BIG DATA MANAGEMENT AND BUSINESS VALUE IN THE COMMERCIAL BANKING SECTOR IN KENYA

BY

WINNIE MBALUKA

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE IN MASTER OF BUSINESS ADMINISTRATION

UNIVERSITY OF NAIROBI

NOVEMBER, 2013.

DECLARATION

This research project is my original work and has not been presented for a degree in any other University.

Signature: _____ **Date**: _____

WINNIE W. MBALUKA

D61/63167/2011

This project proposal has been submitted for presentation with my approval as University of Nairobi supervisor.

Signature: _____ **Date**: _____

Dr. KATE LITONDO

Lecturer

Department of Management Science

University of Nairobi.

ACKNOWLEDGEMENTS

This research paper has been written in order to finish my master “Business Administration” track “Management Information Systems” at the University of Nairobi. I have always been interested in technical matters and my goal is to make innovations, especially within the information technology field, more tangible by translating these innovations to business opportunities. I believe the phenomenon big data is such an innovation that in fact is not just one innovation but rather, a sum of innovations that even may create a new era in information technology. The project was not created by the author alone, but relied on the cooperative assistance of many unseen hands. I owe special thanks to God Almighty for seeing me through.

I sincerely acknowledge my project advisor, Dr. Kate Litondo for her invaluable help and constant encouragement throughout the course of this research. I am most grateful for her teaching, advice and suggested approaches that gave me direction of the project. I would not have achieved this far and this paper would not have been completed without all the support that I have always received from her.

In addition, I acknowledge encouragement from the department’s lecturer’s and my colleagues both at work and at school throughout the project work. Special thanks to all interviewees. Without them, getting sufficient data was just impossible.

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

DEDICATION

I dedicate this project to my entire family and my supervisor for their hard work, support and encouragement through the entire project.

ABSTRACT

Kenya's financial services industry is one of the most prominent in the economy, within it banking possibly the most vibrant sector. The financial services industry is amongst the most data driven and increasingly facing the challenge of big data where unprecedented quantities of structured and unstructured data stored in a variety of systems and formats. Identified as the biggest information technology trend for 2012, big data gets global attention and can be best described using volume, variety and velocity. Big data refers to environment in which data sets have grown too large to be handled, managed, stored and retrieved in an acceptable timeframe. Given the enormous growth of data, banks are suffering from their inability to effectively exploit their data assets. Central Bank of Kenya develops appropriate laws, regulations and guidelines that govern the players in the banking sector. The regulatory environment that commercial banks operate within requires these institutions to store and analyze many years of transaction data. In response to new regulations, banks need to have a horizontal view of risk within their trading arms. Providing this view requires banks to integrate data from different trade capture systems, each with their own data schemas, into a central repository for positions counter-party information and trades. Banks have relied on relational technologies coupled with business intelligence tools to handle this ever-increasing data burden. This research paper focused on the extent of big data management in commercial banks, benefits of big data management, challenges in big data management and the effects of big data management on business value. This study adopted a descriptive survey. The population of the study consisted of all the 43 commercial banks and 1 mortgage finance company in Kenya. Primary data was collected using structured questionnaires and it was presented in pie charts, bar chart and percentages. The results of this research clearly indicate a great deal of activity around planning and implementing big data management environments. This study showed that indeed the banking sector in Kenya is in the very early stages of the big data management initiatives and banks are using various big data management techniques and tools and at the same time struggling to keep pace. There were a number of challenges in managing big data that were identified which hinder big data implementations. Finally, this study also showed that more than half of the banks studied have invested or are planning to invest in better understanding and utilization of the data they collect with the rest of the banks under study not having any big data management initiatives. For organizations, this paper can act as an eye-opener as it shows the potential of big data management. When it comes to banks there is need to understand immediate and strategic information needs of the organization before bringing big data management initiatives on board. Additional research might focus on other markets or types of big data analytics such as real-time analytics, which receives an increasing amount of interest. In addition, its impact on competitive advantage and the creation or reinvention of new business models, might be studied too, as value creation and competitive advantage are closely related.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

1.1 Background

Data is the raw, unorganized facts that are input into the computer to be processed (Morley and Parker, 2009). According to Gordon (2007) data is a re-interpretable representation of information in a formalized manner suitable for communication, interpretation or processing. Information is processed data that improves our knowledge, enabling us to take decisions and initiate actions. Data can exist in many forms, representing text, graphics, audio and video or moving pictures. Information is an asset which, like other important business assets, has value to an organization. The only resource that is readily available to senior management is information, which as a business resource supports the day-to-day operational tasks and activities, enables routing administration and management of the business and supports strategic decision making and future planning (Gordon, 2005).

Data can be designated as unstructured or structured (standard) data for classification within an organization. Structured data refers to data that is organized by codes, classifications, or common practice in a standard form that computers can read. Unstructured (nonstandard) or semi-structured data are not organized into tables and cannot be natively used by applications or interpreted by a database. In other words, unstructured data has no identifiable structure and is, therefore, not so easily usable (Ohlhorst, 2012). Data are collected to be analyzed to find patterns and correlations that may not be initially apparent, but may be useful in making business decisions.

Data management is the management of data or information represented in a formalized manner suitable for communication, interpretation or processing. Data management is a corporate service which helps the provision of information services by controlling or co-coordinating the definitions and usage of reliable and relevant data. This corporate service strategically supports the management and use of business data and not limited to operationally supporting the development and maintenance of computerized information system. Data management helps in achieving recognition of data, both structured and unstructured as an enterprise-wide valuable business resource (Gordon, 2007).

Bank is a financial institution licensed as a receiver of deposits and giver of loans both short and long term there are commercial/retail banks and investment banks where by in most countries, they are regulated by the national government or central bank. Service offerings in banking are dependent on the type of bank and the country where it operates or originated from. Some of the services offered by banks are opening and maintaining current and savings accounts for individual and businesses, taking deposits from customers, cashing cheques, extending loans, issuing debit cards, cashpoints cards and credit cards, offering financial advisory services, effecting money transactions such as wire transfers and bankers drafts, selling insurance ,travelling cheques, investments in securities, foreign exchange, asset/liability management and cash instruments among others (Corporation Essvale, 2011).

Banks are collecting and fully operating on data, both structured and non-structured and the rapid growth of unstructured data is causing issues for banks. Bank regulations are a form of government regulation which subject banks to certain requirements, restrictions and guidelines. This regulatory structure creates transparency between banking institutions and the individuals and corporations with whom they conduct business, among other things. Regulation and compliance issues have put a strain on data management platforms because they require the delivery of high volumes of data, with low latency and banks must now meet a more stringent fiduciary responsibility to provide correct data to regulatory agencies. Faulty information can result in dire consequences and catastrophic financial exposure (Goel, 2012).

1.1.1 Big Data

Big data refers to data sets that are too large or too fast-changing to be analyzed using traditional relational or multidimensional database techniques or commonly used software tools to capture, manage and process the data at a reasonable elapsed time (Slack, 2012). The data is too big, moves too fast, or doesn't fit the structures of database architectures (Dumbill, 2012). Big data is overwhelming not only because of its volume, but also because of the diversity of data types and the speed in which it must be managed. Volume, velocity, and variety which are often referred to as the

three V's of Big Data, capture the true meaning of big data (Krishnan, 2013). Big data represents a trend in technology that is leading the way to a new approach in understanding the world and making business decisions. These decisions are made based on very large amounts of structured, unstructured and complex data (e.g., tweets, videos, commercial transactions) which have become difficult to process using basic database and warehouse management tools. There are two types of big data one being data in movement (twitter/Facebook comments, stock market data, sensors) and the other one is data at rest (collection of what has streamed, weblogs, emails, social media, forms, claims and data from disparate systems) Big data is a combination of transactional data and interactive data (Mohanty et al., 2013). Big data analytics refers to a technology and framework for quickly storing, converting, transferring and analyzing massive amounts of constantly updated, huge, varied, structured and unstructured data for commercial gain (Russom, 2011).

The sheer volume of data being stored today is exploding. In the year 2000, 800,000 petabytes of data were stored in the world and by 2020 it is anticipated to hit 35 zettabytes. As the amount of data available to the enterprise is on the rise, the percent of data it can process, understand, and analyze is on the decline (Zikopoulos et al, 2011). The actual quantity will vary by case for each industry sector, organization, and application (IDC, 2009). The volume associated with the big data phenomena brings along new challenges for data centers trying to deal with it: its variety. With the explosion of sensors, and smart devices, as well as social collaboration technologies, data in an enterprise has become complex, because it includes not only traditional relational data, but also raw, semi structured and unstructured data from web pages, web log files, search indexes, social media forums, email, documents, sensor data from active and passive systems and so on (Manyika et al., 2011). Just as the sheer volume and variety of data we collect and store has changed, so, too, has the velocity at which it is generated and needs to be handled (Zikopoulos et al, 2011).

Velocity typically considers how quickly the data is arriving and stored, and its associated rates of arrival. Today's enterprises are dealing with petabytes of data instead of terabytes, and the increase in RFID sensors and other information streams has led to a constant flow of data at a pace that has made it impossible for traditional systems to handle (Manyika et al., 2011). IDC has further added value as the fourth

characteristic of big data. Value refers to the cost of the technology and the value derived from the deployment of the technology (IDC, 2009).

Big data management is the organization, administration and governance of large volumes of both structured and unstructured data (Mohanty et al., 2013). The goal of big data management is to ensure a high level of data quality and accessibility for business purposes. Effective big data management helps companies locate valuable information in large sets of unstructured data and semi-structured data from a variety of sources. As part of the process, companies must make a judgment call on how much data will be retained and what can be disposed of. That requires careful data classification to categorize information according to factors such as its potential uses and business value (Zikopoulos et al, 2011). Additionally, the Centre for Economics and Business Research (CEBR) in UK predicts that the most beneficial sectors from big data analysis will be the financial services, public sector, retail and manufacturing (CEBR, 2012). A wide variety of techniques and technologies have been developed and adapted to aggregate, manipulate, analyze, and visualize big data. These techniques and technologies draw from several fields including statistics, computer science, applied mathematics, and economics (Zikopoulos et al, 2011).

