ESP IJSHMS
ESP International Journal of Science, Humanities and Management Studies
ISSN: 2583-9756 / Volume 2 Issue 5 September 2024 / Page No: 1-23
Paper Id: IJSHMS-V2I5P101 / Doi: 10.56472/25839756/IJSHMS-V2I5P101

Original Article

Strategic Integration of Big Data Analytics for Enhancing Small to Medium Enterprises in Zimbabwe: A Conceptual Framework

Yogesh Awasthi[1], Dumisani Masuku[2], Lorence Dhlakama[3]
[1,3]*Department of Computer Engineering, Africa University, Zimbabwe.*
[2]*Department of Business Sciences, Africa University, Zimbabwe.*

Received Date: 04 July 2024 Revised Date: 06 August 2024 Accepted Date: 05 September 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(BDA) 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 BDA 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 has 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 so as to establish the current operational status of the firms as well as to check if there is the appreciation and presence of fundamental infrastructural processes and infrastructure which relates to BDA. Some opportunity points which can be leveraged to ensure and guarantee business growth were identified from the thirty-eight companies which were surveyed from different industries in the economy. Results of the study point out to a possible growth potential realisable by on boarding and successfully running a BDA 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 BDA system.*

Keywords: *Big Data Analytics, Small to Medium Enterprises.*

I. 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. 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 has been seen to promote growth of a firm [1] 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 of information provide data which can be processed and analysed to understand consumer trends and other opportunities which can be used to grow the businesses, data in its raw and unstructured form is of no value and use for any business owners and unless it is processed intelligently using different tools the business will not benefit from it. SME's have been an important facet to the Zimbabwean economy, a report by the International Finance Corporation shows that Small and Medium Enterprises (SMEs) have been a significant contributor to the Zimbabwean economy and 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 highlights that the SME's contribute 34% of the GDP annually [25]. 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 [29].

A. Statement of the Problem

Zimbabwe has set out an ambitious vision dubbed Vision 2030, the ambitious vision is to transform the country into an upper middle economy by 2030. The plan's priorities include promoting inclusive economic growth, fighting poverty, reorienting the country towards democratization, respect for human and property rights, freedom of expression and association, and the rule of law the aim also is to reduce income inequality, create job opportunities and improve the lives of ordinary people. The strategy is divided into two five-year medium-term plans, the first of which is from 2021 to 2025. The first plan, known as National Development Strategy 1 (NDS1) is focused on stabilizing the economy, modernizing infrastructure, and improving social services.

Several critical industries have been identified as the drivers to attain the ambition including mining and agriculture. The Small to Medium Enterprises can also contribute immensely to the vision and can be earmarked as a growth driver [37], 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 [20].

Prime examples of economies which have used SMEs development as a catalyst for economic development are China, South Korea and Malaysia. By missing out on patterns behind the data companies miss out on opportunities to give better customer experience, offer tailor made services and products. SME's have in recent times in the context of the Zimbabwean economy proven to be aggressive competitors standing neck to neck with established corporates, therefore deployment of BDA will actually scale up the competition, increase market shares of these small entities and thereby resulting in the growth of the Zimbabwean economy. [16] 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.

B. Research Objectives

- To analyze the components of a BDA system.
- To evaluate how the implementation of a BDA framework can improve the competitiveness of a small to medium firm
- To evaluate the risks encountered when implementing BDA.

C. Research Questions

- What are the main processes and infrastructure required for a BDA system?
- How does Data Analytics improve a firm's performance?
- What are the challenges faced by SMEs in adopting BDA capability in their operation?
- How can the BDA framework can be rolled out for an SME?

D. 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**.**

II. LITERATURE REVIEW

A. Theoretical Framework

a) 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 [11]. These limits vary by country, with experts and government agencies setting thresholds for each country or region. Chapter 24:12 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.

b) Importance of Small to Medium Enterprises

i) Job Creation

SME's play a very important and pivotal role according to a study by the International Labour Organisation (ILO) providing two thirds of all formal jobs in developing countries and up to 80% in low-income countries. Unemployment has a big impact on the economy and society of Zimbabwe as it leads to drug abuse, crime, violence and promiscuity, therefore the SME 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 SMEs are a low hanging fruit in creating jobs in the country.

ii) Flexibility and Innovation

SME's are more flexible and innovative than big firms especially in the context of Zimbabwe, strategic flexibility contributes to firm growth in Small to Medium Enterprises [10][27], the study also outlines enables and barriers of strategic flexibility, these include strategic orientation, positive impact of slack resources and the negative effect of lack of competent employees. SME's have been seen to be more flexible than bigger firms which allows them to be more agile and adapt to the market conditions, 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.

iii) Competition in Markets

SME's play a very crucial role in increasing competition in the market for larger firms. By being more flexible, these smaller firms 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. SMEs stimulate competition for the design of products, prices, and efficiency. By stimulation competition for the design of products, pricing and efficiency it prevents a monopolist market where one company has total control over a product and a service resulting in unfair pricing to customers who will not have a choice. Without Small to Medium Enterprises large corporations will hold monopoly in all the activity areas therefore by providing competition SMEs will create a more competitive and healthy economy. Small Medium Enterprises also help large organisations in some specific areas of operation so that that they are better able to supply the market so that when the SMEs are removed the big firms will be forced to be involved in other 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.

iv) Economic Growth

The SME sector is the backbone of many economies in the world including Zimbabwe as they drive economic growth through job creation, creation of new products and services, some of these products can be exported to regional and international markets. SME's also come with innovative solutions which lead to improved productivity and efficiencies.

c) Challenges faced by SME's in Zimbabwe

i) The Operating Environment

Several legal, economic and social developments are a direct result of the political decisions made by the government 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 [3]. The political environment has been seen to greatly affect investor confidence, and where 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 deter investors. Government support plays an important role in the development and growth of small and medium-sized enterprises [8] this includes creating a favourable business environment within a stable political climate.

To stimulate economic growth through entrepreneurship, governments need to focus on supporting entrepreneurs and from the government's end this is achievable through the right mix of policies, consistent policies ensure stability and will enable SMEs to plan and forecast without any fear of future unforeseen eventualities. Government measures are needed to protect small and medium-sized enterprises from both internal competition (from large corporations) and external competition. Regulatory constraints such as licensing and regulations also pose a serious threat to small businesses.

ii) Finance and Operational Constraints

All small to medium-sized enterprises need financing in order to run their day to day operations, restock, open more routes to the market, on-board new technologies and survive in the market. Cash flow problems are a barrier to most businesses in Zimbabwe, and liquidity problems and inability to service debts are the two main causes of financial difficulties for businesses. Several small and medium-sized enterprises rely on internal funding sources during the first years of their lifecycle but are they grow in size and expand their activities they require access to external funding. Therefore lack of facilities for SMEs hinders 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 Small to Medium Enterprises. 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.

iii) 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 [18]. Big Data can also be viewed as high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information process for enhanced insight and decision making [17]. These two definitions bring out the nexus between Big Data Analytics and Business Intelligence in an organisation, where companies can efficiently manage data and incorporate it into their decision making processes, this will result in growth in sales and profitability. Big data analytics is also defined as the process of examining 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 [19]. Big data analytics is used in several industries such as the healthcare, finance, retail, and manufacturing for process improvement, system improvement, and analytical purposes. The three Vs of big data analytics, which explain how big data differs from traditional data management, are volume, velocity and variety.

- Volume: Refers to the amount of data generated. Big data is characterized by vast amounts of data that are generated and can be difficult to manage using traditional data management techniques.
- Speed: This refers to the speed at which data is generated and processed. Big data is often generated in real time, and businesses need to be able to quickly process this data to gain insights
- Diversity: This refers to the different types of data that are generated. Big data comes in many forms, including structured data, semi-structured data, and unstructured data

iv) Data Analytics and its types

Data analytics is the process of evaluating or examining data sets to postulate the underlying information they contain, this is done with the aid of specialized systems and software. There are four main broad types of data analytics, these are;

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

BDA involves extensive analysis of data to meet different needs, the different types of data will require therefore require different approaches. Analytics technologies are often classified into four categories [6]. Figure 1 shows the different types of data analytics and what they entail.

