasons require the retrieval of data at the lowest level of complexity. This covers retrieving facts and looking for things that are known.

• **Learning:** At a higher level, these capacities facilitate the gathering of knowledge and the interpretation of data to facilitate comprehension. Features including data

consolidation, aggregation, and comparison are supported. Additional elements could facilitate data sharing through social functionality. Simple searches for particular items, such celebrities or their use in retail advertising, are examples of learning exercises. Applications for big data search are supposed to look through all pertinent data and provide an integrated perspective.

• Research: At the pinnacle of decision support systems, information can be gathered, processed, and combined. This covers tools for evaluation as well as support for negations and exclusions. At this stage of analysis, tools impact forecasting and planning and help real-world discoveries. Higher up, his study (finding) looks for significant correlations, such the impact of the weather or the season on a product's sales during a specific event. Furthermore, there are instances of big data being used for strategic business decision-making, particularly at the highest levels. These tasks can also be (partially) automated at a higher level to produce prescriptive and even predictive analytics. The latter describes decisions that are automatically determined and put into action based on the outcomes of automated (or manual) analysis.

These capabilities, which include complex event processing (CEP), are more likely to be present in the environment and are outside the purview of conventional decision support systems. Depending on the circumstances of your application, security frequently needs regular upkeep, like the aerospace sector. Determining the intervals between maintenance is a strictly economic matter, though, as there are instances in which the cost of a machine breakdown is not catastrophic. Predictive analytics is predicated on the idea that time can be anticipated more correctly if you have enough sensor data about a specific machine and a sizable enough database of sensor and error data for that machine or a similar type of machine.

This approach offers the following cost savings:

• As planned maintenance time is achieved, longer maintenance intervals are avoided as "unnecessary" disruptions to production (or activity). Based on available sensor data, predictive algorithms can increase the length of maintenance intervals.

• Fewer failures occur earlier than planned maintenance, thereby lowering the total number of failures, thanks to sensor data and predictive maintenance that necessitates early maintenance.

• By anticipating potential failures and providing enough prior notice, predictive maintenance can lower failure costs, schedule maintenance or replacement, and minimize downtime. One crucial part of using big data is visualizing the analysis results, which includes using the right visualization tools to show trends and other predictions. A crucial component of data mining and machine learning is feature selection, subset selection, and numerous cycles are needed to evaluate various configurations. A fast and precise evaluation of the quality of the results is made possible by high-quality visualization, since the settings are assessed in light of the analysis results that are provided. By contrasting the outcome with other outcomes on the test dataset, the predictive quality of the model is verified. Visualization is a crucial component of data analysis because, without it, this can be a costly and time-consuming procedure. In order for data scientists or corporate decision makers to be able to draw conclusions from the analysis using the results of the analysis in later phases in a data usage scenario, for instance, you need to The chosen visual display may hold significance both controllable and efficient.

It may be computationally demanding and prohibit interactive use of the visualization, depending on its complexity. However, in many big data use cases, exploratory search is crucial to the analytical outcomes. Sometimes, like with an airplane engine, the

outcomes of large data analysis are limited to a single instance. However, analytical datasets frequently require interactive exploration and interpretation since they are as complicated as the underlying data, stretching the boundaries of traditional statistical display techniques.

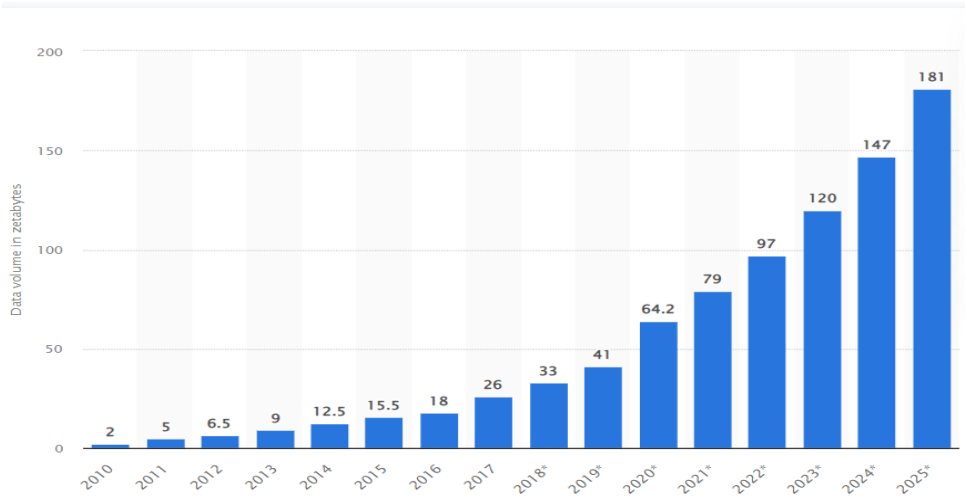
Shneiderman's seminal work on visualization (Shneiderman, 1996) identifies seven types of tasks: overview, zoom, and filter, details on demand, retrieval, history, and extraction. Data models, which are different from standard data mining and reporting applications and are utilized in many machine learning methods, are another area of visualization. Such data models are evaluated against well-understood data sets before being employed for tasks like classification, clustering, recommendation, and prediction. Model configuration and validation are aided by visualization. Lastly, the enormous dataset sizes are a persistent issue for visualization tools, fuelled by advancements in GPU and display technology as well as the sluggish uptake of immersive visualization settings like virtual reality, augmented reality, and caves. The disciplines of science and information visualization both address these issues.

2.2.5 Data analytics and its importance:

The internet of things and other advancements in the IT industry has allowed for the capturing of data, storage and transmission of information (Loebbecke & Picot, 2015).

Figure 7 below shows the evolution of data;

Figure 7: The evolution of Data Storage



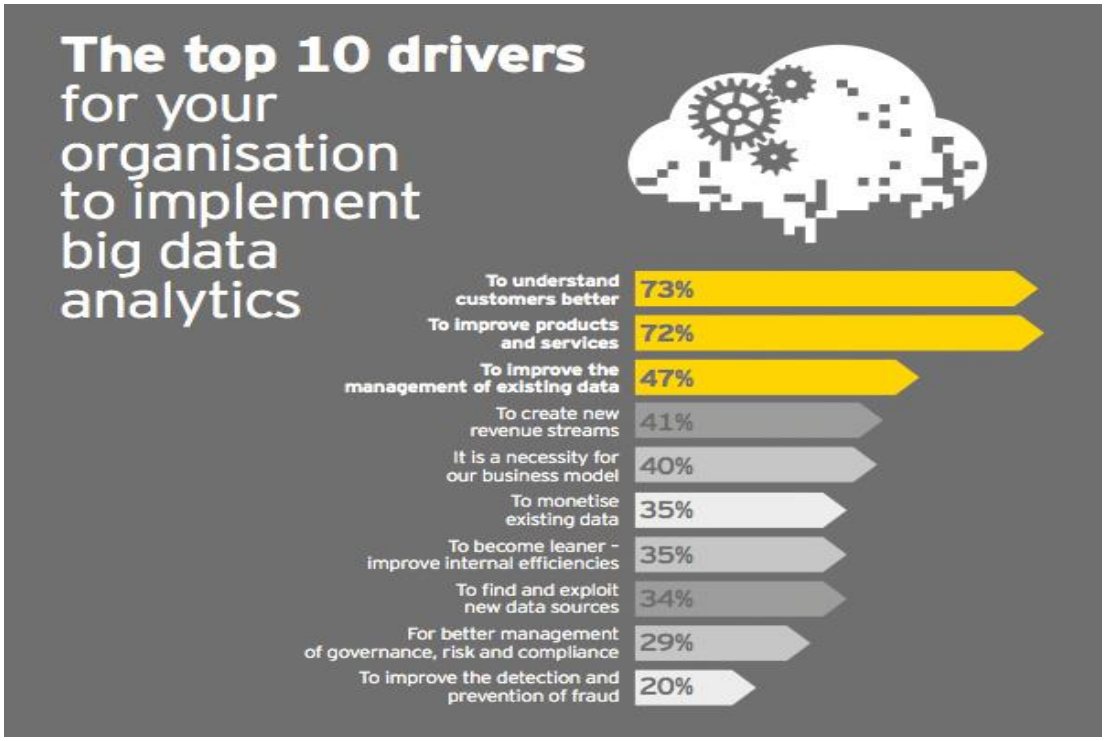
Note: The growth of data storage. From: *Note:* Cloud Storage. From: Keizer, J. A., Dijkstra, L., & Halman, J. I. (2002). Explaining innovative efforts of SMEs.: An exploratory survey among SMEs in the mechanical and electrical engineering sector in The Netherlands. *Science Direct*, 23.