The big data techniques include visualization, text analytics, in-memory analytics, predictive analysis, graph analytics, mobile business analytics, video analytics , audio analytics, A/B testing, association rule learning, classification, cluster analysis, crowd sourcing, data mining, data fusion and data integration, ensemble learning, genetic algorithms, machine learning, natural language processing, network analysis, optimization, pattern recognition, predictive modeling, regression, sentiment analysis, signal processing, spatial analysis, statistics, complex event processing, SaaS-based business analytics supervised learning, simulation, unsupervised learning, time series analysis and Social network analysis. Most big data environments go beyond relational databases and traditional data warehouse platforms to incorporate technologies but not exhaustive include business intelligence, Hadoop, HDFS, MapReduce, Mashup, Hive Hadoop ,MongoDB ,Pig Apache,Riak ,CouchDB ,SCOPE, Non-relational database, stream processing, NoSQL, Cassandra and massively parallel-processing (MPP) databases among others (Manyika et al., 2011).

1.1.2 Business Value of Big Data Management

Business value is being used in management and financial economics as an informal term that includes all forms of value that determine the health and well-being of the firm in the long run (Bomarius, 2009). According to the value chain analysis, value is created when value activities can be performed at lower cost or result in more sales. To achieve this, according to Schumpeterian's innovation theory and resource-based view, resources can be combined to create new products, services, and production methods among others (Sundbo, 1998). Data is the new oil, and like oil, data needs to be refined before its get value. Big data can be used to create value across sectors of the economy, bringing with it a wave of innovation and productivity gains. The true value of big data lies not just in having it, but in being able to use it for fast, fact-based decisions that lead to real business value (Minelli et al., 2012). Companies are relying on data to run businesses, and are using analytics in a predictive way to enable them to be more proactive in their decision making and less reactive (Dumbill, 2012).

There are five broad ways in which using big data can create value. First, big data can unlock significant value by making analysis of large data set more transparent, user friendly and usable at much higher frequency (Schönberger and Cukier, 2013). In the past the process of collecting and analyzing large sets of information was difficult, cumbersome and time-consuming. Second, as organizations create and store more transactional data in digital form, they can collect more accurate and detailed performance information (in real or near real time) on everything from product inventories to personnel sick days, and therefore exposes variability and boost performance. Leading companies are using data collection and analysis to conduct controlled experiments to make better management decisions; others are using data for basic low-frequency forecasting to high-frequency now casting to adjust their business levers just in time (Rahman and Ramos 2013).

Third, big data allows ever-narrower segmentation of customers and therefore much more precisely tailored products or services. For example, instantly generating information about a target customer base for tailored promotions, taking advantage of a particular market condition, will mean the difference between a successful marketing campaign and a costly marketing failure for a bank (Mruthyunjayappa, 2011). Fourth, sophisticated analytics can substantially minimize risks, unearth

valuable insights that would otherwise remain hidden, improve decision-making by analyzing entire datasets from customers, employees, or even sensors embedded in products (Minelli et al., 2012). Finally, big data can be used to improve the development of the next generation of products and services. For instance, banks currently perform analysis of customer data to obtain insight to improve customer offers and to deliver targeted personalized offers (Manyika et al., 2011). However, the influx of data presents many barriers to effective analytics, and to the creation of business insight for most decision makers (Adhikari, 2012).

1.1.3 Commercial Banks in Kenya

Financial liberalization of early 1990s in Kenya opened the banking industry to a number of players leading to stiff competition and weakening of financial performance of a number of commercial banks leading to collapse of some. In Kenya, the banking sector plays a dominant role in the financial sector, particularly with respect to mobilization of savings and provision of credit. In Kenya, there are 43 licensed commercial banks and 1 mortgage finance company with a total asset of Kshs. 510.6 billion. The Banking industry in Kenya is governed by the Companies Act, the Banking Act Laws of Kenya Chapter 488, the Central Bank of Kenya Act and the various prudential guidelines issued by the Central Bank of Kenya (Government of Kenya, 2011). Central Bank of Kenya (CBK) prudential guidelines requires banks to store and analyze many years of transaction data. Banks should provide data on its liquidity ratio and conduct detailed analysis of all available data to assess the level of anti-money laundering risk within each high risk category. When assessing the anti-money laundering risk banks are required to obtain quantitative data pertaining to the institution's activities, assign different weights to each product/service, customer/entity and geographical based on the institution's ability to identify the customer, number and volume of transactions and past history (CBK, 2013).

Central Bank of Kenya requires banks to use computerized system, for internal or external application database checks to check for any inconsistencies in the information provided particularly those containing known fictitious application/ fraud

information. Consumers' financial and personal information should be protected through appropriate control and protection mechanisms. The protection mechanisms should define the purposes for which the data may be collected, processed, held, used or disclosed (especially to third parties). The mechanisms should also acknowledge the rights of consumers to be informed about data-sharing, to access data and to obtain the prompt correction and/or deletion of inaccurate, or unlawfully collected or processed data. The Central Bank of Kenya risk management guidelines requires institutions to ensure that sufficient items are captured in the system logs to facilitate effective internal controls, system trouble shooting, and auditing while taking appropriate measures to ensure time synchronization on all logs. Sufficient disk space should be allocated to prevent logs from being overwritten. System logs should be reviewed for any exception. The review frequency and retention period for transaction logs or database logs should be determined jointly by ICT function and pertinent business lines, and approved by the IT Steering Committee (CBK, 2013).

The Banking industry in Kenya is governed by the Companies Act, the Banking Act, the Central Bank of Kenya Act and the various prudential guidelines issued by the Central Bank of Kenya. The banking sector was liberalized in 1995 and exchange controls lifted. The globalization of financial markets has led to stiffer competition of the local banks for market share, from foreign owned banks. The Kenyan banking sector remains sound and resilient whereas the Kenyan financial sector is developing and deepening faster than the overall economy. This is according to the Central Bank of Kenya in a just released Banking Sector Development Report for the second quarter ended 30th June 2012 (CBK, 2013).

The Central Bank of Kenya has put forward risk management guidelines for banks so that in events of financial risk, there is no direct loss of earnings or imposition of constraints on the bank's ability to meet is business objectives. The risk management program of each institution should at least contain the following elements of a sound risk management system which is an adequate risk monitoring and management information systems (CBK, 2013). In accordance with the Basel Core Principles for Effective Banking Supervision, 'Risk Management Processes' requires that banks and banking groups must have comprehensive risk management processes (including Board and senior management oversight) to identify, evaluate, monitor and control or

mitigate all material risks and to assess their overall capital adequacy in relation to their risk profile. These processes should be commensurate with the size and complexity of the institution (IMF, 2012).

1.2 Statement of the Problem

The value of data is underestimated, especially in many Kenyan banks. This is probably due to a lack of time and resources. Organizations can create more value by leveraging data they have and create, but due to a lack of knowledge, this is not happening on a large scale. This is changing due to certain developments in different fields, both technological and non-technological. Data is changing and new technologies make it possible to analyze this data easier, faster and cheaper. Hence, it becomes easier to extract valuable insights from data, even for the small banking institutions (Cotte, 2012). Swoyer (2012) argued that existing definitions of big data miss out on the most important thing, which is the way big data creates value for organizations. Burgel (2012) also further argues this by saying that existing definitions lack the implication aspect of big data management. The velocity, volume and variety give a comprehensive description of the evolution of data but simply miss the consequence part of big data management and how it affects an organization.

This study tries to fill that gap by showing connections between big data management and creation of business value in the commercial banking sector in Kenya, something which has not been done before. This research is also due to a further recommendation from a research on effects of big data analytics on organizations' value creation. There was need to study more international cases to find out if the results would be more generic, hence improving the validity of the results obtained from the research (Mouthaan, 2012).

Data can create significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector and creating substantial economic surplus for consumers (Lehdonvirta & Ernkvist, 2011). Thus, the present study is also justified by the fact that big data helps the banking industry to 'manage down' the cost of large scale data management , complying with new regulatory requirements of the Central Bank of Kenya and helping banks to derive business

value (Oracle, 2012). Big data management is a big challenge for financial services firms. The sheer amount of storage required to keep it strains IT infrastructure, crowding physical storage capacity. Worse, Big Data overwhelms management capabilities and also there are insufficient data governance practices. Embracing big data management initiatives can help banks to identify strategic opportunities and at the same time acquire a competitive advantage (Ohlhorst, 2012).

1.3 Objectives

The general objective was to evaluate the influence of big data management on business value in commercial banks and specifically to:

a) Determine the extent to which commercial banks are using big data management tools.
b) Establish the benefits of big data management.
c) Determine the challenges in management of big data.
d) Establish the effects of big data management on business value.

1.4 Value of the Study

The study will offer a basis for other academic investigations by future academicians and researchers into the area of big data management in Kenyan commercial banks**.** Academicians will benefit from the information of the study as it will contribute to the existing body of knowledge. The study will also provide further background information to research organizations and scholars who will want to carry out further research in this area.

The study will also benefit the top management in the banking industry since they are the ones who will use it to make informed decisions on how to analyze, predict and deliver more targeted, personalized and effective offers to their customers. They will be able to deliver improved actionable insight during customer interactions with their contact centers. Big data will provide the Banking Fraud Investigation Division with a

scalable, integrated, secure and cost effective platform to more quickly prevent, detect and mitigate internal and external frauds.

Big data can be of great help to politicians in that it played a large role in Barack Obama's successful 2012 re-election campaign. There is no doubt that big data helped President Obama win comfortably over his Presidential opponent Mitt Romney in 2012. While President Obama already had a huge database of 13 million email listings from 2008, Mitt Romney had to build his social media campaign from scratch as the Republicans did not indulge in social media marketing in the year 2008. On the other hand, President Obama's social media strategy in 2012 involved tapping those 13 million people who supported him in 2008 Presidential elections through emails, Facebook, Twitter and other social media platforms.