1. Descriptive Analytics

Descriptive analytics is a form of data analysis that breaks down large amounts of data into smaller usable pieces of information to help companies understand what happened during a particular operation, process, or series of transactions. It is used to identify trends and relationships in data and answer the question, "What happened? Descriptive analysis is the simplest form of data analysis because it describes trends and relationships but does not provide deeper explanations. Because it is relatively easy to access, it is likely to be used routinely in your organization. Simple statistical software such as Microsoft

Excel and data visualization tools such as Google Charts and Tableau can help you analyze data, identify trends and relationships between variables, and present information visually. Descriptive analysis is particularly useful when using trends as a starting point for further analysis to communicate changes over time and drive decision-making. In descriptive analytics information which is provided will be computed from the past by calculating descriptive quantities e.g. the average, mode and the standard deviation. The results are presented in graphical format that is charts and lines. This is the most popular type of analytics which is 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?” [16].

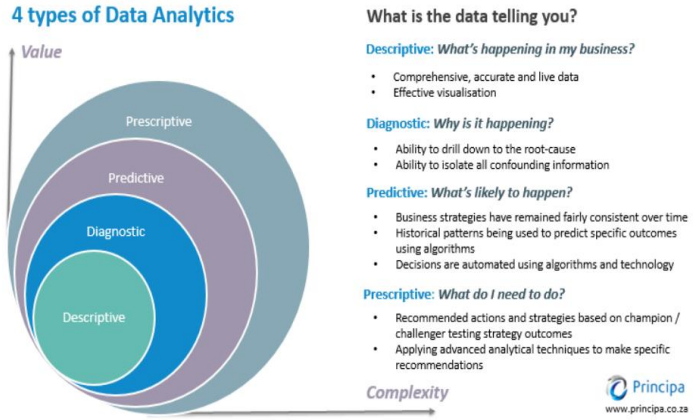

Figure 1: Types of Data Analytics

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 [34]. The commonly used method of predictive analytics is regression modeling where the objective is to create a model in which parameters are computed in a way which minimizes residuals, other methods include decision trees, neural networks and support vector machine. The most relevant question predictive analytics can answer is “What would happen if…?” [20]. Predictive analytics is a field of advanced analytics that uses historical data, statistical algorithms, and machine learning techniques to predict future events. This involves analyzing data to identify patterns and relationships, and using those patterns to predict future outcomes. Predictive analytics can be used in various fields such as finance, healthcare, and marketing. Predictive analytics differs from prescriptive analytics, which determines the best course of action based on predictions made by predictive analytics [14][15]. Prescriptive analysis uses optimization algorithms to determine the best decisions based on the predicted results. Predictive analytics is also different from descriptive analytics, which focuses on analyzing past data to understand what happened and why. Descriptive analytics is used to gain insight into past trends and patterns, while predictive analytics is used to make predictions about future events

3. Diagnostic Analysis

Diagnostic assist to answer the important question, which is why did it happen [7]? This is done by an analysis of the historical data and combining multiple sources in search of patterns, correlations and trends. Diagnostic analysis is a type of data analysis that examines data or content to answer the question, "Why did that happen? It features techniques such as drill-down, data discovery, data mining, and correlation. Diagnostic analysis can be performed manually using algorithms or using statistical software such as Microsoft Excel. This is used to identify behaviours, trends, patterns, and find out why certain outcomes occur. The process of diagnostic analysis can be seen as the logical next step after identifying trends using descriptive analysis. Hypothesis testing, the difference between correlation and causation, and diagnostic regression analysis are some of the concepts you need to understand before delving into diagnostic analysis. Because of its focus on cause and effect, diagnostic analysis is sometimes called root cause analysis. The goal is to identify and explain anomalies and outliers. Although diagnostic analysis has many benefits, it is important to be aware that it also has some limitations and drawbacks. For example, it depends on the quality of the data.

If the data is incomplete, inaccurate, or skewed, the conclusions drawn from it may also be incorrect. The main objective of diagnostic analytics is to:

- Identification of anomalies: that is the areas which require investigation, these can be singled out for querying and highlighting important questions which cannot be answered by just looking at the data can for example why have sales gone down for a particular season.
- Determination of causal relationships: After anomalies have been identified advanced statistical techniques are used to

determine whether there any relationships.

Current developments in diagnostic analytics use machine learning techniques to augment the analysis. Computers process vast amounts of data to identify patterns, pick anomalies, and show outliers or uncharacteristic trends [32].

4. Prescriptive analytics

An emerging technology which goes beyond the previously discussed methods in that it goes beyond to recommend action paths and calculates the most probable outcome for each decision path. The method does not only predict possible future outcomes but will outline various possible 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 [23]. Prescriptive analytics is a type of data analysis that seeks to answer the question, "What do I need to do to make this happen?". This involves using technology to analyze raw data and assist businesses in making better decisions. Prescriptive analysis takes into account information about possible situations and scenarios, available resources, past and current performance, and suggests a course of action or strategy. It can be used to make decisions for any time period, from immediate too long term. Prescriptive analysis is the opposite of descriptive analysis, which examines decisions and outcomes after the fact. Work with predictive analytics that uses data to determine short-term outcomes. When used effectively, it allows companies to make decisions based on predictions that are weighted with facts and probabilities, rather than intuitive conclusions. Normative analysis is not infallible because it is only valid according to its inputs. We rely on artificial intelligence (AI) techniques such as machine learning to understand and evolve the data we collect to constantly adapt. Machine learning allows us to process the vast amounts of data currently available. When new or additional data becomes available, computer programs automatically adapt to use that data in a process that is much faster and more comprehensive than human ability. This new digital world is generating vast amounts of data and opening up new paradigms. Because of the availability of high computing power and large amounts of data, this data can be utilised to make data-driven decisions. The main advantage of data-driven decision-making is that it is based on observations of past trends that have produced favourable outcomes. Therefore, data analysis is the process of manipulating data to extract useful trends and hidden patterns that help derive valuable insights for business prediction.

v) The Data Analytics Life Cycle

The data analysis lifecycle is a process that includes six phases as shown in Figure 2 below.

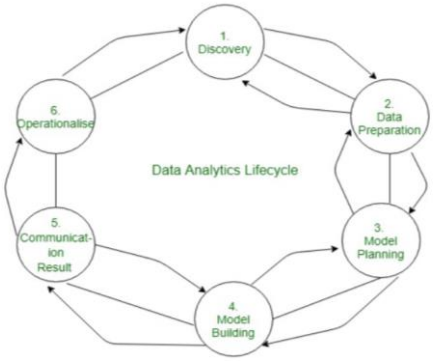

Figure 2: Data Analytics Life Cycle

Note: Data Analytics life cycle. From: "Life Cycle Phases of Data Analytics

- Discovery: In this phase, the data science team investigates the problem to gain context and understanding. The team also develops an initial hypothesis that can later be tested using data.
- Data Preparation: At this stage, the data is inspected, pre-processed, and prepared before modelling and analysis. Data preparation tasks may be performed multiple times and not in a predefined order.
- Model planning: In this phase, the data science team examines the data to learn the relationships between variables and select the key variables and the most appropriate model. The team develops datasets for training, testing, and production purposes.
- Model building: In this phase, the team develops data sets for testing, training, and production purposes. The team also considers whether existing tools are sufficient to run the model, or whether a more robust environment is needed to run the model.
- Communication Results: After running the model, the team should compare the modelling results to established criteria for success and failure. The team considers the best way to communicate findings and results to various team members and stakeholders, taking into account caveats and assumptions. Teams must identify key results, quantify business value, and create a narrative to summarize and communicate results to stakeholders.
- Operationalize: In this phase, the team sets up a pilot to more fully communicate the benefits of the project and deliver

the work in a controlled manner before scaling the work across your organization. This approach allows teams to understand model performance and associated limitations in a small production environment and make adjustments before full deployment.

1. Big Data Components and Value Chain

A Big Data Analytics system is complex and therefore requires several components so as to function effectively, 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 3 V's of Big Data Analytics represent the challenges and the problem which BDA seeks to solve.

2. Volume of data

Handling large volumes of data for industries such as Global supply chains and financial systems is a tremendous task as this involves high volumes of data generated every second. The volume of data refers to the size of data sets which an organization collects for processing and analysis. The data sets are now in larger sizes such as terabytes and petabytes. Big data analytics is therefore the process of analysing these large volumes of data to derive insights and trends. The data storage system in a BDA system platform arranges the collected information in a convenient format for further analysis and extraction of value. For this purpose, the data storage subsystem should provide two sets of features:

- The storage infrastructure which must accommodate information reliably.
- The data storage system which provide access and an interface to query and analyse the vast amounts of the data. Data storage can be divided into hardware infrastructure and data management.