According to Statistica, over the next five years up to 2025, global data generation is expected to grow to more than 180 zettabytes. The installed base of storage capacity is expected to rise at a compound annual growth rate of 19.2% in tandem with the growth of the data volume. Unstructured data has a vast amount that is always expanding, and it contains value that can be used.

2.2.6 Drivers of BDA

Big Data is today's biggest buzzword, and with the quantity of data generated every minute by consumers and organizations around the world, Big Data analytics holds a huge potential (Gandomi, A. and Haider, M, 2015)

Figure 8: Drivers of Big Data Analytics



Note: EY: Only third of businesses use big data strategically. From: https://www.consultancy.uk/news/1375/ey-only-third-of-businesses-use-big-data-strategically

2.2.7 Factors to consider when selecting an Analytics solution

2.2.7.1 Internal Factors

The following comprises the factors which should be considered when choosing the solution to use for a SME firm;

a. The expected outlay costs appropriate to achieve the analytics goals.
b. The elaborate strategy and the analytics road map to realize the goals.
c. A forecast of the expected payback of the investment.
d. The needs of the company from a data and analytics perspective, that is, what are the targets does the solution seek to reach.

e. Stage of the organisation’s analytics maturity? Capability assessments, does the infrastructure support the solution (Krol & Zdonek, 2020)

These factors will point out to the overall preparedness of the organisation from an internal standpoint.

2.2.7.2 External Factors

a. What analytics frameworks are the competitors using now, and how can the company leverage its own analytics resources to counter and become more competitive in your market?
b. Any laws (prohibitive) which relate to the type of data to be collected and analysed.
c. The plan which is available to create benefit for the customers in the present and in the future.

2.2.8 Value Creation potential of BDA

2.2.8.1 Understanding Customer Needs

In order to preserve or grow profit margins and stay competitive, businesses are implementing more and more targeted marketing strategies. Predictive models are therefore frequently utilized in precision marketing to comprehend and satisfy client demands and expectations. Predictions derived from transaction records and customer data are being used more and more to analyze consumer behaviour and preferences and adjust product supply chains (SCs). In the current economic climate, building business relationships and comprehending client needs are essential to a company's growth. In the business sector, the phrase "people by from people" (BDA) is significant since it helps SMEs comprehend their customer base by anticipating and predicting consumer desires.

The SMEs will subsequently be able to interact with potential customers and enhance their value proposition through micro-targeting. By identifying their top consumers, developing customer profiles, and gathering as much data as possible, it aids organizations in understanding their clients. Utilizing big data analytics tools, the gathered data is analysed, and the insights obtained are then applied to enhance the client experience. SAS claims that by using analytics to gauge consumer satisfaction, big data analytics helps businesses to provide customers what they want, when they want it. Additionally, big data analytics allows more businesses to create cutting-edge new goods that adapt to their customers' changing needs. According to Forrester, businesses risk being exposed if they do not figure out how to use customer insights and measure the financial benefit of their customer experience (CX) initiatives. According to a different Adobe study, businesses deemed "CX leaders" have a threefold higher chance of succeeding in their objectives. Forrester, Adobe, and other forecasters are right about CX, but many forecasts are only a stand-in for actuality. Every industry, from retail, logistics, and healthcare to banking and insurance, benefits from having a strong customer experience.

Big data is necessary to build the scalable, frictionless, and human-centred tailored experiences that customers need. In order to better understand their customers and enhance the customer experience, businesses are investing more and more in intelligent CX technologies that make use of AI, machine learning, and sophisticated analytics.

2.2.8.2 Improving Operational Efficiency

By utilizing the data at hand to analyse customer behaviour, business Big Data Analytics (BDA) plays a pivotal part in the expansion of organizations. This process aids in the application of predictive analytics to address essential issues that affect the

customer experience. Big Data analytics assist companies in analysing the trends in the market and will suggest optimum inventory levels and management of supply chains (Seyedan & Fereshteh, 2020), In turn, this will increase operational effectiveness. Business operations now include information systems that use data to guarantee timely delivery, cost-effective ordering, etc. By analysing customer behaviour based on purchase data and assisting with the use of predictive analytics to determine the average checkout wait time, big data management increases operational efficiency. Operators use data and analytics to save operating and maintenance costs, boost productivity, prolong asset life, and improve operational efficiency, according to KPMG. By employing analytics to monitor customer requirements and satisfaction, big data analytics enables organizations to provide customers what they want, when they want it.

More businesses now have the chance to create cutting-edge new items that cater to the shifting wants of their clientele thanks to big data analytics. To sum up, big data analytics may assist businesses in increasing operational efficiency through cost reduction, staff productivity, supply chain optimization, and behaviour analysis of customers. Businesses can build new and inventive goods, enhance existing services, and tailor consumer experiences by utilizing customer information.

2.2.8.3 Improved Decision Making

Big Data analytics gives businesses insights into factors like consumer behaviour and market trends, which helps them make better and more informed decisions. It also helps businesses find new opportunities, enhance their offerings, and find new markets for their operations. All of these benefits will help businesses, especially Small and Medium-Sized Enterprises, expand and grow.

2.2.8.4 New Revenue Streams

Through insights into consumer behaviour and industry trends, Big Data Analytics helps businesses find new revenue streams. By meeting or satisfying customer wants, businesses can create new products or services that bring in additional revenue.

2.2.8.5 Competitive Advantage

By employing the insights derived from a business diagnostic analysis (BDA), small and medium-sized enterprises (SMEs) can set themselves apart from their rivals, thereby meeting client demands and expanding their market share.

2.2.8.6 Better Customer Experience

By utilizing insights into consumer behaviour and preferences, BDA assists organizations in enhancing and providing a better customer experience. This would enable businesses to better customize their offerings to match the wants and needs of their clientele, which will boost client happiness and loyalty.

2.2.9 Barriers to Big Data Analytics Adoption for Zimbabwean SME's

2.2.9.1 Data barriers

Big Data analytics is a potent tool that supports businesses in making data-driven decisions that improve operations. BDA entails a number of dangers that must be managed to ensure the data's safety and security. Data storage, scalability, complexity, and quality are common issues with data that prevent small and medium-sized businesses from implementing big data analysis.

2.2.9.2 Knowledge and skills barriers

Large-scale data analytics will require knowledgeable employees who can analyze and comprehend the data. Inaccurate analysis and bad decision-making may arise from the SME sector's lack of qualified workers in the field.

2.2.9.3 Regulatory barriers

Regulatory barriers are those aspects of law, ethics, and privacy that organizations need to be aware of in order to comply with data protection laws that specify how companies must use data. Many large firms rely heavily on the sensitive data they collect, and if it falls into the wrong hands—such as cybercrime or hackers—it may seriously harm the company's operations and reputation. Consequently, businesses should prioritize safeguarding the security and privacy of their data against malevolent intrusions. Big data is difficult to carry around in a pocket; businesses must manage large servers to store and safeguard this sensitive data. Businesses must undertake this difficult and dangerous procedure in order to safeguard their large data.

2.2.9.4 Technical barriers

These speak of the setup needed to make BDA possible. They are made up of software and hardware, and in order to enable cloud-based services, a high-bandwidth internet connection and security measures may be necessary.

2.2.9.5 Organisational barriers

These speak to problems with the structure, culture, and lack of a strategy within the organization as well as the support of upper management. Financial barriers that a small and medium-sized enterprise (SME) faces include the price of acquiring Big Data Analytics technology and the resources needed to deploy and utilize it. Big Data analytics can help businesses increase sales and income by offering insightful knowledge about things like consumer behaviour and industry trends. However, careful planning and execution are needed for the adoption and use of a BDA framework. Small and medium-sized businesses must take into account the following elements when putting a Big Data Analytics framework into practice:

1. **Data Quality**: Small to Medium Enterprises should ensure that the data they collect is accurate consistent and complete.
2. **Data Security**: Data should be secure from unauthorized access.
3. **Infrastructure**: This will include the hardware, software and network infrastructure.
4. **Skills**: This will include skills in areas like data management, data analysis, and data visualization.