Big Data brings the potential to transform the work of government agencies by helping government agencies operate more efficiently, create more transparency and make more informed decisions. The data stores that various government agencies accumulate over the years offer new opportunities for agencies which can use Big Data technologies to extract insights and keep track of citizens' specific needs. In turn, these insights could then be used to improve government services.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

The literature gathered focuses on extent of big data management in commercial banks, benefits of big data management, challenges in big data management and effects of big data management on business value.

2.2 Extent of Big Data Management in Commercial Banks.

Big Data technology is used to store, convert, transmit and analyze large quantities of dynamic, diversified data (which may be structured or unstructured data) for the purpose of commercial benefit (Borkar et al., 2012). Big data technology applications need to be able to undertake real-time, high-complexity analysis of vast amounts of data, to help business enterprises perform decision-making within the shortest possible timeframe (Bryant, Katz, and Lazowska, 2008). In the financial services industry, while there is a great deal of discussion around big data, many banks are just beginning to consolidate and utilize many of the internal data elements at their disposal, such as debit and credit transactions, purchase histories, channel usage, communication preferences, loyalty behavior, etc. Banking is a data-driven business, and banks must be the trusted custodian of excess customer and corporate data. Emerging from the Great Recession, banks are faced with the overwhelming task of effectively running their business and also meeting the requirements of present and future regulations. The drivers of change include regulatory reforms still pending following recent failures, ailing consumer and business sentiment, the continuing economic crisis in some markets and rapid expansion in others, the availability and cost of new technology and business models, and industry consolidation (IDC, 2009).

Regulatory compliance and risk management demands often lead to large IT outlay. In response to new regulations, banks need to have a ‘horizontal view’ of risk within their trading arms. Providing this view requires banks to integrate data from different trade capture systems, each with their own data schemas, into a central repository for positions counter-party information and trades. Regulatory pressure however dictates that this entire process of risk analysis is done many times every day. Moreover,

various risk scenarios need to be simulated, and it's not uncommon for the simulations themselves to generate terabytes of additional data every day (Lemieux, 2012). Up until now, despite possessing a large stash of structured data sources in their databases, banks found it virtually impossible to glean 360-degree views of their customers for intelligent decision making but this is now a thing of the past (Ramachandran and Malladi, 2013). The recent emergence of technologies designed to address big data, however, has ensured that today, banks have the capacity to process huge volumes of unstructured data, typically in the form of emails, SMS text messages, customer feedback among others (Ohlhorst, 2012). Customer relationship management initiatives drive to effectively manage customer interactions, and in the process improve customer retention, customer loyalty and revenue per customer. Related to this customer relationship management initiatives are the building of data marts and data warehouses and achieving the single customer view to better utilize customer information. Banks are moving from a product-centric approach to a client-centric approach with a 360-degree understanding of their clients to better manage and maintain client relationships. To remain competitive in cut-throat business environment and create tangible value for customers, banks require an integrated data analysis solution that can process and analyze the deluge of unstructured data generated every day to make effective, customer-focused decisions (Ramachandran and Malladi, 2013).

In addition to technology and regulatory pressures, firms continue to be squeezed by global economic instability, expectations to increase revenue and reduce costs, and persistent pressure to optimize capital and enhance customer engagement (Goel, 2012). The performance of key markets and economies will continue to impact the financial performance of financial services institutions. Given the current short-term forecasts, this means a continuing cost-conscious environment in established markets and a higher relative spend in emerging economies with uneven spending worldwide overall. Increasing revenue has become challenging for banks in particular as they seek ways to replace revenue lost to regulations, including interchange and penalty fees (IDC, 2009). Regulation and compliance are also key drivers in the march towards an improved data management platform. The emergence of Basel II, Sarbanes-Oxley and other key risk and compliance considerations has forced firms to place high priority on production of accurate and timely data to feed internal risk

management systems (Tarantino, 2008). These factors have put a strain on data management platforms because they require the delivery of high volumes of data, with low latency and these institutions must now meet a more stringent fiduciary responsibility to provide correct data to regulatory agencies. Faulty information can result in dire consequences and catastrophic financial exposure.

In order for a bank to differentiate itself from others in the competitions, it is imperative to make use of the large pool of available data to offer targeted products and services designed to cater for its customers' specific needs. The advent of technology-enabled social media platforms and their most addictive popularity among millions of users is resulting in the generation of increasingly large volumes of unstructured data on the internet and mobiles. With advent of Hadoop, MapReduce and other big data management related technologies, banks can now leverage on its processing and storage capabilities by complementing it with natural language processing and semantic analytics. This can help them to more accurately capture their customers' needs and preferences (Liu, 2010). There are many insights a bank can draw from customer complaints which include: getting an indication of the customers mood based on his tone, knowing the level of customer service they are offering at their branches, knowing what is making the customer to be unhappy, knowing any inconveniences caused to the customer among others (Ramachandran and Malladi, 2013).

Customer expectations have changed as a byproduct of increased competition. Today's customer expects banks to understand their needs and offer them products that align to their specific requirements. The emergence of social media has played a significant role in bringing about change. While social media has encouraged individual customers to have bigger, more unique expectations than ever before, it also gives the banks an ideal platform to identify and understand these expectations (Liu, 2010). To add to the complexity, banking has witnessed an increasing shift from 'brick and mortar' branches to computers and mobile phones, with modes of payment moving from cash to electronic transfers. While greatly enhancing customer convenience, tis has also had the undesired effect of increasing opportunities for fraud. By using cutting-edge tools to pull and analyze data from structured and unstructured sources, it is possible to create a pattern to trace fraud (Gadd ,Karstedt

and Messner, 2011). Most unstructured data generated in banks include e-mails, documents, customer feedback, customer complaints, internet blogs and websites. This data is mainly composed of text rather than numbers. However, the effort made to utilize this unstructured data to create significant customer intelligence is minimal (Ramachandran and Malladi, 2013). Most analytic-based solutions for converting unstructured data (like tweets) into meaningful customer insights would typically involve technology elements like text mining, natural processing and stochastic-based algorithms (Liu, 2010).

2.3 Benefits of Big Data Management.

These are the things banks stand to gain by analyzing every bit of data available. There are immense benefits that the harnessing of big data can bring to an organization or to society at large; by tapping into information that has so far lain hidden. Big data systems are being implemented in multiple enterprise sectors, including commerce, science, and society (Bryant, Katz, and Lazowska, 2008). Big data serves an important role in enterprise risk management and fraud detection patters in that digital information offers managers a window into the array of risks facing their organizations.

Big data serves an important role in enterprise risk management and fraud detection patterns. With the financial crisis of 2009, risk profiling and fraud detection became top priorities (Gadd ,Karstedt and Messner, 2011). Expanding the use of alternative channel insight and increasing the velocity of data capture, the use of data beyond the institution's firewalls provides an enhanced snapshot of household finances and spending behaviors. For instance, with the addition of alternative device transactions and the ability to track changes in behavior beyond what is occurring with a client's credit account, banks can isolate new fraud or risk triggers (Hurwitz et al., 2013). This added insight and enhanced algorithms provides advantages in effectively reducing risk, managing credit exposure and allowing for timely intervention where necessary. Several challenges in the fraud detection pattern are directly attributable to solely utilizing conventional technologies. Big data can be highly utilized in credit card monitoring to build fraud models (Zikopoulos et al., 2011).

Big data helps in real-time decision making and this leads to better decisions. Big data technology enables analytics to be implemented in (near) real time. The extensiveness of electronic trading has meant that Capital Markets need to capture time-series data and merge it with real-time event processing systems. Leveraging the huge volumes of data on electronic grading for real-time decision support and risk management is now possible thanks to big data management techniques. On-demand risk analytics is now the goal of global banks, especially at a trading desk level. Aggregation of global positions, pricing calculations, and value at risk (VaR) fall within the realm of big data and due to the mounting pressure to speed these calculations, big data technologies enables faster calculations by allowing real time access to in-memory positions data (Sammer, 2012). The stock market is increasingly relying on high-frequency trading to settle accounts. Even Twitter feeds have been mined as program traders have discovered that twitter feeds are correlated to stocks prices (Russell, 2011).

The ability to better understand consumers, seamlessly matching "right-time" offers to a customer's or prospect's needs, allows a financial institution to optimize the management of profitable, long-term customer relationships. Banks are using big data to track, segment and offer new data-driven targeted services to their customers, in order to increase revenue, reduce churn, better predict customer behavior and increase customer engagement. As banking and payments have moved onto mobile and online channels, the opportunities for fraud have expanded. Big Data helps banks analyze credit quality, monitor fraud, reduce customer churn (Liu, 2010). The ability to predict that a particular customer is at a high risk of churning, while there is still time to do something about it, represents a huge additional potential revenue source for every business. Banks are able to develop a customer profile which encompasses engagement, preferences, sentiment and other behavioral aspects (Pagani, 2005).

Big data has churn prediction modeling techniques that attempts to understand the precise customer behaviors and attributes which signal the risk and timing of customer churn. More and better data should also lead to improved customer segmentation and more persuasive communications with customers. Powerful customer segmentations can serve as a blueprint for everything from product development to organizational design (Macmillan, 2006). Banks need to keep pace

with the preferred channels of interaction like mobile and social, and as financial institutions mature, so should their marketing analytics strategies. The addition of a vast amount of relatively unstructured online insight provides an enhanced view to this end, potentially improving both effectiveness and efficiency of marketing efforts. Big Data has the power to help banks use "social listening" to understand customer sentiment, interact and communicate with customers in real time using the channel they prefer (Freedman, 2005).