3. Storage Infrastructure

The hardware infrastructure is responsible for the physical storage of the collected data. Storage units are classified based on the specific technology; the popular storage technologies include the following.

- Random Access Memory (RAM): This is a form of data storage associated with volatile types of memory, these loses information when powered off. Modern technologies have seen the emergence of static RAM (SRAM), dynamic RAM (DRAM), and phase-change memory (PRAM). DRAM is the popular form of computer memory.
- Magnetic Disks and Disk Arrays: Magnetic discs are the primary component in modern storage systems. A HDD comprises one or more rigid and rapidly rotating discs with magnetic heads which are arranged on a moving actuator arm to read, interpret and write data to the surfaces. A HDD retains data even when powered off and has significantly lower per capacity cost, however they read and write operations are slower. Due to the high costs of a single large capacity disk, disk arrays line up several disks to create a larger capacity, throughput, and high availability at a lower cost.
- Storage Class Memory: This refers to non-mechanical storage media e.g. such flash memory. Flash memory is used to build solid-state drives (SSDs) and unlike HDDs, SSDs have no mechanical components, hey run with no noise, and have low access times and less latency than HDDs. SSDs however remain more expensive per unit of storage than HDDs.

These storage systems have different performance metrics, these can be leveraged to build a scalable and high-performance big data storage system.

4. Database Technologies

Various database systems are used for different scales of datasets and applications. Traditional relational database systems can not address the volumes and scale challenges required by big data. The three models of data are the key-value stores, column-oriented databases, and document databases.

5. Speed of data

Handling streams of high frequency incoming real-time data such as information from sensors, IoT and electronic trading is a very strenuous process. The speed of data in BDA is also referred to as Big Data Velocity. Big Data Velocity is the rate at which large sets of data are acquired and processed. Big data is frequently available in real-time and it is produced continually with high speeds so it requires distinct processing techniques. Velocity refers to how quickly data is ingested.

i) Big Data Value Chain

The data which is handled in a BDA system comes in different syntactic formats such as spreadsheets and schemas. The European Commission views the data value chain as the centre of the future knowledge economy, bringing the opportunities of the digital developments to the more traditional sectors (e.g. transport, financial services, health, manufacturing, retail). Figure 3 below shows the data value chain.

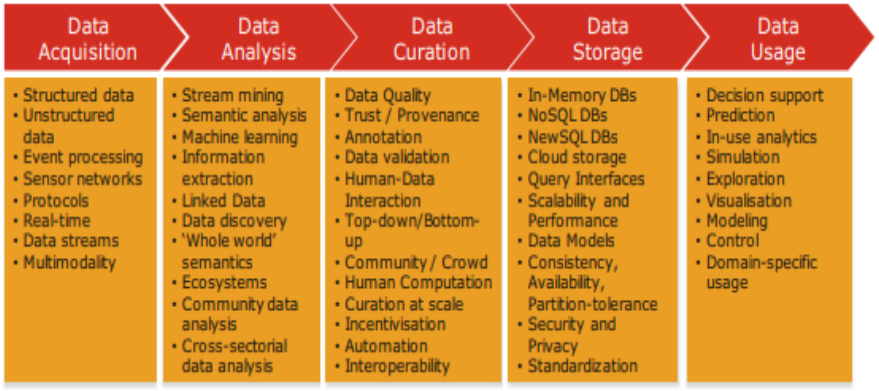

Figure 3: Big Data Value Chain

ii) Data Acquisition

Big Data Acquisition is concerned with making raw data acquired amenable for use in decision-making as well as domain-specific usage. Data analysis involves exploring, transforming, and modelling data with the goal of highlighting relevant data, synthesising and extracting useful hidden information with high potential from a business point of view. Related areas include data mining, business intelligence, and machine learning [14] [15]. Oracle in their article Oracle Enterprise Architecture white paper suggested a three-step approach for data processing.

- The content from the different sources of data is retrieved and stored in a scalable storage solution such as the NoSQL database or the Hadoop Distributed File System. The stored data is then processed, organized and stored in an SQL-capable BDA software and then it is further analysed using a big data analytics algorithm. The core of data acquisition is gathering data from information sources so as to keep it in a scalable and a big data-capable storage. Three elements are critical to achieve this
- Getting the protocols which will allow the data streams of any type to be collected (that is structured, semi structured and unstructured).
- The frameworks where the data will be collected from the sources by using different protocols.
- The technologies which will allow storage of the data which would have been retrieved by the frameworks.

There is an exponential increase in the data being gathered today from several sources. All this data needs to be aggregated to assist businesses to confidently answer questions and have a picture of the market and the trends. The two main technologies of acquiring data are;

1. Protocols

Organizations or firms which rely internally on Big Data Analytics have developed enterprise-specific protocols. The commonly used open protocols used for data acquisition are:

1.1 AMQP (Advanced Message Queuing Protocol)

AMQP was designed to cover the gap and satisfy the need for an open protocol which could fulfil the requirements of big firms with respect to data acquisition. AMQP operates four different layer namely a transport layer, a messaging layer and a transaction and security layer; it uses data sets such as integers, symbols, strings, etc. The transport layer relates how messages will be processed. An AMQP network consists of nodes which are connected through links, these messages are extracted from senders and can be forwarded by relays, or be consumed by nodes. The transport layer backs and supports different types of route exchanges for example the topic exchange. The transaction layer will allow coordinated outcomes of independent transfers and the encryption of the content of AMQP messages.

1.2. Java Message Services (JMS)

JMS allows Java related programs to produce send and receive enterprise messaging system's messages and notifications. JMS offers two messaging alleys these are the point-to-point and publisher-subscriber. Advanced Message Queuing Protocol is compatible with JMS.

i) Software Tools

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

- Storm
- S4
- Kafka
- Flume

- Hadoop

The use of each tool will need a deep understanding and thorough knowledge of the internal functions and the use of the software. The architectural diagram Figure 2 shows an overall picture of the complete big data workflow highlighting the data acquisition part.

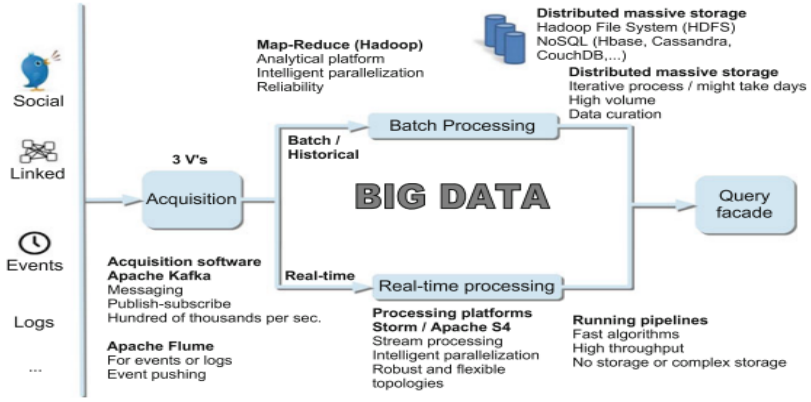

Figure 4: Big Data Value Map

ii) Data Analysis

Data comes in an array of formats and one way to compare the different data formats is the amount of structure it contains. The more structured a dataset is the easier it will be for machine processing; semantic representations will allow machine reasoning. Big Data Analytics is a subset of Big Data which deals with the structuring of data so as to enable precise decision making. Data Analysis within the overall Big Data value chain is very critical, data in its raw form which in many instances is unstructured and contains many different formats is transformed so that it will be ready for data curation and subsequently, data storage and usage, without Big Data Analysis the bulk of the extracted data will not be in a usable state and therefore meaningless. As shown in Figure 1 the following are the critical elements covered under Big Data Analysis, stream mining and semantic analysis, machine learning and information extraction, linked data and data discovery, ecosystems and community data analysis, cross sectorial data analysis and data quality amongst other factors.