2.2.10 Case Study: Use of BDA in Supply Chain Management

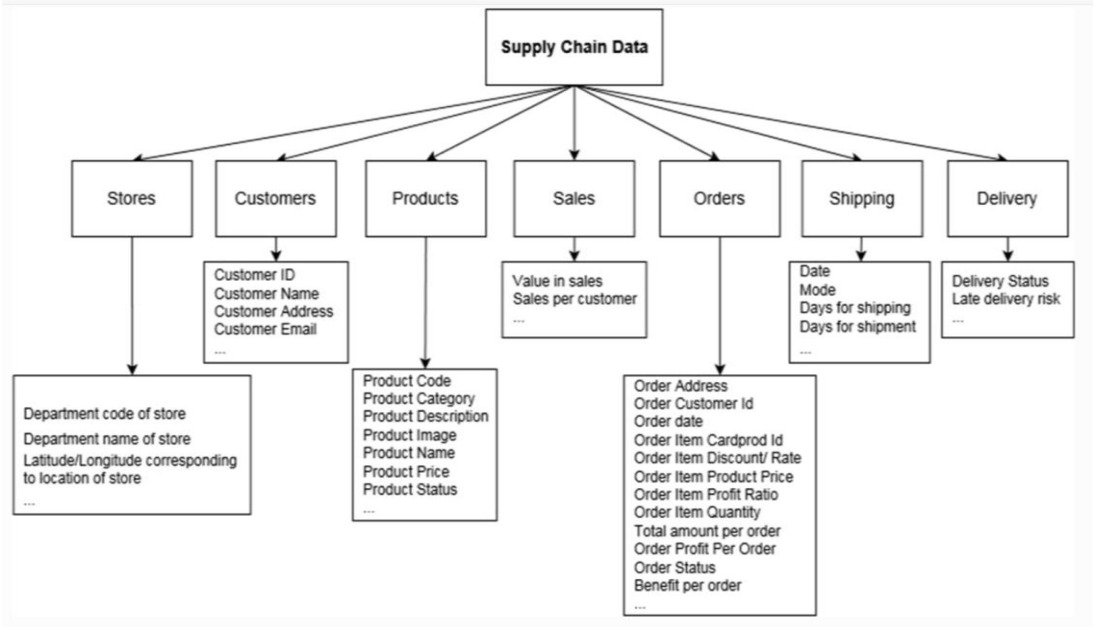
Supply chain management (SCM) focuses on the flow of goods, services, and information from origin to customer through a series of interconnected entities and activities (Kramer , Bitsch , & Hanf, 2021). The presumption that capacity, demand, and cost are well-known and constant variables presents a significant hurdle in SCM problems. In actuality, though, this is not the case because of unknowns brought on by changes in lead times, organizational risk, supply transportation, and customer demand. Specifically, demand uncertainty affects SC performance the most and has a broad impact on transportation, inventory, and production planning. Demand forecasting is a crucial strategy for handling supply chain unpredictability in this way SCM demand forecasting makes use of a number of statistical analysis methods, including regression and time series analysis. Big Data Analytics (BDA) can better represent consumer needs, facilitate the evaluation of SC performance, raise the efficiency of SC improvement, and decrease response time thanks to the growth of information technology and greater processing efficiency. It has become popular as a means of making more precise forecasts. Demand forecasting in today's ever growing and irregular global supply chains necessitates the use of big data analytics (and machine learning) techniques due to the nature of demand data. Because there are many suppliers, products, and customers at various points in the chain for various

purposes, supply chain data is highly dimensional and extensive (items, supplier capabilities, orders, delivery, customers, retailers, etc.). Demand forecasting based on finding statistically significant trends (defined by mean and variance features) has replaced more conventional (statistical) methodologies. Several organizations have been forced to adjust to the constantly shifting demands in the supply chain by cleverly utilizing past data. Big data analytics approaches are used to generate this capability by finding the underlying links between demand data across the supply chain network and extracting prediction rules. Processing these methods calls for sophisticated machine-programmed algorithms and is computationally demanding. Machine learning/data analysis algorithms enable accurate (data-driven) demand forecasting and align supply chain activities with these projections to boost efficiency and customer satisfaction, since supply chain management (SCM) efforts strive to meet customer demand while minimizing the total cost of supply.

2.2.10.1 Data in Supply Chains

Data can be grouped into customer, shipping, delivery, order, sales, store, and product categories in the context of the supply chain. The supply chain data classification is displayed in Figure 9. As a result, a variety of (and disaggregated) sources, including manufacturing, sales, inventory, warehousing, and transportation, provide the SC data. In this way, established projections may underestimate or overestimate demand due to shifts in the level of competition, price volatility, technical advancements, and consumer commitments. Thus, to improve demand-forecasting accuracy, supply chain data must be carefully analysed to learn more about suppliers, customers, industry trends, and technological advancements. Supply chain expenses can be reduced by identifying trends and patterns in such data and applying them to enhance the precision of forecasts made in the future.

Figure 9: BDA in SCM



Note: Big Data Application in Supply Chain Management. From: Seyedan , M., & Fereshteh, M. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Springer Open*, 2.

The analysis of supply chains is now very complicated this is due to;

- The diversity of supply chain structures based on product homogeneity or heterogeneity.
- The overall increase in diversity of supply chain units.
- The interconnection among these units.
- Inaccuracies in these components' dynamic behaviour;
- Insufficient knowledge of the Supply Chain units.
- Increasing synchronization and coordination among manufacturing departments through networking to achieve high degrees of customisation and adaptability to various client needs.

- Supply chain digitalization techniques and block chain-driven technological adoption (use of technology).

Procurement management (supplier selection, cost reduction, risk management, product creation, production scheduling and control, quality assurance, maintenance, etc.) can be enhanced with the application of big data analysis (BDA). SC management, warehouse selection, diagnosis, and other techniques are being used more and more. Demand management (such as demand forecasting, demand capturing, and demand shaping), inventory management, logistics/transportation (such as intelligent transportation systems, logistics planning, and in-transit inventory management), Demand forecasting in particular is one of the key uses of BDA in SCM for precise forecasting.

2.2.10.2 BDA for Forecasting and Demand Management in SCM

In reality, the phrase "demand management" first appeared in the late 1980s and early 1990s. Demand management is often approached in two ways, backward-looking strategy depends on the ability to meet demand from the past or present, whereas a forward-looking strategy considers possible demand over the coming years. The three main areas of concentration for Future Demand Management are marketing strategy, data management, and demand forecasting and planning. Predicting the volume and timing of client requests is known as demand forecasting and planning. By promptly satisfying consumer requests, this kind of forecasting seeks to increase customer happiness. When resources are allocated according to demand, industrial processes (as well as related supply chains) become more robust and efficient, which lowers waste and inventory.

Several methods for demand forecasting and planning have been put forth in the literature and in actual practice in light of the aforementioned facts.

These strategies include benchmark-based evaluations, statistical techniques (such moving averages), and spreadsheet models. The most popular tool for demand planning and forecasting nowadays is Excel. Spreadsheet models used in demand forecasting are sometimes plagued by the inability to handle massive data sets. Furthermore, straightforward statistical methods like moving averages or exponential smoothing are unable to extract, analyse, or resolve the complexity and uncertainty of supply chain management (SCM)—that is, the diversity and unpredictability in supply and demand. Traditional SC demand forecasting and planning methods have had a difficult time cutting expenses and inventory levels during the last ten years.

The suggested technique raised SC expenses at suppliers' expense even though it sometimes improved daily payments. Large-scale data processing is now simple, quick, and efficient thanks to the advent of big data and advanced computer analytics, which has allayed worries about data collecting and storage via cloud services. There are now more options for data-driven demand forecasting and planning thanks to the development of new technology in data processing and storage as well as the quantity of high-quality data. Using data mining algorithms and tools that can mine the data, analyse the findings, and discover more about the linkages involved has now significantly increased the accuracy of demand forecasts.

2.2.11 Framework for Implementing a BDA system

SME’s tend to benefit from BDA by improving performance and also having the ability to make informed and data-driven decisions. Implementing BDA systems is complex and can be also risky to implement however a unified model for adoption can allow for smooth on-boarding (Maroufkhani & Wan , 2020). The technology, organization, and environment (TOE) model is combined in this model, which may be incorporated into an organization's strategy to improve its performance in the market

and financially. The following model, which includes the following crucial steps, will be applied in the context of Zimbabwe:

a) An evaluation of the company's data management procedures and systems; an identification of the data sources; an audit of the instruments used to collect the data and the data's information quality.
b) Evaluating the data storage methods.
c) Being aware of the industry
d) The company's product line.
e) Analysis of competitors.
f) The supply chain.
g) Margin of products
h) Define the performance metrics that will be assessed
    - Specifying the KPI (sale volumes, revenue, etc.) that will be monitored.
    - Determining the performance as of right now.
i) Appropriate BDA Technology Selection:
j) The right BDA technology is chosen in accordance with the business's goals. The infrastructure of software and hardware will be a part of the technology.
k) Establish a system and specify the procedures for gathering, storing, and processing data.
l) User training: Educating and preparing people to manage and comprehend data.
m) Monitoring and Evaluation
n) Keeping track of the KPIs that have been identified and adjusting as needed to meet the goals. The framework is not all-inclusive; adjustments and other

enhancements may be made in accordance with the needs of the business and in reaction to environmental changes.