Layering upon this added insight, the ability to leverage precisely timed geo-locational and event-based targeting at the point of sale (either at the bank or an offer delivered to a smartphone by a merchant partner) provides both marketing and payments advantages that few industries can match (Hurwitz et al., 2013). The impact can break through the wall of promotional noise, leading to improved revenues along with an enhanced customer experience. For example, if a customer has a habit of going to a certain area for shopping or lunch on a regular basis, analyzing the data could provide the foundation for offers that are highly personalized even to the type of food the customer prefers, with knowledge as to the likelihood of offer acceptance. Adding device specific capabilities, the offer could be delivered by short messaging service the most logical time for decisioning. This is the same type of model used by Amazon and other retailers (Gadd ,Karstedt and Messner, 2011).

Integrating external customer data requires identity resolution engines that use multiple factors to match the identity of an external individual (facebook account, etc) with that of the bank's customer. There are a number of big data technologies that enable banks to implement an optimized cross-channel strategy and also able to generate recommendations, which are pushed in real-time to smartphones, Facebook, web-site and call-center channels. At the same time, banks must be mindful of new regulations limiting the use of customer information for marketing purposes (Loshin, 2010). Banks want to ensure their company's reputation is protected by screening potential and existing customers for social, environment and ethical lapses, and a few banks have started to use Big Data to analyze a customer's commercial activities to determine risk potential (Liu, 2010).

Analyzing social media, news, financial reports is a preventive measure against doing business with people and organizations that could produce bad publicity and at the

same time managing incidents of fraud (Mohanty et al., 2013). Correlating data from multiple, unrelated sources has the potential to catch more fraudulent activities for instance the potential of correlating Point of Sale data (available to any credit card issuer) with web behavior analysis (either on the bank's site or externally), and potentially with other financial institutions or service providers such as First Data or Society for Worldwide Interbank Financial Telecommunication, to detect suspect activities (Zikopoulos et al., 2011).

Log analytics is a common use for an inaugural big data project. The information technology department in the bank likes to refer to all these logs and trace data that are generated by the operation of the information technology solutions as data exhaust (Schönberger and Cukier, 2013). Enterprises have lots of data exhaust, unstructured information which is a by-product of the online activities of internet users, and it's pretty much a pollutant if it's just left around for a couple of hours or days in case of emergency and simply purged. Some of the benefits derived from data exhaust are that it has been transformed into value-added click-stream data that records every gesture, click, and movement made on a website (Zikopoulos et al., 2011). Data base administrators in banks derive terrific value by using it for performance optimization analysis. IT departments need logs at their disposal for emergencies and then discarding them as soon as possible. These logs help in looking for rare problems which happen once in a while and are more difficult to diagnose and prevent from occurring again (Babbin et al., 2011)

Big data makes customer experience more exciting in that its capabilities provides banks with the ability to understand their clients at a more granular level and more quickly deliver targeted personalized offers (Pagani, 2005). This enables higher offer and cross sell acceptance rates that improve customer profitability, satisfaction and retention. While banks currently perform analysis of customer data to obtain insight to improve customer offers, many are unable to take advantage of today's mountain of internal and external big data. These leads to increased revenue through better offer response rates, improved cross-selling and greater product penetration, higher asset/balance values and increased customer advocacy and lower campaign and infrastructure costs. Big data platform enables banks to deliver improved actionable insight during customer interactions with the contact center (Liu, 2010). Being able to

anticipate your customer needs and resolve them before they become problems allows banks to deliver timely, concise and actionable insight to contact center agents. This can lead to increased sales, improved customer satisfaction and a reduction in operating costs. Fighting fraud, financial crimes and security breaches, in all forms, is among the most costly challenges facing the finance industry (Zikopoulos et al., 2011).

As quickly as new technology is used to identify fraudsters, they themselves are identifying new ways of defrauding financial institutions. Big data technologies provides a scalable, integrated, secure and cost effective platform to more quickly prevent, detect and mitigate internal and external frauds. Exploding volumes of data required for counterparty risk evaluation, combined with the need for more comprehensive and timely risk metrics, have pushed current infrastructure solutions to their limit (Gadd, Karstedt and Messner, 2011). Big data solution enables banks to more effectively estimate credit exposures and support over the counter derivatives trading. Banks can use big data solutions to leverage publicly available unstructured weblog data and social media feeds and also analyze portfolio data, such as mortgage lending, against huge volumes of data in the public domain, such as property values and other banks' loans, to model various scenarios and use these for product innovation or look into ways of mitigating risks. From a foundational level, big data could provide the insights to develop segmentation strategies based on transactional, behavioral and even social profiles. This would allow the organization to provide a highly personalized, consistent experience regardless of the channel selected by the customer, eliminating traditional silos that create a challenge today (Lacy, Diamond and Ferrara, 2012).

2.4 Challenges in Management of Big Data.

Although the Big Data challenge is daunting, it is not insurmountable, and the opportunity is compelling. Several issues stand in the way of financial services firms making optimal use of the data they gather. Applying big data management faces several challenges related with the characteristics of data, analysis process and social concerns. The first challenge appears in terms of privacy. The privacy and security is

the most sensitive issue, with conceptual, legal, and technological implications. This concern increases its importance in the context of big data. In its narrow sense, privacy is defined by the International Telecommunications Union (Gordon, 2005) as the "right of individuals to control or influence what information related to them may be disclosed". Regulatory requirements may force organizations to work around data privacy provisions and governance which can interfere with cross-border transfer of data (Gordon, 2007).

Privacy can also be understood in a broader sense as encompassing that of companies wishing to protect their competitiveness and consumers and stages eager to preserve their sovereignty and citizens. In both these interpretations, privacy is an overarching concern that has a wide range of implications for anyone wishing to explore the use of big data for development in terms of data acquisition, storage, retention, use and presentation. Cloud based solution as class of business software application which has emerged whereby company data is managed and stored in data centers around the globe have common issues of safekeeping and management of confidential company data. These solutions often offer companies tremendous flexibility and cost savings opportunities compared to more traditional on premise solutions but it raises a new dimension related to data security and the overall management of an enterprise's Big Data paradigm (Schönberger and Cukier, 2013).

Another challenge, indirectly related with the previous, is the access and sharing of information. Organizational inflexibility may prevent easy sharing and cross-referencing of data across departments and functions (Eyler, 2009). It is common to expect reluctance of private companies and other institutions to share data about their clients and users, as well as about their own operations. Obstacles may include legal or reputational considerations, a need to protect their competitiveness, a culture of secrecy, and more broadly, the absence of the right incentive and information structures. There are also institutional and technical challenges, when data is stored in places and ways that make it difficult to be accessed and transferred. Another very important direction is to rethink security for information sharing in big data use cases. Many online services today require us to share private information (i.e., facebook, linkedin, etc), but beyond record-level access control we do not understand what it

means to share data, how the shared data can be linked, and how to give users fine-grained control over this sharing (Russel, 2011).

The characteristics of big data structures are also a crucial point that cans constraint the performance of the system. Managing large and rapidly increasing volumes of data has been a challenging issue for many decades. In the past, this challenge was mitigated by processors getting faster, which provide us with the resources needed to cope with increasing volumes of data. But there is a fundamental shift underway now considering that data volume is scaling faster than computer resources (Zikopoulos et al, 2011). Considering the size issue, we also know that the larger the data set to be processed, the longer it will take to analyze. The design of a system that effectively deals with size is likely also to result in a system that can process a given size of data set faster. However, it is not just this speed that is usually meant when we refer to speed in the context of big data. Rather, there is an acquisition rate challenge in the extraction, transformation and loading process (Golfarelli & Rizzi, 2009). Typically, given a large data set, it is often necessary to find elements in it that meet a specific criterion which likely occurs repeatedly. Scanning the entire data set to find suitable elements is obviously impractical. Rather, index structures are created in advance to permit finding qualifying elements quickly (Manyika et al., 2011).

Volume is the most obvious challenge which range from capture, cleansing, storing, searching, sharing and analyzing. Techniques that worked with smaller data volumes typically can't scale up to big data, which requires new architectures such as massively parallel processing (MPP), Hadoop, new database approaches, and new tools that are not simply extensions or repurposing of existing data warehousing or business intelligence architectures (Sammer, 2012). Speed is important, especially as organizations are often looking to make real-time decisions based on analysis of current data. To process large quantities of data within tolerable elapsed times, Big Data requires state-of-the art technologies, including MPP databases, data mining grids, distributed file systems, distributed databases; cloud computing platforms, the Internet and scalable storage systems (Inmon, 1997). Flexibility matters because different deployment platforms offer advantages in dealing with different types of data and tasks. As most data centers can only support a limited number of platforms,

organizations typically commit to one platform or another, with the risk of choosing one that might not serve them well as requirements or conditions change.

Understanding and utilizing Big Data is a daunting task in most industries and companies that deal with big data just to understand the data that is available to be used, determining the best use of that data based on the companies' industry, strategy, and tactics. Also, these types of analyses need to be performed on an ongoing basis as the data landscape changes at an ever-increasing rate, and as executives develop more and more of an appetite for analytics (Krishnan, 2013). Since much of the technology that is required in order to utilize big data is new to most organizations, it will be necessary for these organizations to learn about these new technologies at an ever-accelerating pace. It is estimated that there is a need for approximately 140,000 to 190,000 more workers with "deep analytical" expertise and 1.5 million more data-literate managers, either retrained or hired. Therefore, it is likely that any firm that undertakes a big data initiative will need to either retrain existing people, or engage new people in order for their initiative to be successful. The associated cost of taking on big data projects is also an issue for most organization and this can also be attributed to lack of support from top management in order to get the funding for these projects (Minelli et al., 2012).

Finally, working with new data sources brings a significant number of analytical challenges. The relevance and harshness of those challenges will vary depending on the type of analysis being conducted, and on the type of decisions that the data might eventually inform. The big core challenge is to analyze what the data is really telling us in a fully transparent manner. The challenges are intertwined and difficult to consider in isolation, but according to King and Powell (2008), they can be split into: getting the picture right (i.e., summarizing the data); interpreting or making sense of the data through inferences; and defining and detecting anomalies.