iii) Stream Mining

Data stream mining is the process of mining or acquiring knowledge from continuous and high-velocity data sets that enter a system in a stream. This is a subset of the general concepts found in machine learning, knowledge extraction, and data mining. Stream mining allows for the analysis of large amounts of data in real time using limited resources. MOA is the most popular open source framework for highly active data stream mining, it covers the collection of machine learning algorithms namely classification, regression, clustering, outlier detection, and recommendation systems. For data streams, the distribution of samples received can change over time due to the inherent temporal nature of the stream. Stream mining algorithms are computationally fast and lightweight.

iv) *Semantic-Analysis*

Semantic analysis is a natural language processing solution which analysis the meaning of words and phrases with the view of understanding the intended purpose of a sentence or paragraph. This is the process of pulling out meaning from text and translating it into a concrete plan of action. Semantic analysis can be used to implement operational improvements, serve customers, and streamline organizational processes for knowledge management systems. Semantic Web technologies enable the automatic processing of data through advanced reasoning techniques. The technologies have been successfully applied in many scenarios for heterogeneous data integration, knowledge-level data analysis, and data visualization.

v) Linked Data

Linked Data is a set of design principles used for sharing machine-readable, interrelated data. This is one of the foundations of the Semantic Web which is also known as the Web of Data 1. Linked Data provides best practices for creating links between datasets that are understandable to humans as well as machines. When combined with open data (data which can be freely used and distributed) it is called Linked Open Data (LOD). RDF database Ontotext like GraphDB is an illustration of LOD, it can process large sums of datasets from disparate sources and it links them with open data, aiding knowledge discovery and effective data-driven analysis. Big data analysis is a crucial part of the big data value chain, the process is crucial in handling large amounts of data, which are heterogeneous in terms of encryption mechanisms, structure and the underlying semantics. Big data analysis includes techniques and tools, some of which are old mechanisms such as large-scale inference

that have been structured to meet the problems posed by the three V's. (E.g. stream inference). It is clear that providing a scalable data analysis and the inference mechanisms together with an ecosystem of accessible and usable tools is of great importance in Data Analysis. Business models and processes are being changed for economic and social benefit and big data analysis is an important part of the overall big data value chain, without big data analysis, the rest of the chain would not work.

vi) Data Curation

Data curation outlines the methodological and technical support to data management which address data quality issues and maximize the usefulness of data. Data curation is the active and continuous management of data throughout its lifecycle. Curation activities enable data discovery and retrieval, maintain quality, add value, and ensure long-term reuse. Data curation is an important part of the growing number of data sources and data generation platforms and it is emerging as an important data management process. The data curation process can be categorized into different activities such as content creation, selection, classification, transformation, validation, and storage. The selection and implementation of a data curation process is a multidimensional issue, depending on the interaction between incentives, economics, standards, and technical aspects. The section analyzes the dynamics of data where data curation finds its place, it also considers the future needs and emerging trends in data curation. Using data which has been generated under different requirements has in-built costs in addressing data quality and data heterogeneity issues, the data may be incomplete or may require transformation to be usable. 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 [14][15]. Data curation can be conducted on consumer data this can be done by selecting and transforming data resources to be commensurate with the data consumer's requirements. Data curation activities rely on the challenges that arise in the context of big data, in particular the diversity of data. It is costly to deal with data diversity even for small quanta of data, the challenges are not faced because of the size of data which now comes in petabytes but more so because of the data which is diverse and thereby requiring vast amounts of resources to make sense of it. Data Diversity in the context of BDA is the tendency of data management to handle data from a string of sources. Data curation is becoming a more common data management activity which was earlier restricted to small research houses where it was used to address data quality challenges.

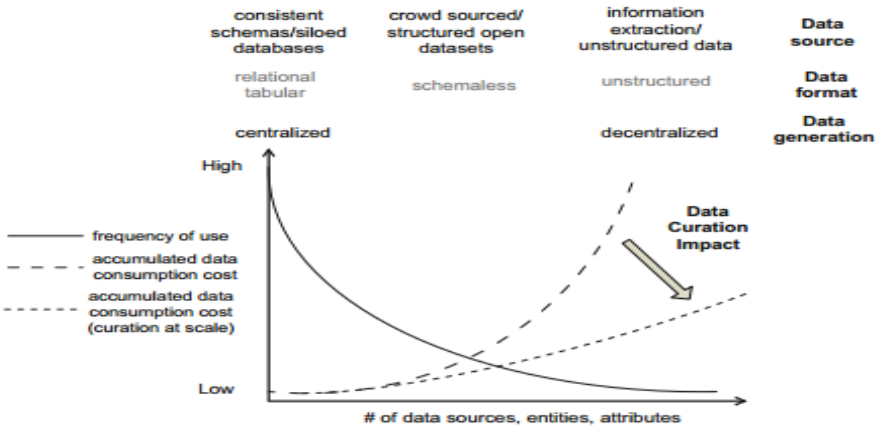

Figure 5: Data Curation

vii) Data Curation Platforms

1. Data Timer

Data Timer aims to replace the current platforms such as ETL (Extract, Transform, and Load) process with a much more automated data integration. The system utilizes sets of algorithms to spontaneously map schemas and reproduce entities. Human experts and crowds will then be used to review integration updates.

2. ZenCrowd

ZenCrowd was developed to solve the problem of tying up and linking named entities which come in text form to a knowledge base, it therefore bridges the gap which exists between automatic and manual links by refining and improving the results of automatic links with people.

3. CrowdDB

This system resolves queries which arise from SQL which cannot be resolved by database management systems or any of the search engines. CrowdDB permits vague operations using humans such as the ranking and profiling of items by significance and comparing the equivalence.

4. Qurk

Qurk is similar to CrowdDB except that it aims to reduce the cost and latency by applying techniques namely batch processing, filtering, and output matching.

5. Wikipedia Bots

The system processes programmed algorithms so as to access the quality of text articles which are called bots. The bots flag and highlight articles which require further reviews by experts.

6. Data Storage

It is the storage and management of vast amounts of data so as to meet the requirements of applications that require the access to the data. An idyllic big data storage system should have the capability to store unlimited amounts of data and should have the capability to handle both high-speed random data cop with different data models efficiently. Big data storage technologies are the storage technologies which address the three V's of Big Data which are volume, velocity and Variety challenges. These storage facilities are an improvement of the traditional storage frameworks which are not effective in storage the volumes of data which the BDA system can handle. Hadoop is commonly considered as the most efficient and cost effective solution in data storage. Big data storage system resolves the problems of handling voluminous data by leveraging the distributed shared architecture allowing a system to handle increased storage requirements, this is done by the expansion into new nodes which provide computing power and storage. New machines can be integrated flawlessly to a storage cluster; the storage system handles the distribution of data between nodes. Storage solutions must be developed to solve the challenges which relate to velocity and variety of data, speed is crucial as it relates to query latency that is the time it takes to receive feedback on a specific request. Big data storage is now a business imperative and standard therefore a scalable storage technology at enterprise level where it is supposed to manage unlimited amounts of data is a necessary investment by every firm which wishes to take advantage of Big Data Analytics. Compared to the traditional database management systems which utilizes row-based storage and costly caching strategies the new big data storage technologies come with reduced operational complexity and costs, and superior scalability which is important for businesses. Big data storage is used same way as the traditional relational database management systems for online transaction processing solutions (OLTP) and structured or semi-structured data warehouses, the strength in particular strength lies in processing unstructured and semi-structured data at scale. This section assesses the current state of data storage technologies capable of handling large amounts of data and identifies trends related to data storage.

Types of storage systems include:

1. Distributed file systems

File systems for example the Hadoop file system (HDFS) can reliably store large amounts of unstructured data on commodity hardware. Although there are some better performing file systems available the HDFS is a very important part of the Hadoop framework. This system was designed for large data files and it is suitable for quicker data collection and bulk processing.

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) [13].

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 [27].

4. Cloud Storage

With the increase and popularity of cloud computing, the influence of big data is also increasing. Organizations such as Microsoft and Google are creating and building their own cloud platforms, other organization like IBM, Dell and Cisco are jumping onto OpenStack which is an open source cloud system. Cloud storage is a distributed file system with complicated architecture [12], it is deployed over the cloud computing infrastructure which is based on affordable, virtualized and inconsistent and therefore unreliable physical hardware. Cloud storage supports large server scale, and has efficient heavy data storage [2]. 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 [1].