2.3 Relevance of the Theoretical Framework to the Study

The theoretical framework outlined the foundation to the research and was guided by the objectives set out for the study, that is, the theory was structured so as to expand on the research objectives. The section analysed in depth the components of a Big Data Analytics system touching on the software and the hardware requirements which entrepreneurs or SME's require to on-board a Big Data Analytics system. The theoretical framework also evaluated how the implementation of a BDA system can rev up and drive growth of a Small to Medium enterprise. The section proceeded further to analyse the risks which are encountered when implementing the system and the implementation framework thereof. The chapter elaborated on all the theory on Big Data Analytics and cited applicable examples of application in several industries so as to understand the applicability and the usefulness of Big Data Analytics in business and also in the context of Small to Medium Enterprises.

Summary

Establishing specific performance indicators should be based on a variety of organizational situational elements, including the organization's objectives, its external environment, and its internal structure. When thinking about leveraging Big Data Analytics to improve process innovation capabilities, it is crucial to remember that the strategy needed to get the best outcomes varies depending on the various BDA combinations and applications. BDA Processes that are efficient in carrying out activities like continuous diagnostics, action planning, and BDA implementation and assessment are necessary to support organizational decision-making for growth. According to Lewin's theory of organizational development (OD), processes serve to

impart knowledge and skills to an organization, with the main goal of the process being to enhance the organization's capacity for problem solving and change management. According to Burckhardt, organizational dynamics encompass internal dynamics, wherein a group of people collaborate to enhance organizational efficacy, capacity, and work output, as well as the adaptability of culture, policies, practices, and procedural needs.

Three ideas make to the definition of organizational development (OD), which is defined as the system-wide application and transfer of behavioural science insights into the planned development, improvement, and enhancement of strategies, structures, and processes that lead to organizational effectiveness. Shared beliefs regarding rules, practices, and processes are part of organizational culture, which is defined as the atmosphere or distinctive character of an organization. Leadership, communication, participative management, and role clarity are other cultural traits. An organization's shared basic beliefs, values, norms, behavioural patterns, and artefacts—visualized as an assumption pattern—that a group has acquired via resolving challenges with internal integration and external adaptation make up its organizational culture.

Four components make up organizational capabilities (OC): resource acquisition, organizational structure, production subsystems, and performance. OC describes an organization's capabilities, such as creating goods and services and upholding organizational procedures. The performance of an organization is influenced by its climate and culture. For this reason, in order to successfully implement a Big Data Analytics system, organizations must foster a positive culture and be open to accepting change and reaping its benefits.

CHAPTER 3 METHODOLOGY

3.1 Introduction

The section evaluated the selected procedures and approaches of gathering the data and streamlined this data to obtain convincing and satisfactory results to answer the research objectives. The methodology is an important part of the research because it defines how the research will be conducted and how the data will be obtained and analysed to achieve a satisfactory result (Bryan & Bell, 2011). The chapter covered the research design, sampling techniques which were employed, the data collection instruments and the analysis.

3.2 Research Design

The research used the descriptive Qualitative Research Approach and used selected companies within the SME sector to explore and find out the answers to the research questions, the qualitative research design was chose and was found to be necessary in the context of the study. The selected pool of companies were contacted to explore the importance and challenges in the adoption of BDA among SMEs in Zimbabwe.

3.3 Population and Sampling

3.3.1 Population

The population encompassed all the Small and Medium Enterprises in Harare and its environs, the unit of analysis were the owners and the managers of the Small to Medium Enterprises. The target for the study was drawn from the following industries, Agriculture, Consultancy, Engineering Services, Manufacturing and Mining, Personal Care and Retail and the estimated population for the formal and established SMEs in this sector was estimated to be 1,000 in the Harare Metropolitan region.

3.3.2 Sample Size

A sample size of 100 companies was pulled up for this study using convenience sampling where the companies which are in Harare and its environs were considered for easier engagement and accessibility. The sample of 100 companies was used which is 10% of the overall population which make the sample representative for the purposes and intent of the study (Anokye , 2020). A computation from Raosoft (http://www.raosoft.com/samplesize.html) for the computation of the sample size gives a sample size of 214 for a permissible response rate of 50% at a confidence level of 90% and a margin of error of 5%. Convenience sampling methods place primary emphasis on generalizability (i.e., ensuring that the knowledge gained is representative of the population from which the sample was drawn) (Etikan, Musa, & Sunu, 2015).

3.4 Data Collection and Instruments

Data collection refers to the process of gathering information, the data collected answered the research questions posed in Chapter 1. The data was collected using Qualitative Data Collection procedures, that is interviews and questionnaires where pertinent questions were framed and shared to the respondents.

3.5 Data Collection Procedure

The researcher focused on getting qualitative data from the target groups, in-depth interviews looked for quality and in-depth answers which looks deeply into the respondents' experiences, opinions and feelings. A questionnaire was used to collect data from participants in the research, the questionnaire solicited to capture questions to answer the critical questions to the research namely;

- Concept familiarity.
- General preparedness of a firm to adopt Big Data Analytics.

- Evaluation of data storage mechanisms (databases and quality of data).
- Willingness to invest in the system.

Information was collected from Owners and in cases where owners were not available to respond Managers were used. Interviews were also used to evaluate and have better understanding of business processes and explain in depth the features, benefits and quick and long term wins realisable through the adoption of Big Data Analytics for the specific organisation.

3.6 Analysis and Organisation of Data

The collected data from the survey was first cleaned and validated, incorrect responses and erroneous responses to specific questions where removed. The data was processed and presented in the form of graphs (pie charts and tables) to show the relevant trends and patterns from which pertinent relationships and correlations were drawn.

3.7 Ethical Consideration

The process was conducted in a rational manner without any cohesion or solicitation to favour or sway the respondents' answers, no imposition of views was done in the process.

3.8 Summary

The design plan was carefully integrated to allow for the coherent collection of data ensuring that the results collected best portrayed the applicability of the framework to the SME's in Zimbabwe (results scalability)

CHAPTER 4 DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.1 Introduction

The section presents the findings and interpret the results of the study. The data which was collected is presented in the form of tables, pie charts and graphs. Data will be presented and analysed followed by a discussion of the trends as revealed by the data. Not all the respondents which were given the questionnaire managed to answer and complete all the questionnaires and the response rate recorded for the study was at 68%. The chapter ends up with a summary.

4.2 Data Presentation and Analysis

4.2.1 Response Rate Evaluation

Table 1: Research Response Rate

Number of Questionnaires Administered	**Number of questionnaires completed**	**Percentage Response Rate**
214	146	68%

The response rate was lower than the expected target of 100% due to various factors in the target population which include;

- Access to internet – this factor affected responses to the online questionnaire where many respondents cited the aspect as a deterrent, the physical questionnaire covered the gap for respondents which could be accessed in Harare

- Access to a suitable gadget to complete the questionnaire – some respondents did not have the gadgets which could permit them to complete the questionnaire.

The physical distribution of the questionnaire allowed for explanation by the researcher line by line for the respondents to fully understand the questions and thus respond accurately, for the online respondents diligence was done to explain the purpose of the survey and clarifying some grey areas to also ensure accurate responses. Responses from a mix of industries were solicited to cover the whole range of businesses which fall under the Small to Medium Enterprises sector, these have been categorized into the following.

Table 2: Industry Categorization

Industry/Category	**Areas of Specialization**
Agriculture	• Horticultural Farming, Poultry and Cattle Fattening
Consultancy	• Logistics procurement and logistics consultancy • Software and information systems
Engineering services	• Garage equipment installations. • Electrical engineering
Manufacturing and Mining	• Small scale mining • Manufacturing
Personal Care	• Hair and Cosmetics
Retail	• Plastic ware, hardware, motor spare, clothing • Cake business

Table 2 shows the main categories of the respondents which gave feedback to the questionnaire.