2.5 Effects of Big Data Management on Business Value.

Big data has improved consumer dialogue within the banking sector. Banks were in firm control of customer relationship but not anymore. Consumers now have transient relationships with multiple banks and these banks no longer have a complete view of

their customer's preferences, buying patterns and behaviors. Traditional relational database technologies are not able to store, process and handle huge amounts of data. This problem is aggravated by the fact that social networks now capture very valuable psychographic information like consumer's interests, activities and opinions. Gaining a fuller understanding of a customer's preferences and interests are prerequisites for ensuring that banks can address customer satisfaction and for building more extensive and complete propensity models. This now forces the banks to bring in external sources of information, information that is often unstructured. Bringing together transactional data in customer relationship management systems and payments systems, and unstructured data both from within and outside the firm requires new technologies for data integration and business intelligence to augment the traditional data warehousing and analytics approach. Big data management therefore plays a pivotal role in enabling customer centricity in this new reality (Oracle, 2012).

Big Data tools allow bank clerk to check customers profile in real-time and learn which relevant products or services s/he might advise. Big Data also have a key role in uniting the digital and physical shopping spheres: a retailer could suggest an offer on a mobile carrier, on the basis of a consumer indicating a certain need in the social media (Lacy, Diamond and Ferrara, 2012). Big data has helped to understand how others perceive a firms products so that they can adapt them, or for marketing, if need be. Analysis of unstructured social media text allows banks to uncover the sentiments of their customers and even segment those in different geographical locations or among different demographic groups (Versace and Massey, 2012). On top of that, big data lets banks test thousands of different variations of computer-aided designs in the blink of an eye so that you can check how minor changes in, for instance, material affect costs, lead times and performance. This can then raise the efficiency of the production process accordingly (Freedman, 2005).

There is increased performance of risk analysis. Success not only depends on how people run their companies. Social and economic factors are crucial for the company's accomplishments as well. Predictive analytics, fueled by Big Data allows one to scan and analyze newspaper reports or social media feeds so that they permanently keep up to speed on the latest developments in their industry and its environment. Detailed health-tests on ones suppliers and customers is another goodie

that comes with Big Data (Westphal, 2008). Banks can map the entire data landscape across their company with Big Data tools, thus allowing them to analyze the threats they face internally. They are able to detect potentially sensitive information that is not protected in an appropriate manner and make sure it is stored according to regulatory requirements. With real-time Big Data analytics they can, for example, flag up any situation where 16 digit numbers – potentially credit card data - are stored or emailed out and investigate accordingly. Banks are able to market using social media information, both content and relationship, to move from sampling to full dataset analysis, from demographic segments to markets-of-one, and from longer-term trending of historical data to near real-time reaction to emerging events. Prediction of customer behaviors and outcomes of proposed actions allow new business models to be created and tested, ultimately driving increased revenue is now happening because of big data. Big data enables banks to develop business model or advance the ones they have because of direct or indirect interaction with large consumer markets (Mohanty et al., 2013).

Cost containment in real-time becomes viable as electronic event monitoring from automobiles to smartphones, fraud detection in financial transaction data and more expands to include larger volumes of often smaller size or value messages on ever-shorter timescales (Zikopoulos et al., 2011). Big data analysis techniques on streaming data, before or without storing it on disk, have become the norm, enabling faster reaction to specific problems before they escalate into major situations. The spate of recent regulations is unprecedented for any industry. Point-in-time liquidity positions currently provided by static analysis of relevant financial ratios are no longer sufficient, and a more near real-time view is being required. Efficient allocation of capital is now a major competitive advantage, and risk-adjusted performance calculations require new points of integration between risk and finance subject areas (Oracle, 2012). Banks who deal in capital markets need to capture time-series data and merge it with real-time event processing systems to get a more accurate view of risk exposures across asset classes, lines of business and firms in order to better predict and manage systemic interplays. Many firms have also moved to a real-time monitoring of counterparty exposure, limits and other risk controls. From the front office all the way to the boardroom, everyone is keen on getting

holistic views of exposures and positions and of risk-adjusted performance (Jacobsen et al, 2012).

As devices that consumers can use to initiate core transactions proliferate, so too do the number of transactions they make. Not only is the transaction volume increasing, the data points stored for each transaction are also expanding. In order to combat fraud and to detect security breaches, weblog data from bank's internet channels, geospatial data from smart phone applications, etc., have to be stored and analyzed along with core operations data. Up until the recent past, fraud analysis was usually performed over a small sample of transactions, but increasingly banks are analyzing entire transaction history data sets. Similarly, the number of data points for loan portfolio evaluation is also increasing in order to accommodate better predictive modeling (Gadd ,Karstedt and Messner, 2011).

2.6 Value Creation Theories

In their research, Amit & Zott (2001) introduced a model illustrating how value is created within e-business organizations. Besides examining dozens of organizations, they used existing theories on value creation to define their model. In order to define big data analytics in terms of value creation, the same theories are used to describe how value can be created when an organization is successfully performing its value activities. Next, these theories will be discussed.

2.6.1.1 Value Chain Analysis

Value chain analysis identifies the activities of the organization and then studies its economic implications (Amit & Zott, 2001). Value is defined as the amount that buyers are willing to pay for a product or service. The organization is profitable if the value it commands exceeds the cost of performing these value activities. Hence, an organization becomes more valuable when either the value activities can be performed at lower costs or results in more sales. These linkages not only connect value activities within an organization but also connect value chains between

organizations. This means that the value of an organization can also be influenced by its strategic partners, for example a supplier. An organization‘s value chain is a system of interdependent activities connected by linkages Value chain analysis explores both primary activities (activities involved in the physical creation of the product, its marketing and delivery to buyers and its support and servicing after sale, hence activities having a direct impact on value creation) and support activities (Porter & Millar, 1985).

2.6.1.2 Schumpeterian’s Innovation Theory

In Schumpeterian’s innovation theory, innovation is the source of value creation, emphasizing the importance of technology and research & development (R&D). According to this theory, value can be created by introducing new goods, new production methods, new markets, new supply sources and the reorganization of industries through technical innovation (Amit & Zott, 2001) resulting in economic development. Existing products, services and technologies can be combined resulting in new products or services, which in turn can be the start of the creation of another product or service. This vicious circle is the foundation of the rise of complete new markets and hence, creative destruction (also called “Schumpeter’s gale”) referring to the creation of entire new markets destroying old markets. An important aspect of Schumpeterian’s innovation theory is the way how existing products, services and technologies can be combined resulting in a new product or service. Schumpeterian’s innovation theory originates from the 1930’s, an industrial period in which “combining existing products and services” referred to the combination of tangible and physical goods (Moran, 1997).

Innovation, more broadly, can occur in a wide variety of areas, including the underlying business processes, marketing image and strategic positioning of the firm, all of which may be impacted by how the underlying resource of a firm are combined and configured. New products or services are created by combining and reconfiguring resources. These resources can be categorized in two groups: input resources and knowledge-based resources. Input resources mostly refer to tangible recourses like people, equipment, property rights and capital while the latter refers to an intangible ingredient needed to combine the resources productively and in a value creating way:

knowledge. Knowledge-based resources operate on the input resources by providing both specialized understanding of the separate inputs as well as coordinative understanding of how the various input resources fit together to provide value to the firm (Galunic & Rodan, 1998).

2.6.1.3 Resource-Based View

The resource-based view is closely related to Schumpeter's innovation theory, especially Galunic and Rodan's extension described in the previous section. In the resource-based view, resources are defined as anything which could be thought of as strength or weakness of a given firm (Wernerfelt, 1984) and according to this theory, an organization is a bundle of resources and capabilities (Amit & Zott, 2001). An organization is creating value when marshaling and uniquely combining a set of these resources and capabilities in order to reduce the organizational costs or increase its revenue compared to what would have been the case if the firm did not possess those resources (Amit & Zott, 2001).

2.6.1.4 Strategic Networks

Strategic networks are long-term, purposeful arrangements among distinct but related for-profit organizations that allow those firms in them to gain or sustain competitive advantage (Jarillo, 1988) and can exist in forms of strategic alliances, joint ventures, long-term buyer-supplier partnerships and other ties (Amit & Zott, 2001). The relationships between organizations within a strategic network are essential to their competitive position. With strategic networks, organizations often outsource existing functions to other organizations. Strategic networks theory is arguing that networking can create value by enabling access to resources (information, markets, technologies, activities, etc.). Also strategic networks offer risk sharing, the opportunity to scale and scope more easily, knowledge sharing and learning (Amit & Zott, 2001).