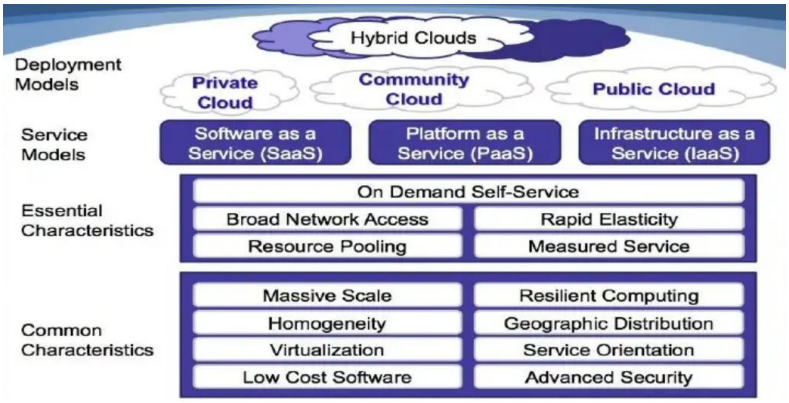

Figure 6: Cloud Storage Note: Cloud Storage.

a) Data Usage

Big data will is poised to significantly impact the efficacy of future data-based decision making processes. The main factors which will influence this are:

- The time perspective of the decision/recommendation which will range from short term to long term,
- Different databases applicable for different scenarios ranging from past historical data to current data.

New data-driven applications significantly impact the development of the new markets. A potential obstacle to that development is the need for new critical insights into the use of big data. Key examples of predictive analytics applications 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 [11]. Even greater savings are possible when data flags issues before scheduled maintenance is reached, avoiding failures, repair costs, and downtime. Information bases are not only sensor data but will also include other environmental and context data, these including machine usage information (for example high load). Because predictive analytics relies on new sensor and data processing infrastructure, leading manufacturers are changing their business models and investing in new infrastructure themselves (to achieve economies of scale) and increasing their customers' we rent machines. Industry 4.0 is becoming popular in the manufacturing industry in the use of cyber-physical systems, this is as a result of the evolution of the old manufacturing processes, the presence of a large number of sensors and other data streams, and on the other hand, the pressing need to connect all data with the use of communication networks and usage scenarios. It leads to Potential benefits. Industry 4.0 represents the introduction of IT into manufacturing and poses many challenges 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[16]. This is done automatically through data analysis. Smart data and service integration a further development of the Industry 4.0 scenario described above focuses on services that solve the task at hand. To be able to use smart services to address big data usage issues, technical and organizational aspects need to be considered. Privacy and data protection issues, regulatory issues, and new legal challenges (such as ownership issues for derived data) all need to be addressed. On a technical level, there are several aspects that need to enable service interaction. At the hardware level, from individual machines to facilities to networks. Conceptual level from smart devices to intelligent systems and decision making. At the infrastructure level, from IaaS to PaaS, SaaS, new services for big data consumption, 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 [22]. This means there is an increased need for data analysts to explore datasets and analysis. This is addressed through new dynamic techniques of visual analysis and his data visualization, but requires a new user interface with new functionality for exploring data. The unified data consumption environment provides support for: B.'s historical mechanism and the ability to compare different analyses, different parameter settings, and competing models. Decision Support Current decision support systems use these techniques insofar as they are based on static reports, but with sufficient dynamic usage to realize the full potential of exploratory search. However, these groups include the following business objectives in order of complexity:

- Search: At the lowest level of complexity, data is retrieved only for a variety of purposes. This includes fact retrieval and searching for known items such as: B. Verification Purposes.
  Other features include navigation through records and transactions.
- Learning: At the next level, these capabilities support knowledge acquisition and data interpretation to enable understanding. Supported features include data comparison, aggregation, and consolidation. Additional components may support social functionality for data sharing. Learning examples include simple searches for specific items

(knowledge acquisition), such as celebrities or their use in advertising (retail). Big data search applications are expected to search all relevant data and present an integrated view.

- Research: At the highest level of decision support systems, data can be analyzed, collected, and synthesized. This includes support for exclusions, negations, and tools for evaluation. At this level of analysis, real-world discoveries are supported and tools influence planning and forecasting. At a higher level, his research (discovery) seeks to find important correlations, such as the effect of season or weather on the sales of a particular product at a particular event. Additionally, examples of the use of big data for strategic business decision making, especially at high levels, are provided. At a higher level, these functions can also be (partially) automated to provide predictive and even prescriptive analytics. The latter refers to automatically derived and implemented decisions based on the results of automated (or manual) analysis.
- However, such capabilities are beyond the scope of typical decision support systems and include complex event processing (CEP) more likely to be included in the environment. Depending on your application scenario, security often requires frequent maintenance, such as: B. Aerospace industry. However, there are times when the cost of machine failure is not catastrophic and determining maintenance intervals is a purely economic issue. The assumption underlying predictive analytics is that if you have sufficient sensor information about a particular machine and a sufficiently large database of sensor and error data for that machine or a general type of machine, then time will be more accurately predicted.

This approach offers the following cost savings:

- Longer maintenance intervals as 'unnecessary' interruptions to production (or work) are avoided when scheduled maintenance time is reached. Predictive models can extend maintenance intervals based on current sensor data.
- Based on sensor data and predictive maintenance that requires early maintenance, fewer failures occur sooner than planned maintenance, thereby reducing the number of failures.
- Predictive maintenance with a certain amount of advance warning can predict potential failures, reducing failure costs, scheduling maintenance/replacement, and reducing downtime. Visualization of analysis results, including displaying trends and other predictions with appropriate visualization tools, is an important aspect of big data utilization. Selection of relevant parameters, subsets, and features is a key element of data mining and machine learning, and many cycles are required to test different configurations. Since the settings are evaluated based on the presented analysis results, high-quality visualization allows a quick and accurate assessment of the quality of the results. When validating the predictive quality of the model by comparing the result to other results on the test dataset. Without visualization support, this can be an expensive and time-consuming process, making visualization an important element in data analysis. To use the results of data analysis in subsequent steps in a data usage scenario, for example to create a large set to enable data scientists or business decision makers to draw conclusions from the analysis, you need to The visual presentation chosen can be important. Manageable and effective.

Depending on the complexity of the visualization, it can be computationally intensive and prevent interactive use of the visualization. However, exploratory search in analysis results is essential in many cases of big data use. In some cases, the results of big data analysis only apply to one instance, such as an aircraft engine. However, analytical datasets are often as complex as the underlying data, pushing the limits of classic statistical visualization techniques, and requiring interactive exploration and analysis. Shneiderman's seminal work on visualization identifies seven types of tasks: overview, zoom, and filter, details on demand, retrieval, history, and extraction. Another area of visualization concerns data models, which are used in many machine learning algorithms and differ from traditional data mining and reporting applications. When such data models are used for classification, clustering, recommendation, and prediction, their quality is tested against well-understood data sets. Visualization supports validation and configuration of models and their parameters. Finally, the huge size of datasets is a continuing challenge for visualization tools, driven by technological advances in GPUs, displays, and slow adoption of immersive visualization environments such as caves, VR, and AR. I am. These aspects are covered in the fields of science and information visualization. T. Becker The next section describes the application of visualization for working with large amounts of data, called visual analytics. Presents a number of research questions related to visualization in general.

d. 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 [30]. 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. In line with the growth of the data volume, the installed base of storage capacity is also predicted to increase at a compound annual growth rate of 19.2%. There is enormous volume of unstructured data which keeps growing and there is value stored in the data that can be exploited.