Figure 10: Distribution of Respondents by Industry

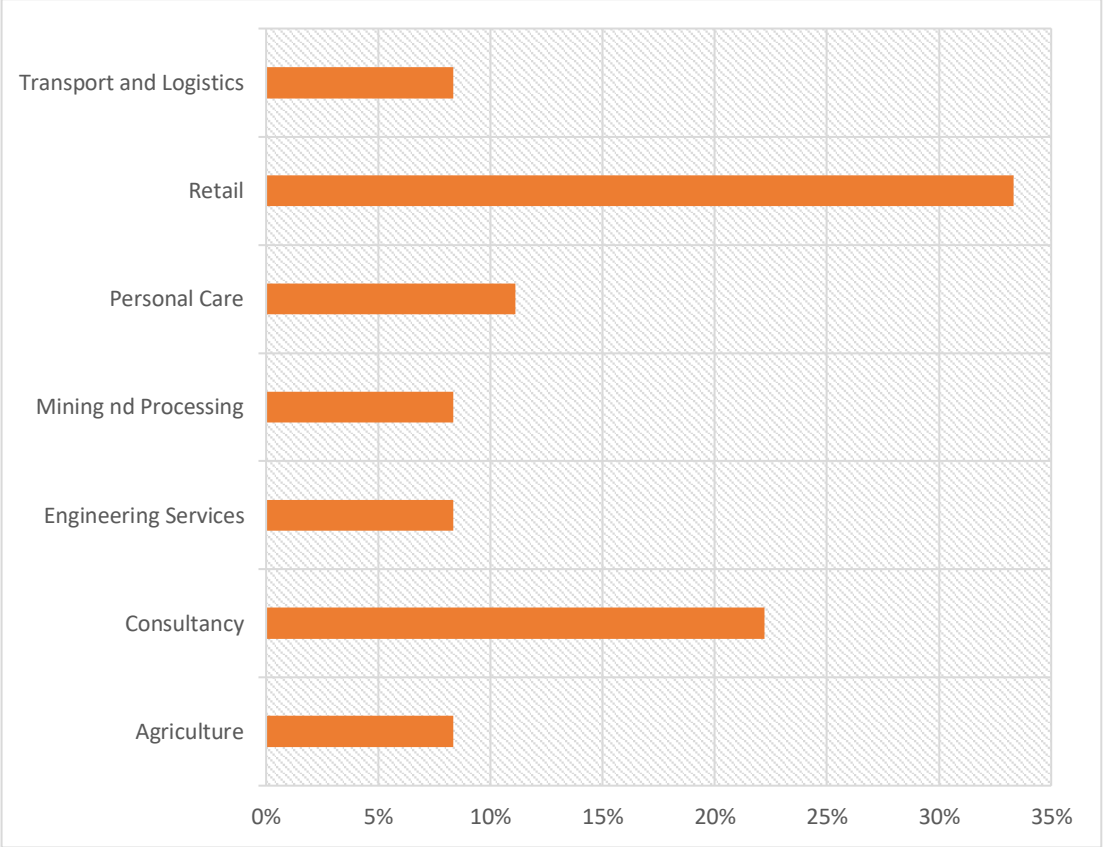
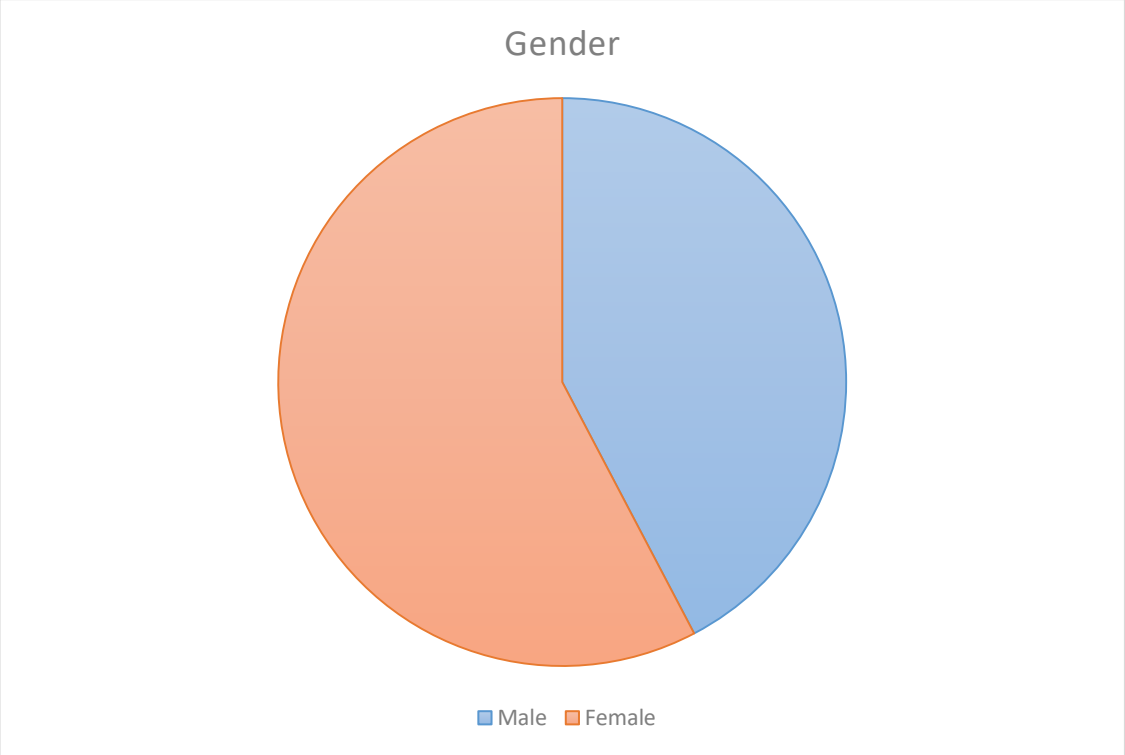


Figure 10 shows the demography of the frequencies of the seven broad categories of the companies which were analysed from the questionnaires received. The retail sector is the most dominant (33%) followed by the consultancy field with a frequency of 22%.

4.2.2 Demographics

4.2.2.1 Gender Demography

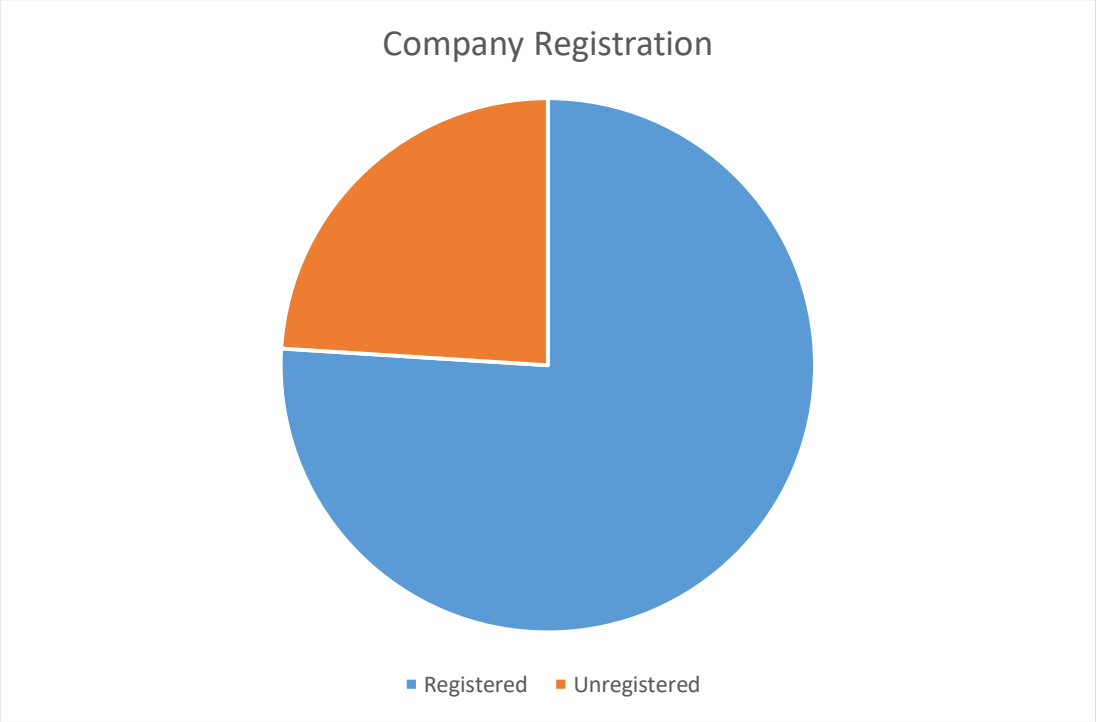
Figure 11: Gender Demography



From Figure 11 the data shows that 57% of the respondents were female and 43% were male, these are the respondents who were able to complete the questionnaires shared to them and send them back.