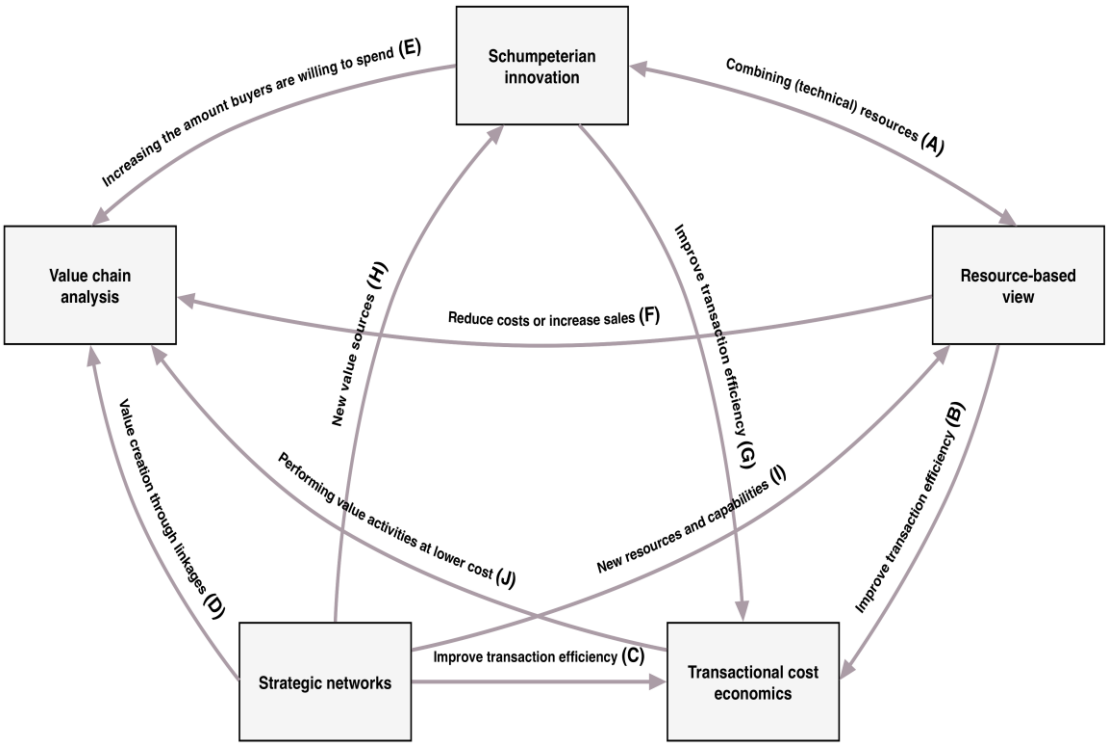
2.6.1.5 Transaction Cost Economics

In transaction cost economics theory, value creation occurs when increasing the efficiency of a transaction as enhanced efficiency reduces costs. Hence, this theory is more focused on cost and effort reduction. Value can be derived from the attenuation of uncertainty, complexity, information asymmetry, and small-numbers bargaining conditions (Amit & Zott, 2001). Also reputation, trust and transactional experience can increase transaction efficiency. Important is the broad definition of a transaction: it is defined as a transfer of a good or service across a technologically separable interface (Amit & Zott, 2001).

2.6.2 Linking Value Creation Theories

So far the different theories on value creation are introduced. In this section connections between these theories are exposed leading to a comprehensive view of how a value activity might create value. Figure 1.0 illustrates the connections between the theories discussed. Next, these connections will be discussed in more detail including examples.

Figure 1.1: Value Creation Theories Connected.



According to the Schumpeterian's innovation theory, innovation is the ultimate source of value creation. This innovation comes from the combination of existing physical goods or knowledge-based resources (Galunic and Rodan's extension). Hence, the connection (A) between Schumpeterian's innovation theory and the resource-based view becomes clear: both argue that combining resources is a major value creation driver. This might have another goal than just creating new products and services: it also can improve primary or secondary activities leading to transaction efficiency improvements resulting in a decrease of costs and hence, indicates a connection with the transaction cost economics theory (B & G). This also indicates a connection between the transaction cost economics theory and the value chain analysis theory since, according to the latter, value creation occurs when the value activity can be performed at lower costs. An example would be an organization creating an improved version of its factory machines leading to energy savings and an increase in capacity per hour.

This organization might also invest in a long-term buyer-supplier partnership by exclusively renting these machines from another organization leading to the same efficiency improvements. This indicates a connection (C) between the strategics network theory and the transaction cost economics theory. Strategic networks also influence the set of resources for a particular organization, since a strategic partnership with another organization opens a new range of resources that can be used to create unique combinations. This indicates a connection between the strategic network theory, Schumpeterian's innovation theory (e.g. by creating a new supply source) (H) and resource-based view theory (e.g. access to specific knowledge) (I). Since, according to the value chain analysis theory, an organization's value chain is also depending on inter-organizational connections called linkages, a clear connection between this theory and strategic networks theory exist (D). Schumpeterian's innovation theory might influence the value chain of the organization (E) by improving organizational efficiency through the combination of (new) resources.

Also, by innovating, organizations have the ability to sustain or increase the amount that buyers are willing to spend for a (new) product or service, leading to added value. Such an innovation is often temporary as competitors will try to catch up. By

protecting innovations, a (temporary) monopoly emerges, restraining competitors to do so. The resource-based view theory also connects with the value chain analysis (F) as combining resources might reduce organizational costs (e.g. due to innovation or strategic networks) or increase sales (e.g. due to new products or a new marketing strategy) By connecting the earlier described value creation theories it becomes clear that there is a significant overlap between these theories, also indicated by Amit & Zott (2001).

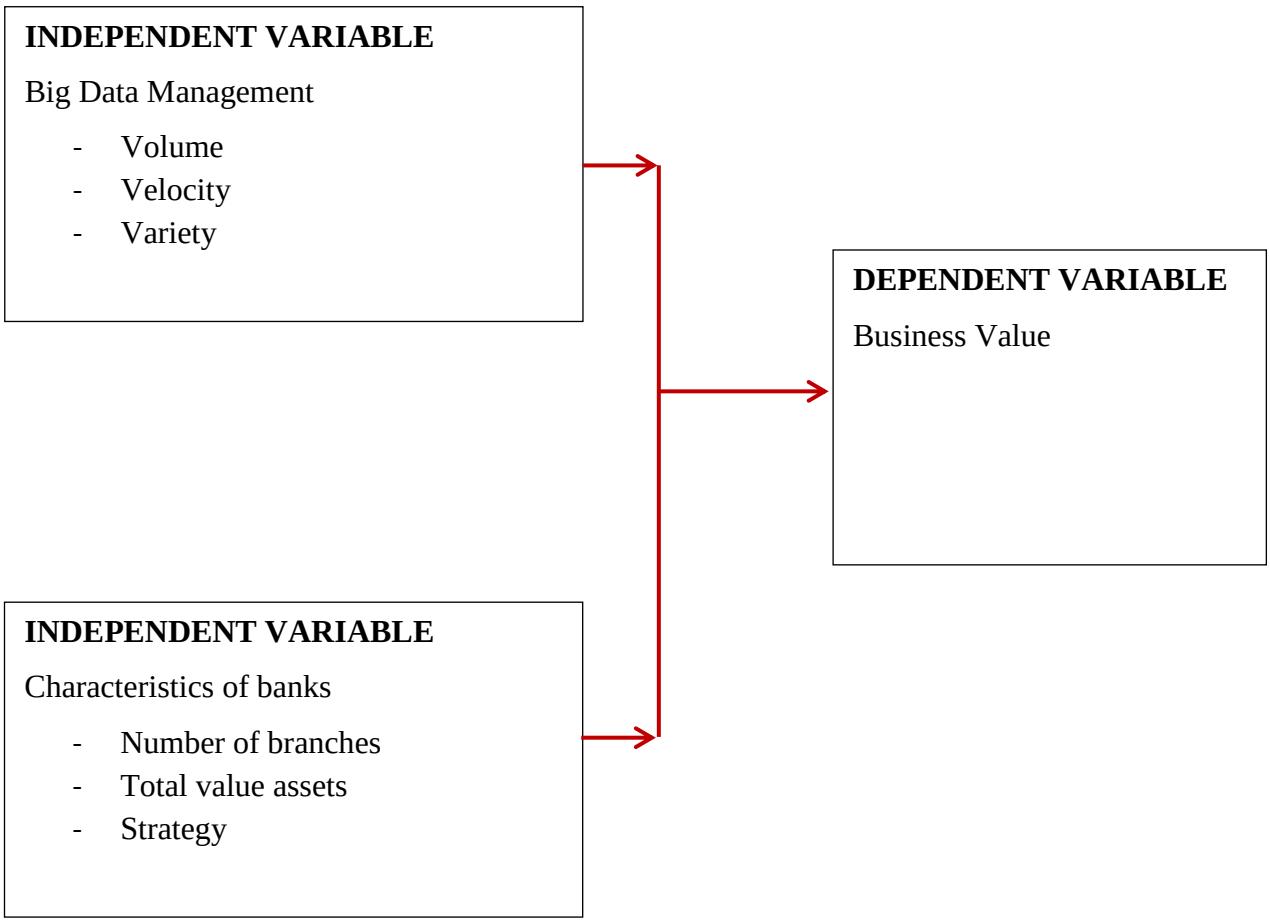
To summarize and answer the second sub-research question, value creation within organizations is caused and affected when an activity acts as a value creation driver. This happens when it Creates, supports the creation of, or improves a product or service, of which the amount buyers are willing to pay is higher than the cost to create and leverage it (value chain analysis), or; Creates, supports the creation or improves a product or service (possibly leading to entire new markets) by combining xisting resources (Schumpeterian's innovation theory and Galunic and Rodan's extension), or; Combines resources in an unique way reducing organizational costs or increasing its revenue (transactional cost economics theory), or; is a result of a strategic partnership leading to competitive advantage (strategic networks theory), or; reduces costs by improving transaction efficiency (transaction cost economics theory).

Remarkable but quite obvious is the fact that most theories directly link value with profit, either by decreasing costs or increasing sales. Only Schumpeterian's innovation theory is not directly referring to this. It indeed argues that technological innovation is the source of economic development but not necessarily for that particular activity. Many new products and services failed to be profitable but yet, value still might have been created; perhaps in forms of knowledge-based resources or technical developments. These resources can be used for the creation of a new product or service which eventually can be profitable, hence leading to economic development (Moran, 1997).

2.7 Conceptual Framework

The paradigm illustrates the conceptual framework of the study and how big data management has influence on the business value in commercial banks.

Figure 2.1: Conceptual Framework



Source: (Researcher, 2013)

Big Data can be best described using the three V's: volume, variety and velocity. These are the three defining properties or dimensions of big data. Volume refers to the amount of data, variety refers to the number of types of data and velocity refers to the speed of data processing. According to the three V's model, the challenges of big data management result from the expansion of all three properties, rather than just the volume alone. Organizations employ big data management strategies to help them contend with fast-growing pools of data, typically involving many terabytes or even petabytes of information saved in a variety of file formats.

Big data can generate value to organizations that fully harness its potential. The various banks characteristics: number of branches, total value assets and strategy are key in order for an organization to get full benefits of big data. The more the number of branches a bank own, the more data the organization generates. These huge volumes of data need to be analyzed for the organizations to know how they are going to strategize in terms of marketing and product innovation. Big data management initiatives are costly and the more value assets a company has the easier it is to invest in big data tools and technologies.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter provides a description of the research methodology that was followed in completing the study. The chapter has described the research design; target population, the data collection methods and data analysis techniques.