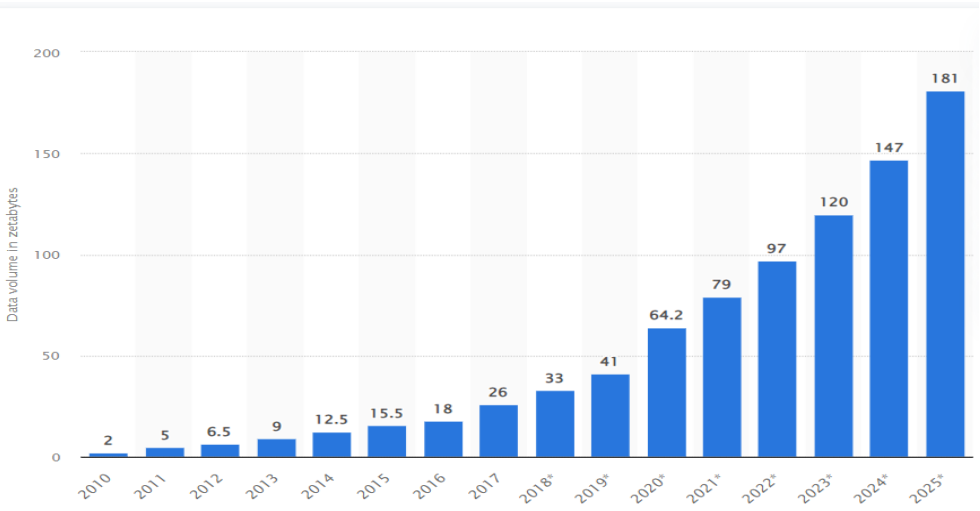

Figure 7: The Evolution of Data

e. 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 [17]

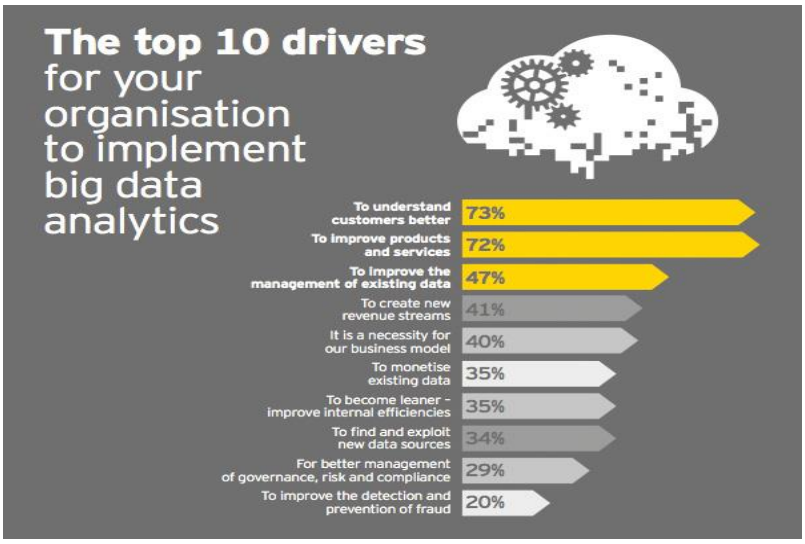

Figure 7: Drivers of Big Data Analytics

f. Factors to consider when selecting an Analytics solution

i) Internal Factors

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

a. The expected investment costs suitable to achieve the analytics goals.
b. The elaborate plan and the analytics road map to achieve the goals.
c. A forecast of the future benefits and rewards of the investment, where the envisioned benefit of the implementation of the solution should lead to.
d. The needs of the organisation from a data and analytics perspective, what are the targets does the solution seek to achieve.
e. Stage of the organisation's analytics maturity? Capability assessments, does the infrastructure support the solution [22]

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

ii) External Factors

a. What competition is currently doing in terms of their analytics frameworks and how the firm can use its own analytics resources to counter and increase competitive in your marketplace?
b. Any prohibitive laws relating the type of data to be collected and analysed.
c. The Plan available to create additional value for the customers in the present and in the future.

g. Value Creation potential of BDA

i) Understanding Customer Needs

Today, companies are taking increasingly precise marketing measures to remain competitive and maintain or increase profit margins. Therefore, predictive models are widely used in precision marketing to understand and meet customer needs and expectations. There is increasing focus on using predictions from customer data and transaction records to analyse consumer behaviour and preferences and manage product supply chains (SCs) accordingly. In the current operating environment, fostering business relationships and understanding customer needs is very crucial for firms to grow. The term "people by from people" is important in the business world, BDA is used to foresee and predict the demands of consumers and will assist SMEs in understanding their customer universe. Micro-targeting will then enable the SMEs to engage potential clients and will help in value proposition. It helps businesses better understand their customers by creating customer personas, identifying their best customers, and collecting as much data as possible. Big data analytics software is used to analyse the collected data and use the insights gained to improve the customer experience. According to SAS, big data analytics enables companies to provide customers with what they want, when they want it, by measuring customer satisfaction through analytics. Big data analytics also gives more companies the opportunity to develop innovative new products that meet the changing needs of their customers. Forrester stated that companies must learn how to leverage customer insights and quantify the business impact of their customer experience (CX) efforts, or find themselves in a vulnerable position. Another Adobe report found that companies considered "CX leaders" are three times more likely to achieve their business goals. Many predictions are only a proxy for reality, but Forrester, Adobe, and others are accurate about CX. Customer experience is a competitive advantage in every industry, from finance and insurance to healthcare, logistics and retail. Big data is essential to creating the personalized experiences customers expect, at scale, without friction, and with the human element intact. Businesses are increasingly investing in smarter CX tools that leverage AI, machine learning, and advanced analytics to understand their customers and improve the customer experience.

ii) Improving Operational Efficiency

BDA plays a crucial role in the growth of companies by way of analysing customer behaviour using the available data and thereby helps in implementation of predictive analytics to solve some of the critical issues which are important for 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), this will turn improve the operational efficiency. Information systems have been built in into business processes utilizing data to ensure economic ordering, timeous delivery etc. Big data management improves operational efficiency by analysing customer behaviour based on purchasing data and helps implement predictive analytics to calculate average checkout wait time. According to KPMG, operators use data and analytics to improve operational efficiency, increase productivity, extend asset life, and reduce operating and maintenance costs. Big data analytics allows businesses to give customers what they want, when they want it, by using analytics to measure customer needs and satisfaction. Big data analytics gives more companies the opportunity to develop innovative new products that meet the changing needs of their customers. In summary, big data analytics can help companies improve operational efficiency by analysing customer behaviour, optimizing supply chain efficiency, reducing costs, and improving employee efficiency. By leveraging customer insights companies can develop innovative new products, improve services, and create personalized experiences that meet customers' changing needs

iii) Improved Decision Making

Big Data Analytics help firms to make better and informed decisions by providing insights into factors such as market trends and customer behaviour, this can help companies to identify new opportunities, improve products and services and it can also help companies to identify new opportunities in the spaces where they operate in, this will assist businesses especially the Small to Medium Enterprises to scale up their operations and to grow.

iv) New Revenue Streams

BDA assists firms to identify new revenue streams, through insights into the market trends and customer behaviour, companies can develop products or add services which meet or satisfy the needs of their customers and thereby generate new streams of revenue.

v) Competitive Advantage

Using the insights generated from a BDA, SME's can differentiate themselves from their competitors which will enable them to satisfy the needs of the customers in turn growing and gaining more market share.

vi) Better Customer Experience

BDA helps organisations to improve and offer better customer experience by leveraging on insights into customer behaviour and preferences. This will assist companies to tailor their services and products to meet the demands and needs of

their customers, this will subsequently result in increased customer satisfaction and loyalty.

h. Barriers to Big Data Analytics Adoption for Zimbabwean SME's

i) Data barriers

Big Data Analytics is a strong tool which helps firms to make data-driven decisions which can help their business operations. BDA comes with multiple risks which need to be addressed to guarantee the safety and security of the data. Common problems with data itself include data storage, scalability, data complexity and data quality as the barriers for a Small to Medium Enterprises to adopt Big Data Analysis.

ii) Knowledge and skills barriers

Big data analytics will require skilled personnel who are capable to analyse and interpret the data. There is a shortage in the SME sector of skilled personnel in the field and this can result in inaccurate analysis and poor decisions making.

iii) Regulatory barriers

Regulatory barriers refer to the legal, ethical, privacy aspects which organisations should take note so as to ensure compliance data protection legislation which dictates how businesses should utilize data. The sensitive data collected by companies is the backbone of many big companies, and if it leaks to any wrong hand, like cybercrime or hackers, it can badly affect the business and its reputation. Therefore, companies should mainly focus on protecting their data's privacy and security from malicious attacks. Big data is not easy to store in pockets; companies need to manage big servers to hold this crucial information and protect it from the outside world. It's a very challenging and risky process, but it's a need for businesses to keep their big data protected.