4.2.2.2 Business Registration

Figure 12: Share of Registered and Unregistered Respondents



76% of companies surveyed are formally registered, 24% of the companies are unregistered. The high percentage of registered companies allows of the study to be valid under the formal context where these registered companies are directly linked to the growth of the economy as they remit all the applicable taxes. On the other hand the engagement and consideration of the unregistered companies enables a fair assessment of the informal niche which is also very critical to the study.

4.2.2.3 Classification of businesses by revenue

The business were classified into four broad categories to allow for refined analysis using the classes, the 4 categories were;

- US$0 – US$5,000
- US$5001 – US$10,000
- US$10,001 – US$20,000

- >$20,000

Figure 13: Revenue Bands

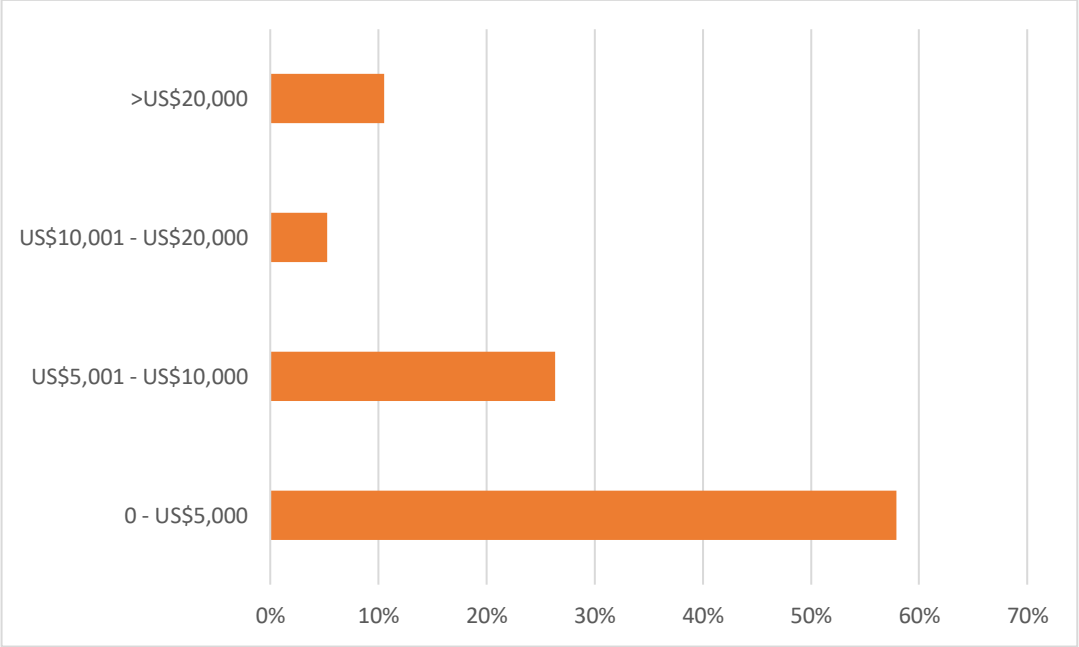


Table 3: Responses and revenue category

Revenue Category	Frequency	Percent (%)	Valid (%)	Cumulative %
0 - US$5,000	85	**58%**	**58%**	58%
US$5,001 - US$10,000	38	**26%**	**26%**	84%
US$10,001 - US$20,000	7	**5%**	**5%**	89%
>US$20,000	16	**11%**	**11%**	100%
Total	**146**	100%	100%	

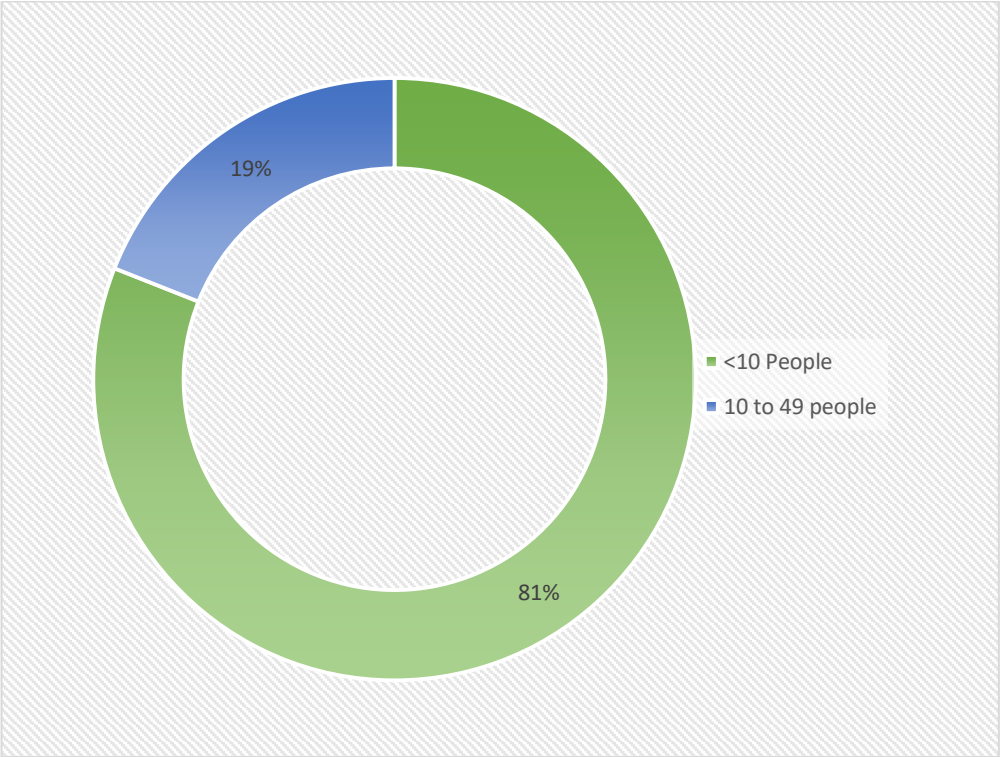
Figure 13 and Table 3 shows that 58% of the respondents earn a monthly revenue of less than US$5,000, 26% are businesses which earn US$5,001-US$10,000, 5% of the

companies earn US$10,001-US$20,000 and 11% of the companies earn a revenue of more than US$20,000 a month.

4.2.2.4 Firm Size and Concept Familiarity

The firm size was also analysed looking at the employees directly employed under the firms, Figure 13 shows the distribution.

Figure 14: Firm Size



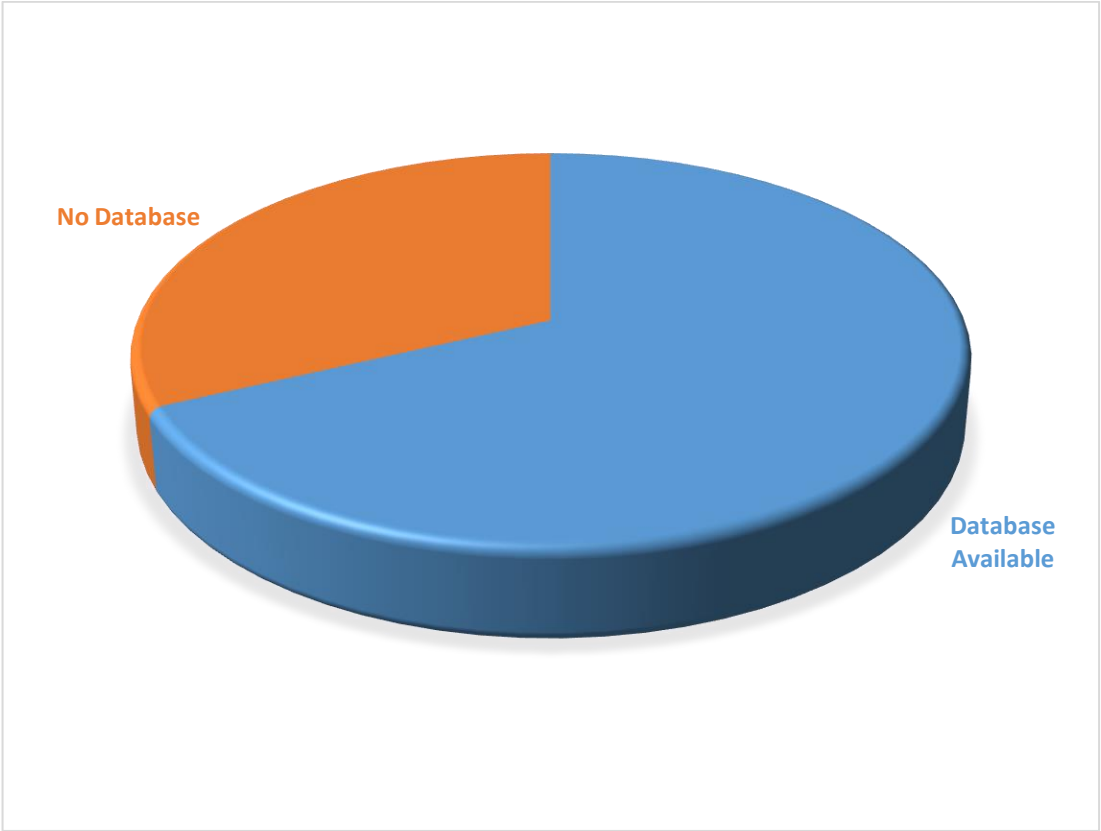
81% of the respondents run companies which employ less than 10 people and 19% of the respondents employ between 10-49 people in their organisations. This shows that many Small to Medium Enterprises in Zimbabwe employ less than 10 people, from the data collected for the survey there is no correlation between business type and firm size.