3.2 Research Design

This section describes research design and methodology in terms of population, sampling and administration of research instruments, data collection procedures, and the techniques used in data analysis all of which will be utilized in order to answer the research questions and achieve the set objectives (Mugenda and Mugenda, 1999). The design of this study was based on a descriptive survey research in which data was collected for the objectives of the study. The research determined the extent to which commercial banks are managing big data, established the benefits of big data management, and determined challenges in management of big data. The choice of descriptive survey research as opposed to other research designs was motivated by the fact that it provides for a suitable instrument for collecting a large amount of data and it provides a practical framework for collecting a large sample of composing groups and thirdly, survey studies have strong data reliability.

3.3 Target Population

According to Gravetter and Wallnau (2007), a population is defined as the set of all the individuals of interest in a particular study. According to the Central Bank of Kenya (2010) commercial banks licensed to carry out business in Kenya, but not under statutory management are 43 banks and 1 mortgage finance company. These licensed 43 banks and 1 mortgage finance company formed the population of the study where questionnaires were distributed to the senior managers in the Information Technology department leading to a total of 44 respondents. This is because they maintain information technology strategies, implementing, maintaining an

organization's technology infrastructure and overall organization evaluation in terms of technology.

The commercial banks are classified by Central Bank of Kenya according to their asset base. Tier I banks have total assets in excess of KES. 20 billion, Tier II banks with total assets less than KES 20 billion but more than KES 5 billion and Tier III banks with total assets less than KES 5 billion. The population distribution will be as follows:

Table 3.1: Population Distribution

Banks Categorization	Assets (in Billions)	Population
Tier 1 Banks	>20 B	9
Tier 2 Banks	>5B and <20B	16
Tier 3 Banks	<5B	20
Total		44

Source: CBK (2013)

3.4 Data Collection

Primary data was collected using structured questionnaires. According to Mugenda and Mugenda (2003), primary data is data the researcher collects while secondary data refers to data from other sources. The structured questionnaire were administered to the management in the information technology department of the banks using self-administered – a drop and pick method. The questionnaire sought to obtain information pertaining to general information of the respondent and that of the organization, extent of big data management in the organization, benefits, challenges and the effect of big data management on the business value.

3.5 Data Analysis

On receiving the questionnaires from the respondents, the data was checked to ensure completeness, consistency, accuracy, and uniformity. The data was then coded and tabulated to facilitate data analysis. Descriptive and inferential approaches were utilized in data analysis. Descriptive statistics was used in analyzing on extent of big data management in commercial banks, value of big data management, challenges in and effects of big data management on business value.

CHAPTER FOUR: DATA PRESENTATION AND ANALYSIS

4.1 Introduction.

This chapter presents a summary of the results and findings of the study in relation to the research objectives.

4.2 Response Rate.

Table 4.1 presents a summary of the study findings with respect to the response rate sent out to all 43 banks and 1 mortgage finance company.

Table 4.1: Response Rate

Response Rate	Frequency	Percent (%)
Questionnaires Issued	44	100
Questionnaires Returned	31	70.5
Questionnaires not Returned	13	29.5
Total	**44**	**100**

A total of thirty one completed questionnaires were obtained from the listed commercial banks. This represents 70.5% response rate.

4.3 Demographic Information.

An analysis of the demographic data was carried out to have a clear understanding of the socio-demographic characteristics for the sample used.

4.3.1 Gender of the Respondents.

Table 4.2 presents the results of the study with regards to the gender of the respondents.

Table 4.2: Gender of the Respondents

Gender	Frequency	Percent (%)
Female	4	13
Male	27	87
Total	**31**	**100**

From the above table, majority of the respondents are men. The percentage of the male respondents is 87%, while that of the female respondents is 13%. This indicates that most banks have male senior Information Technology managers.

4.3.2 Functional Areas in IT Department.

Table 4.3 presents major functional areas in an IT department.

Table 4.3: Functional Areas in IT Department.

Years	Frequency	Percent (%)
Database Administration	5	16.13
IT security	7	22.58
Operations and Support	11	35.48
Systems Administration	6	19.35
Network Administration	2	6.45

As you would most certainly suspect, most of the respondents reported that they performed operations and support roles and formed 35.48%. 22.58% were IT security senior managers, 19.35% were from systems administration, 16.13% from database administration and the remaining 6.45% from network administration functional unit.

4.3.3 Duration of Bank in the Market.

Table 4.4 presents a summary of data relating to the number of years the bank has been in operation.

Table 4.4: Duration of Bank in the Market.

Years	Frequency	Percent (%)
<10 years	9	29.03
11 -25 years	11	35.48
26 - 50 years	7	22.58
> 50 years	4	12.90
Total	**31**	**100**

Kenyan banks have been in operation for more than one hundred years. 29.03% of the respondent banks have been operating in the market for at least 10 years or less, 35.48% for periods between 11 and 25 years, 22.58% for periods between 26 and 50 years, and 12.90% having operated for more than 50 years.

4.3.4 Asset Value of the Bank.

Table 4.5 presents a summary of data relating to the total asset value of the banks.

Table 4.5: Asset Value of the Bank

Assets (in Billions)	Frequency	Percent (%)
>20 B	5	41.94
>5B and <20B	12	32.26
<5B	14	25.81
Total	**31**	**100**

41.94 % of the respondent banks had an asset value of more than 20 Billion, 32.26% had a total asset value ranging from 5 Billion and 20 Billion, whereas 25.81% of the respondent banks had an asset value less than 5 billion at the time this research was carried out.

4.3.5 Number of Branches in Kenya

Table 4.6 presents a summary of the number of branches owned by the banks in Kenya.

Table 4.6: Number of Branches in Kenya

No. of Branches	Frequency	Percent (%)
<10	5	16.13
11 - 50	15	48.39
51- 100	8	25.81
>100	3	9.68
Total	**31**	**100**

16.13% of the respondent banks have at least 10 or fewer branches. Most banks under study have 11 to 50 branches, 25.81% have branches from 51 to 100 and 9.68% has more than 100 branches countrywide. Big data management can help banks interpret which branches or products are performing the best. The growth of the number of branches by banks has the benefit of availing the banking services to customer's convenience, this being key in bringing the services closer to the customer, it also increases the avenues for revenue generation.

4.3.6 Big Data Management Initiative in the Organization.

Table 4.7 presents the stages of big data management initiative in the banks.

Table 4.7: Big Data Management Initiatives

Big Data Management Initiative	Frequency	Percent (%)
Yes, initiatives in operations	9	29.03
Serious Planning	7	22.58
No initiatives underway	15	48.39
Total	**31**	**100**

Some 29.03% of the respondents already have ongoing big data management initiatives or have implemented the initiative on some scale. 22.58% of correspondent banks have big data management initiatives under serious planning while the remaining 48.39% have no big data initiatives underway.

4.4 Extent to Which Banks are Using Big Data.

4.4.1 Level of Awareness of Big Data Products in the Market.

Table 4.8 shows the level of awareness of the correspondents in regard to their awareness of big data management tools and the visibility of these products in the Kenyan market.

Table 4.8: Level of Awareness of Big Data Products

Level of Awareness	Frequency	Percent (%)
1 (Minimal)	12	38.71
2 (Less than adequate)	9	29.03
3 (Adequate)	6	19.35
4 (More than adequate)	3	9.68
5 (Mastery)	1	3.23
Total	**31**	**100**

The results show that 38.71% of the correspondents felt that their awareness and visibility of big data products on the market was low, while 3.23 % thought that they were fully aware of big data management and its products on the market. 19.35 % fairly rated the awareness and visibility of big data management in the market.

4.4.2 Current Organization's Big Data Management Capabilities.

Table 4.9 shows how the current big data management capabilities in the various banks look like.

Table 4.9: Big Data Management Capabilities

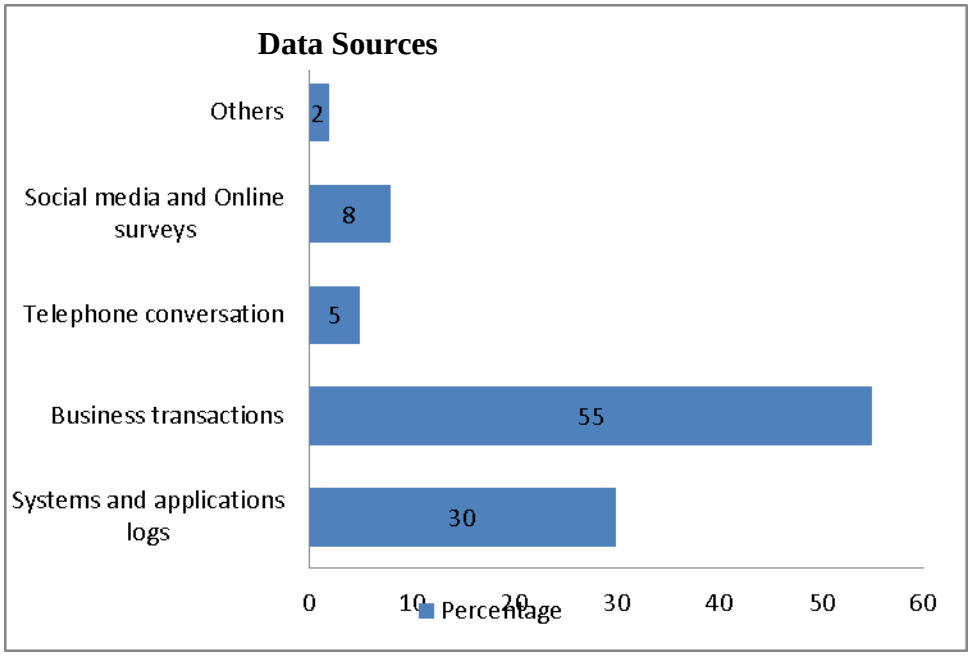
Big data management capabilities	Frequency	Percent (%)
Less than Adequate	13	41.94
Adequate	10	32.26
More than Adequate	3	9.68
World class	0	0.00
Don't know	5	16.13
Total	**31**	**100**

Only 9.68% of respondents ranked their analytic capabilities as adequate. 9.68% of respondents ranked their ability to use data and analytics to transform their business as more than more than adequate. 16.13% didn't know how to rank their capabilities because they do not have any big data initiatives in operations or have not even thought about it.