iv) Technical barriers

These refer to the infrastructure required to facilitate BDA. They consist of hardware and software and may include security and the requirement of high bandwidth internet connection to support Cloud-based services.

v) Organisational barriers

These refer to issues behind the organisational culture, structure, top management support and the lack of a strategy. The financial constraints an SME encounters e.g. cost of procuring Big Data Analytics technology and the resources required to implement and use it. Big Data Analytics is capable of assisting organisations to grow in revenue and sales by providing valuable understanding into factors such as customer behaviour and market trends. Nevertheless, the adoption and implementation of a BDA framework requires meticulous planning and execution. Small to Medium Enterprises will require considering the following factors when implementing a BDA framework:

a) *Data Quality:* SME's should ensure that the data they collect is accurate consistent and complete.
b) *Data Security*: Data should be secure from unauthorized access.
c) *Infrastructure:* This will include the hardware, software and network infrastructure.
d) *Skills:* This will include skills in areas like data management, data analysis, and data visualization.

i. 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 [27]. The model brings together the technology-organisation-environment (TOE) model which can feed into the strategy of the organisation to enhance the financial and market performance. In the context of Zimbabwe the following model will be used which involves the following key steps:

a) An assessment of the of the data management protocols and systems of the firm
- Identification of sources of data.
- Auditing the tools of capturing the data as well as the information quality of the data.
- Assessing how the data is being stored.

b) Understanding the business
- Firm's product basket.
- Competitor analysis.
- Supply chain.
- Product margins

c) Outlining Performance Metrics To Be Measured
- Defining the KPI to be tracked (e.g. Sale Volumes, Revenue etc.).
- Establishing the current performance.

d) Selection of appropriate BDA Technology
- Basing on the objectives of the business, a selection of the appropriate BDA technology.
- The technology will include the hardware and software infrastructure.
- Set up a system and define the how the data will be captured, how it will be stored and processed.

e) Training of users
- Training and equipping of the users to handle and interpret the data.

f) Monitoring and Evaluation
- Tracking of the KPI's identified and making necessary changes in line with the objectives.

The framework is not exhaustive, changes and other improvements can be effected according to the nature of the firm and also in response to changes in the environment.

B. 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.

III. METHODOLOGY

A. 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 choosing and was found to be necessary in the context of the study. The selected pool of companies was contacted to explore the importance and challenges in the adoption of BDA among SMEs in Zimbabwe.

B. Population and Sampling

a) 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 were 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 and the surrounding towns.

b) 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). 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 [4].

C. 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.

D. 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.

E. 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.

IV. DATA ANALYSIS AND INTERPRETATION

A. Data Presentation and Analysis

a) Response Rate Evaluation

Table 1: Research Response Rate

Number of Questionnaires Administered	Number of questionnaires completed	Percentage Response Rate
100	68	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 respondent's 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.

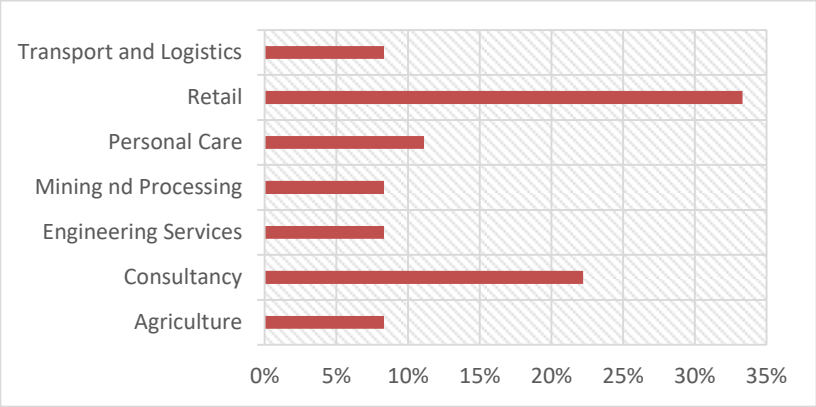

Figure 8: Distribution of Respondents by Industry

Figure 9 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%.

b. Demographics

i) Gender Demography

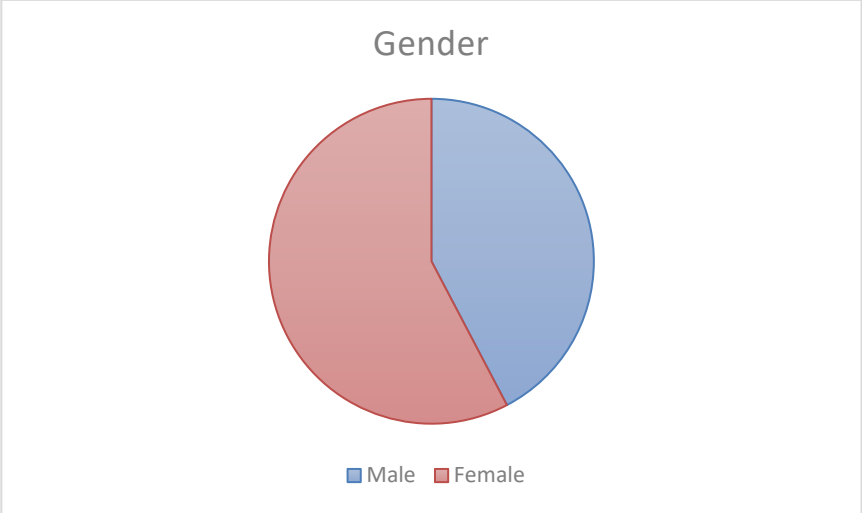

Figure 9: Gender Demography

From Figure 10 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.

ii) Business Registration

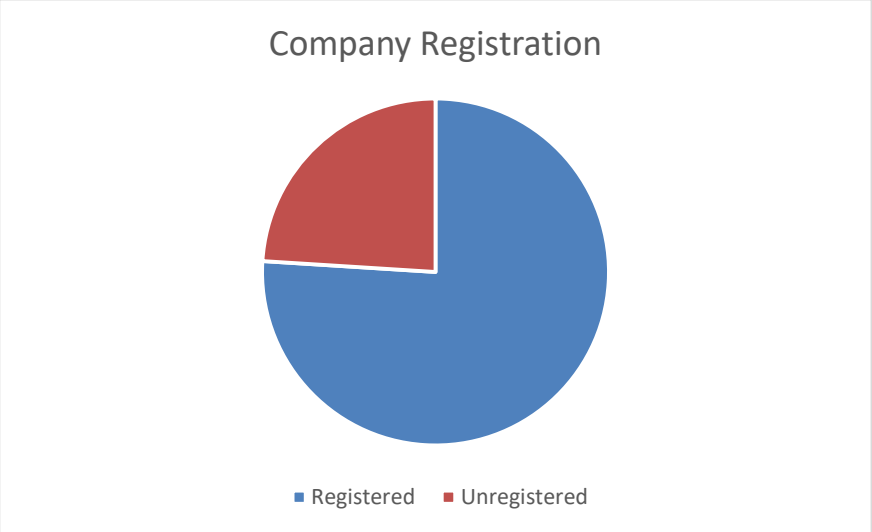

Figure 110: Share of Registered and Unregistered Respondents

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.

iii) 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

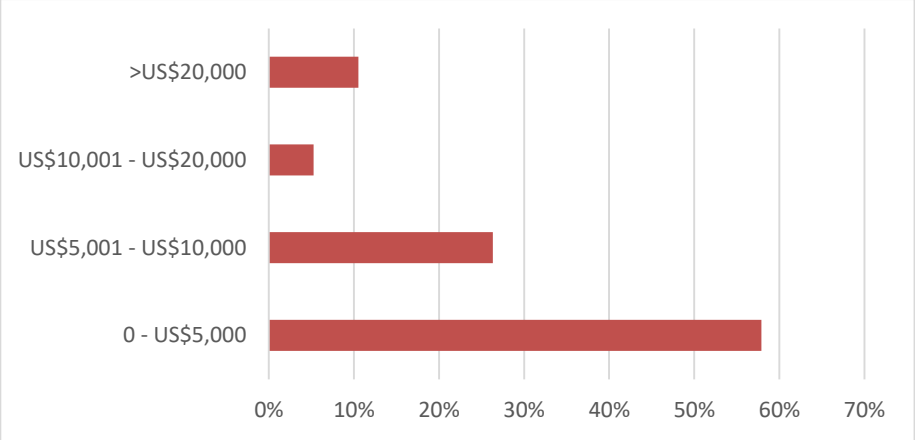

Figure 12: Revenue Bands

Table 3: Responses and revenue category

Revenue Category	Frequency	Percent (%)	Valid (%)	Cumulative %
0 - US$5,000	11	58%	58%	58%
US$5,001 - US$10,000	5	26%	26%	84%
US$10,001 - US$20,000	1	5%	5%	89%
>US$20,000	2	11%	11%	100%
Total	19	100%	100%	

Figure 12 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.

iv) 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.