4.2.2.5 Internal Organisation – Data Base Management and Quality of Data

Internal organisation and the readiness and preparedness of the firm to roll out a Big Data Analytics was analysed under two different surveys.

a. Availability of a customer database.

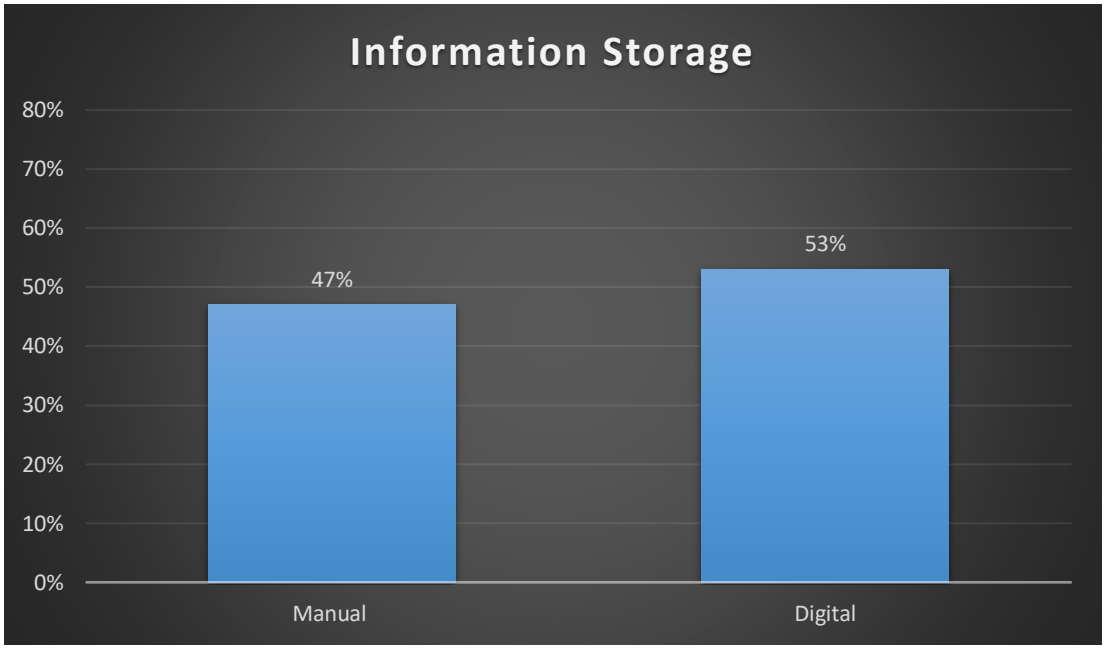
Figure 15: Database Availability



68% of the respondents acknowledged that they keep a database where they capture and store information on their interactions with customers, that is, information on daily sales, orders and quotations. 32% of the companies do not have such a system. Unavailability of a database entails that there is no deliberate effort by the firm to track daily customer interactions and it also predisposes the company to some blind spots in its daily business processes. 90% of the respondents acknowledged that they have the infrastructure to take on a big data analytics system

a. Data capture and information storage (manual or digital)

Figure 16: Information Capture and Storage



47% of the businesses which were surveyed keep their information in manual files and 53% use digital sources to store their data. Storage of information and investment in the infrastructure to store information is the prerequisite for rolling out a Big Data Analytics system. The survey shows the exposure of the bulk of the companies in risks associated with information storage.

4.3 Discussion and Interpretation

4.3.1 Concept Familiarity

The concept of BDA is still fairly known and appreciated, from the data survey it was noted that 63% of the respondents are familiar with the concept, 37% of the respondents have never heard of the concept and the benefits of the concept (new term). This represents a very considerable proportion and points out to the knowledge gap in the current state of the Zimbabwean environment as it relates to the new concept

which has been identified as a frontier for growth and development in developed nations. The gap can be covered by education and roll outs of low costs BDA framework which can be adopted my small firms (low cost model). Development of application which can capture data and show trends, set reminders and can be used for target marketing.

Example

Retail business – a company into cakes can capture the following information as shown below on an app.

- Name of Person (can include totem or nick name)
- Contact details
- Marital Status
- Date of Birth
- Name of Spouse if Any and their Date of Birth.
- Name of Children and their Date of Birth

NB: with this information, the application can send reminders to the Manager and will therefore prompt the company to send birthday reminders a month before (or a birthday countdown), offer discounts, offer add-ons. This will not allow the customer to shop around for options but through sentimental marketing the business can attract and keep the customers.

4.3.2 Database Management

68% of the respondents from the survey indicated that they sort and keep some form of database for all their daily business interactions, however of the number, 53% only hinted that they keep the information in digital form. This reflects very badly on the internal organisation of these companies and also shows the missed opportunities due

to lack of the basic capture and storage of business interactions. Information is to a business what blood is to the body, data collected and processed can show trends, pain points of the business, forecasts, sales trends, seasonal trends and all this information is critical to make informed decisions.

Example

Personal Care Business – a digital database capturing the right information should answer the following questions.

- Who buys from the business (profile, anchor customers) – informs on marketing efforts.
- When do they buy – trends, sales analysis, informs on stocking and right stocking.
- How often do they buy – repeat business, customer loyalty – informs on loyalty programs, targeted and specific.
- In what quantities do they buy – the trend will inform on stocking, bulk pricing for the anchor customers.
- Supply chain considerations – for each product, the sales rate and the lead times should inform the inventory management.

4.3.4 System Investment and Infrastructural Considerations

80% of the respondents alluded that they will be ready to invest in a BDA system if there is guarantee of growth in their sales and subsequently their revenues. This result shows that with a low cost system and a monitoring and evaluation procedure where the growth can be measured and proven to the business owners there is a possibly of broad acceptance and use of the system. 92% of the participants already have the basic

infrastructure to set up a low cost system and 100% of the respondents have basic computer knowledge.

Summary

The data collected from the respondents reflects some glaring gaps in the internal organisation of the Small to Medium Enterprises, data is valuable and is now being treated as a commodity and in some instances as a currency, therefore from the data shown, companies should be able to harness quality data and should be able to process it and gain insights of the business to the end that these companies should be able to grow in market share.

CHAPTER 5 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1Introduction

The research project's thrust was to build a conceptual framework for using Big Data Analytics in the Small to Medium Enterprises sector looking the context of the Zimbabwean economy, by and large the project focused on a deep analyses of the SME sector from a TOE (Technological, Organisational and Environmental) standpoint highlighting the internal organisational aspects that can facilitate the adoption of a BDA framework and the short, medium and long term goals achievable through leveraging on the system. The methodology used for the project was a descriptive qualitative approach, the method was picked so as for the researcher to fully understand the respondents' views as well as to get insights into their business operations. The researcher used a population of Small to Medium Enterprises in Zimbabwe, a sample was selected using convenience sampling. The researcher targeted business owners and senior managers for the survey to ensure that the information was accurate and verifiable. A questionnaire was used for data collection which consisted of clearly structured questions for the respondents to answer.

5.2 Discussion

The study showed that for the adoption of a Big Data model in a firm key aspects should be looked at, these are the Technological aspects, that is, the infrastructural considerations and information systems frameworks, Organisational and Environmental aspects, the organisation should be ready and equipped to manage and use the systems, training is required and the right attitude from users is very important. The study revealed that the majority of small to medium enterprises capture their data and store it in databases, however an analysis of the quality of this data revealed considerable opportunities missed in the data capturing process and the stark inability

of the users to create trends and read the data to understand the business and to make strategic and informed decisions. The key reflections from the survey which was conducted as evidenced by the results outlined in chapter four show that Small to Medium Enterprises stand to benefit from on boarding a big data analysis framework, this is evidenced by the missed opportunities which the study uncovered, these blind spots include;

- Unavailability of a database to safely store information from customer interactions.
- Lack of a clearly planned and thoughtful data collection medium which allows for processing into trends which can be converted into business in the future.
- Storage of data in manual format which cannot be processed and analysed

These among other factors are the key opportunity points which are the BDA system can leverage for business growth. Though seemingly complex and expensive at first sight, a BDA system is a necessary investment, the bulk of the respondents in the survey affirmed that they were ready to make an investment into the adoption of the system as long as there is guaranteed returns. The study also outlined the key elements of the system which should be clearly understood by the entrepreneurs for the roll out of the system, these are data ingestion and storage, not all data is good data, the inception or the starting point of the model is collection of data. The data should be correctly captured and cleaned before it is stored in data warehouses, data processing, BDA applications can be used to process the data, the data can then be reported and interpreted in “easy to interpret” reports which can be understood by a common entrepreneur.

5.3 Conclusions

In Conclusion, there is compelling evidence that the implementation of a BDA system for the SME's is capable of transforming the sector and contribute positively to the economy of Zimbabwe, even though the investment costs may be high for a comprehensive installation of the BDA system however a tailor made product which is affordable and within reach of the SME's can be designed. BDA can be a frontier for economic development by empowering the SME's.

5.4 Implications

The research outlines a conceptual framework for the adoption of a Big Data Analysis for the SME sector in Zimbabwe, the results shows that 32% of the businesses in Zimbabwe do not keep the information of their daily interactions, there are missed opportunities in the grey areas in storage and processing of data. A roll out of a BDA system can result in the growth of an SME enterprise. The implications of the research on the SME industry can be very big if the model is rolled out in a customised format which will be within reach in terms of costs to the businesses. The research finds are generalizable and can be scaled up or applied to different industries in Zimbabwe in the field.

5.5 Recommendations

An exhaustive study into the field of Big Data Analytics is necessary so as to understand the impacts and the benefits to the context of the Zimbabwean industries. Small to Medium Enterprises can benefit from the deployment of a Big Data Analytics system, due to the current state of the economy, a low cost entry model will be necessary with potential to scale the system up for the micro enterprises, the system will be the basic model. The system therefore can be tailored depending on user

specifications which are dependent on firm size and firm performance. The system will benefit the SME's by;

i. Data capture – the SME's will be oriented to intelligently capture useful data from their customers, the data captured should allow the business owners to understand their customers, target marketing for example, a Cake Retailer can capture date of birth for clients and close relatives, when that data is captured it will allow the entrepreneur to profile and understand customers, target clients, send reminders through a system generated message servicing, this will create potential for future repeat business and also create loyalty (emotional marketing).

ii. Data Storage and Processing, making sense of data – for some businesses under the survey, it was noted that the businesses generate a lot of data daily and monthly, the SMEs do not keep data for periods beyond a year, data is lost due to storage capabilities, the volatility of the infrastructure of the storage systems poses risk of data security, there is a huge vulnerability to lose information, the recommended storage system in the framework will ensure that the data is kept for periods longer than three years, the information will allow simulation of historical trends, to understanding the sales trends, inform stocking requirements, recognize the fast selling products and the relevant stockholding. This will allow SME's to better manage the businesses and make informed decisions.

5.6 Suggestions for Further Research

The future research work should focus on designing solutions which can be rolled out for SME's in Zimbabwe. There is a growth of the informal market in Zimbabwe and

deployment of a BDA systems applicable for the various stages of the businesses will be a game changer to grow the economy.

REFERENCES

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of Things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys and Tutorials*, 2349.

Armbrus, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., . . . Zaharia, M. (2010). A view of cloud computing. Communications of the ACM. *ACM Journal on Computing and Cultural Heritage*, 50-51.

Dumbu, E., & Chadamoyo, P. (2012). Managerial deficiencies in the Small and Medium Enterprises (SMEs) in the craft industry: An empirical evidence of SMEs at Great Zimbabwe in Chief Mugabe's area. *European Journal of Business and Management*, 84.

Epstein, E. (2013, January). Big data and the creative destruction of today's business models. *Kearney*.

Etikan, I., Musa, S. A., & Sunu, R. (2015). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 4.

Kramer , M. P., Bitsch , L., & Hanf, J. (2021). Blockchain and its impacts on Agri-Food supply Chain network Management. *MDPI*, 2.

Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, 114-124.

Soltanpoor , R., & Sellis , T. (2016). Prescriptive Analytics for Big Data. *Australasian Database Conference* (pp. 246-326). Sydney: Springer International Publishing.

Alkahtani, A., Nordin , N., & Ullah, R. K. (2020). Does government support enhance the relation between networking structure and sustainable competitive performance among SMEs? *Journal of Innovation and Entrepreneurship*, 1.

Anokye , A. M. (2020). Sample Size Determination in Survey Research. *Journal of Scientific Research and Reports*, 93.

Brozovic, D., Jansson, C., & Boers, B. (2023). Strategic Flexibility and growth of small and medium-sized enterprises: a study of enablers and barriers. *Emeral Publishing*, 15.

Caldeira , M. M., & Ward, J. M. (2003). Using resource-based theory to interpret the successful adoption and use of information systems and technology in manufacturing small and medium-sized enterprises. *European Journal of Information Systems*, 127-141.

Calheiros, R. N., Ranjan, R., Beloglaz, A., De Rose, C. A., & Buyya, R. (28). CloudSim: a toolkit for modeling and simulation of cloud computing

environments and evaluation of resource provisioning algorithms. Software: Practice and Experience. *Wiley Online Library*, 2010.

Cattell, R. (2010). Scalable SQL and NoSQL Data Stores. *Springer Open*, 18-20.

Cavanillas, J. M., Curry, E., & Wa, . (2016). *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe.* New York: Springer Open.

Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). New Horizons for a Data Driven Economy. *Springer Open*, 32.

Fitz-Enz, J. (2010). *The New HR Analytics Predicting the Economic Value of Your Company's Human Capital Expenditure.* New York: American Management Association.

Gandom, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 138.

Gudivada, N. V. (2017). *Data Analytics for Intelligent Transportation System.* Amsterdam: Elsevier Science Publishers.

Hu, H., Wen, Y., Chua, T.-s., & Li, X. (214). Toward Scalable Systems for Big Data Analytics: A Technology tutorial. *IEEE Access*, 654.

Keizer, J. A., Dijkstra, L., & Halman, J. I. (2002). Explaining innovative efforts of SMEs.: An exploratory survey among SMEs in the mechanical and electrical engineering sector in The Netherlands. *Science Direct*, 23.

Krol, K., & Zdonek, D. (2020). Analytics Maturity Models: An Overview. *Information*, 10.

Lepeniotia, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 57-70.

Manzoor , F., Wei , L., & Sahito, N. (2021). he role of SMEs in rural development: Access of SMEs to finance as a mediator. *PLOS ONE*, 35.

María , J. C., & Wolfgang Wahlster, E. C. (2016). *New Horizons for Data Driven Economy.* Madrid: Springer Open.

Maroufkhani, P., & Wan , K. W. (2020). Big data analytics adoption model for small and medium enterprises. *Journal of Science and Technology Policy Management*, 11.

Moniruzzaman, A. (2014). Towards Next-Generation Scalable RDBMS for Online Transaction Processing (OLTP) for big data management. *International Journal of Database Theory and Application*, 127.

Musabayana, G. T., Mutambara, E., & Ngwenya, T. (2022). An empirical assessment of how the governement policies influenced the performance of the SMEs in Zimbabwe. *Journal of Innovation and Entrepreneurship*, 4.

North, D., & Smallbone, D. (2010). The Innovativeness and Growth of Rural SMEs During the 1990. *Taylor and Francis Online*, 145.

Perdanaa, A., Leeb, H. H., Kohb, S., & Arisandi, D. (2021). Data analytics in small and mid-size enterprises: Enablers and inhibitors for business value and □rm performance. *International Journal of Accounting Information Systems*, 13-16.

Ridzuana, F., & Mohd Nazmee Wan Zainon, W. (2022). Diagnostic analysis for outlier detection in big data analytics. *Sixth Information Systems International Conference (ISICO 2021)*, 686.

Seyedan , M., & Fereshteh, M. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Springer Open*, 2.

Wamba, S. F., Gunasekaran , A., Shahriar , A., Steven , R. J.-f., Dubey, R., & Childe, S. J. (217). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 357.

Zhao, L. (2020). *Event Prediction in the Big Data Era: A Systematic Survey*. 2020: Emory University.