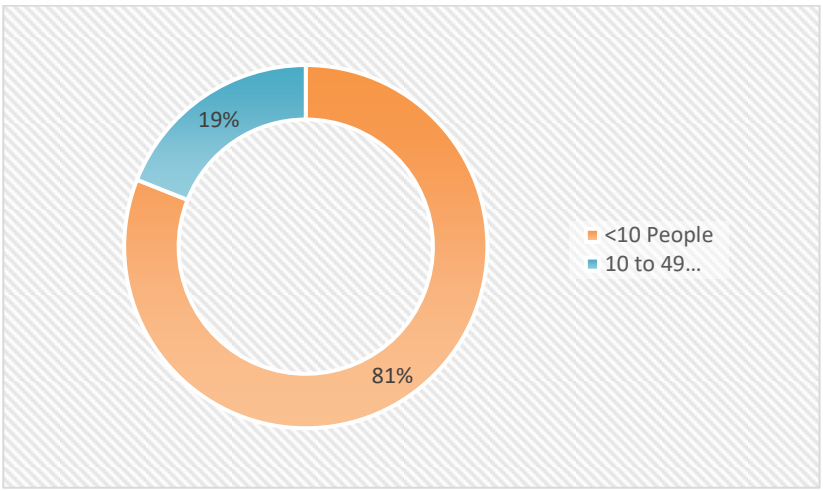

Figure 13: Firm Size

81% of the respondents run companies which employ less than 10 people and 19% of the respondent's 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.

v) 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.

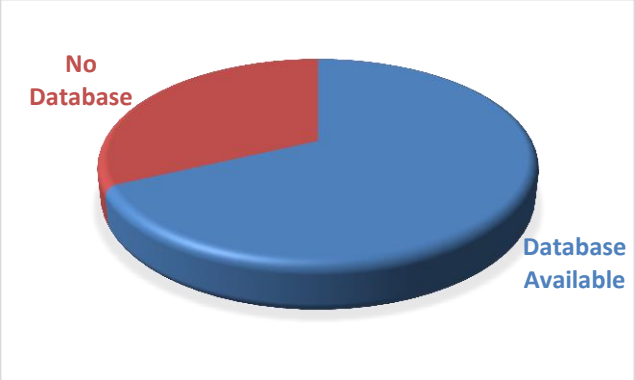

Figure14: Database Availability

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

b) Data capture and information storage (manual or digital)

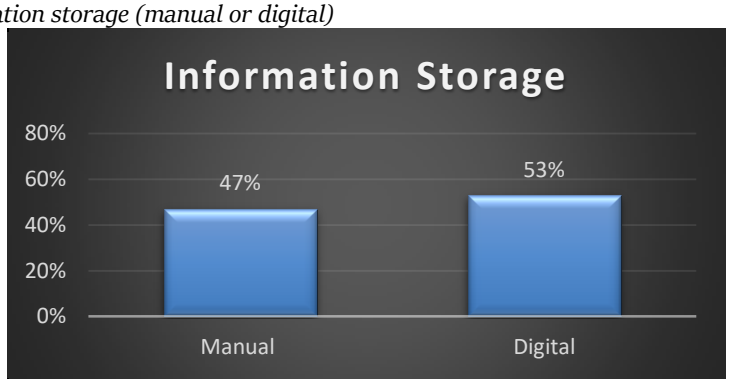

Figure 115: Information Capture and Storage

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.

B. Discussion and Interpretation

a) 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

b) 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.

c) 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.

IV. CONCLUSION

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.

A. 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.

B. 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;

a) 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).

b) 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.

V. REFERENCES

[1] 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, 17(4), 23-49.

[2] Armbrus, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., & Konwinski, A., et al. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50-51.

[3] 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, 4(12), 84-98.

[4] Etikan, I., Musa, S. A., & Sunu, R. (2015). Comparison of Convenience Sampling and Purposive Sampling. American Journal of

Theoretical and Applied Statistics, 4(1), 1-3.
[5] Kramer, M. P., Bitsch, L., & Hanf, J. (2021). Blockchain and its impacts on Agri-Food supply Chain network Management. MDPI, 2, 1-10.
[6] Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. International Journal of Information Management, 37(1), 114-124.
[7] Soltanpoor, R., & Sellis, T. (2016). Prescriptive Analytics for Big Data. In Australasian Database Conference). Sydney: Springer International Publishing, 246-326.
[8] 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, 3(2), 1-15.
[9] Anokye, A. M. (2020). Sample Size Determination in Survey Research. Journal of Scientific Research and Reports, 10(3), 93.
[10] Brozovic, D., Jansson, C., & Boers, B. (2023). Strategic Flexibility and growth of small and medium-sized enterprises: a study of enablers and barriers. Emerald Publishing, 15, 1-20.
[11] 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, 12(2), 127-141.
[12] Calheiros, R. N., Ranjan, R., Beloglaz, A., De Rose, C. A., & Buyya, R. (2010). CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and Experience, 41(1), 1-15.
[13] Cattell, R. (2010). Scalable SQL and NoSQL Data Stores. Springer Open, 18-20.
[14] Cavanillas, J. M., Curry, E., & Wa, W. (Eds.). (n.d.). New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe. New York: Springer Open.
[15] Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). New Horizons for a Data Driven Economy. New York: Springer Open.
[16] Fitz-Enz, J. (2010). The New HR Analytics Predicting the Economic Value of Your Company's Human Capital Expenditure. New York: American Management Association.
[17] Gandom, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 138-154.
[18] Gudivada, N. V. (2017). Data Analytics for Intelligent Transportation System. Amsterdam: Elsevier Science Publishers.
[19] Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology tutorial. IEEE Access.
[20] Epstein, E. (2013, January). Big data and the creative destruction of today's business models. Kearney.
[21] 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, 145-155.
[22] Krol, K., & Zdonek, D. (2020). Analytics Maturity Models: An Overview. Information, 10(3), 1-10.
[23] Lepeniotia, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. International Journal of Information Management, 40(1), 57-70.
[24] Manzoor, F., Wei, L., & Sahito, N. (2021). The role of SMEs in rural development: Access of SMEs to finance as a mediator. PLOS ONE, 35(4), 1-12.
[25] María, J. C., & Wahlster, W. (2016). New Horizons for Data Driven Economy. Madrid: Springer Open.
[26] Maroufkhani, P., & Wan, K. W. (2020). Big data analytics adoption model for small and medium enterprises. Journal of Science and Technology Policy Management, 11(3), 1-10.
[27] 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(2), 1-10.
[28] Musabayana, G. T., Mutambara, E., & Ngwenya, T. (2022). An empirical assessment of how the government policies influenced the performance of the SMEs in Zimbabwe. Journal of Innovation and Entrepreneurship, 4(1), 1-10.
[29] North, D., & Smallbone, D. (2010). The Innovativeness and Growth of Rural SMEs During the 1990s. Taylor and Francis Online, 145(2), 1-10.
[30] Perdanaa, A., Lee, H. H., Koh, S., & Arisandi, D. (2021). Data analytics in small and mid-size enterprises: Enablers and inhibitors for business value and firm performance. International Journal of Accounting Information Systems, 20(1), 13-16.
[31] Ridzuana, F., & Mohd Nazmee Wan Zainon, W. (2022). Diagnostic analysis for outlier detection in big data analytics. In Sixth Information Systems International Conference (ISICO 2021
[32] Seyedan, M., & Fereshteh, M. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. Springer Open, 2(1), 1-10.
[33] Wamba, S. F., Gunasekaran, A., Shahriar, A., Steven, R. J.-F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 82, 357-371.
[34] Zhao, L. (2020). Event Prediction in the Big Data Era: A Systematic Survey. Emory University.