4.4.3 Big Data Management Data Sources

This section sought to find out what data sources the organizations were using to analyze big data. These also helped us in knowing whether the people really knew what kind of data really constitute to being part of big data. Figure 4.1 showed us the various ratings of the current data rankings of their big data management.

Figure 4.1: Big Data Management Data Sources

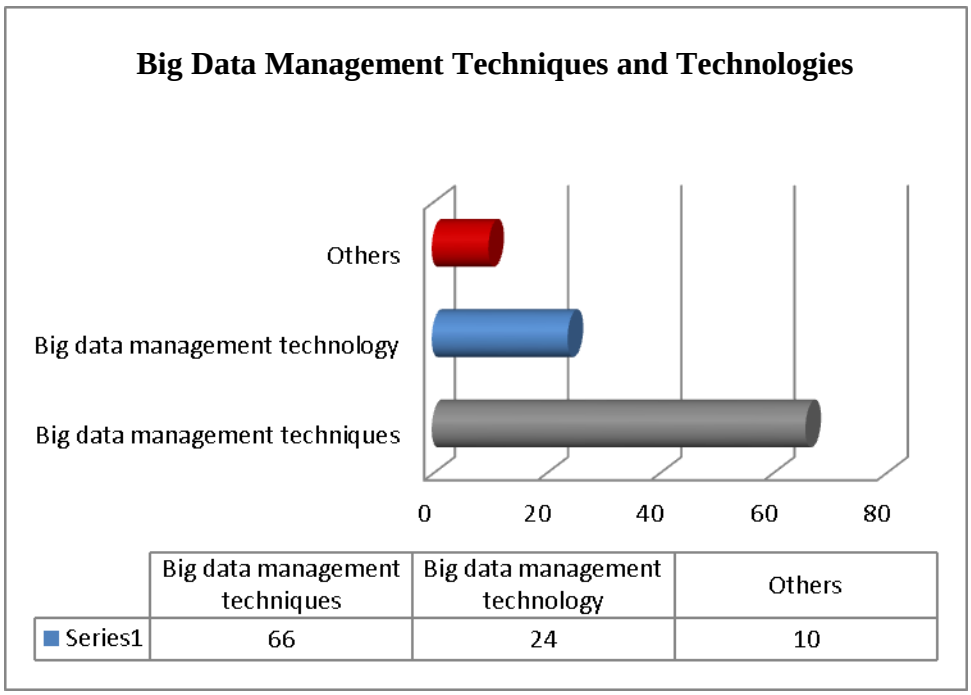


The IT managers from various functions, placed business transactions at the top of the list with 55%, traditional structured data that continues to be of critical importance to companies despite the buzz about unstructured data types. However, system logs were also ranked high at 30%, social media and online surveys was given 8%, telephone conversations were ranked at 5% and others getting a ranking of 2%.

4.4.4 Big Data Management Techniques and Technologies

Fig 4.2 gives an overview of the techniques and technologies the banks have used in managing their big data.

Figure 4.2: Big Data Management Techniques and Technologies

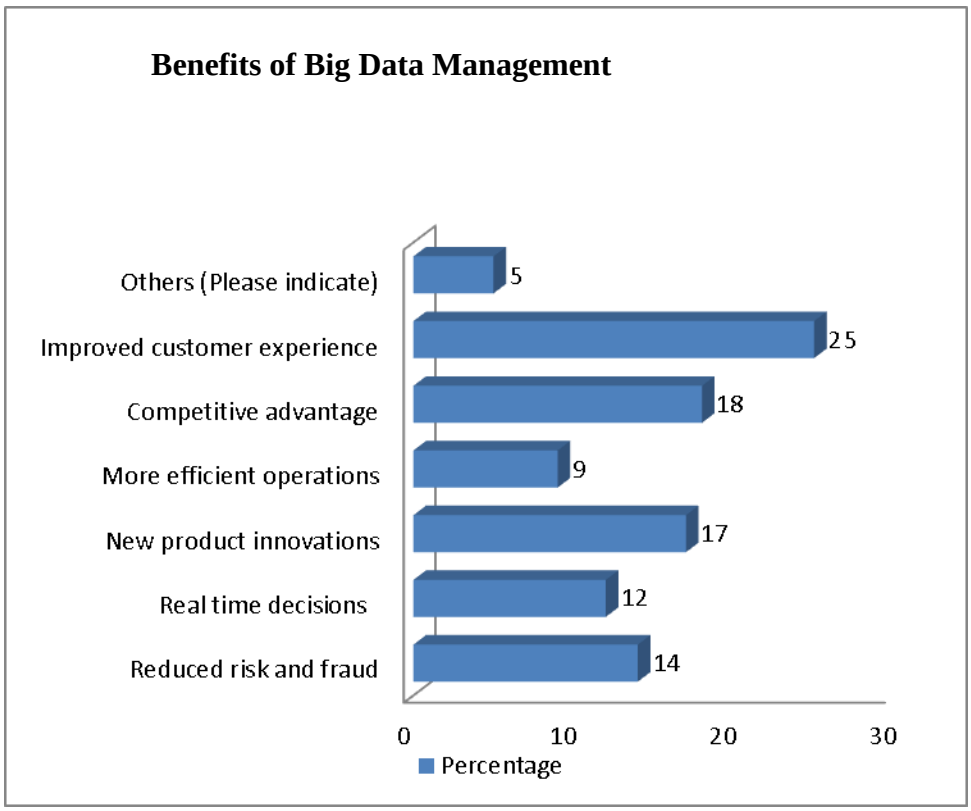


66% of the correspondents indicated that they used big data management techniques, 24% were into the big data management technologies while 10% managed their data with other tools like queries and dashboards.

4.5 Benefits Using Big Data Management

The study sought to find the various benefits that the harnessing of big data can bring to an organization. Figure 4.1 presents a summary of the study findings with regards to the various benefits attributed to big data management.

Figure 4.3: Benefits of Big Data Management



It is clear that most of the people felt that big data management has really improved their customer experience ranking it top with 25%. 5% of people felt that it has helped them achieve other benefits like increased sales, development of business models among others. 18% felt that it gave them competitive advantage, 17% felt that it helped them in new product innovations, 14% used it to reduce risk and detect fraud patterns and detect fraudsters, 12% felt that it facilitated making decisions in real-time and 9% indicated that it played a role in their IT operations.

4.6 Challenges in Management of Big Data

This section identified the major challenges involved in the management of big data. These challenges need to be addressed in order of banks to fully harness big data potential and its values. Figure 4.2 shows how people rated the challenges.

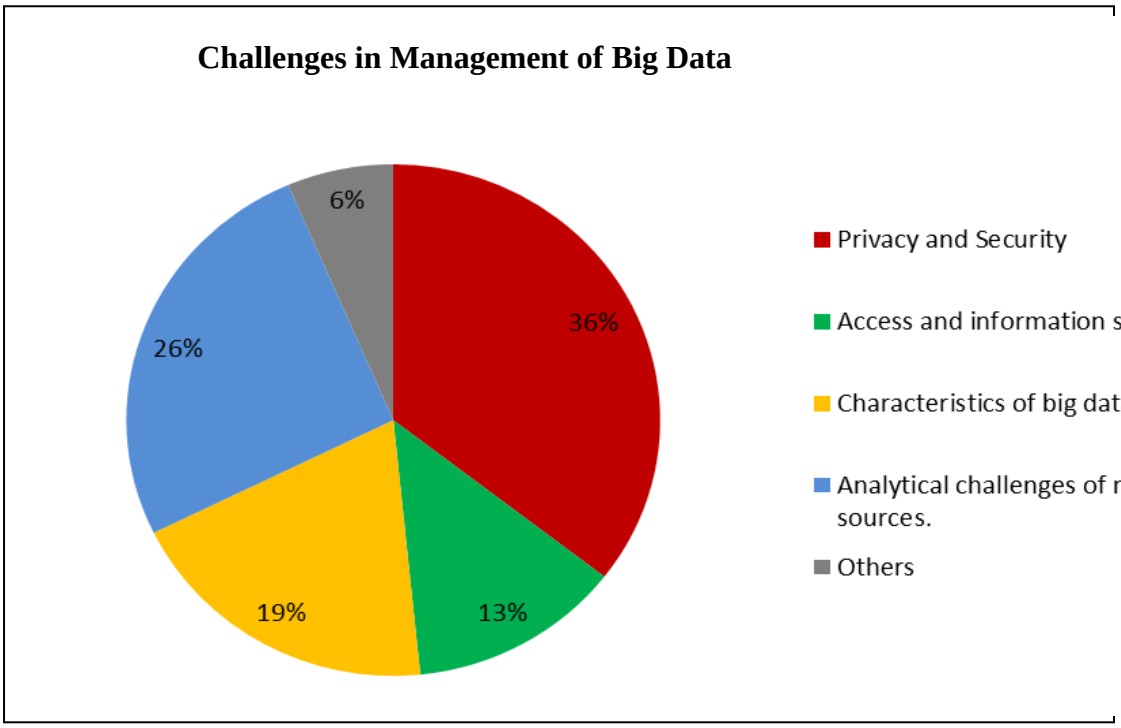


Figure 4.4: Challenges in Management of Big Data

Privacy and security issues stood out to be the biggest challenge with 36%. 13% was attributed to access and information sharing. 19% felt that the volume, size and variety, which are the characteristics of big data was also a big issue. Challenges which come with analyzing new data set got a rating of 26% and other issues like data governance, lack of skills to manage the big data projects, lack of skills to analyze the data, cost of implementing big data and lack of business support from top management also came up in the rating of the challenges.

4.7 Effects of Big Data Management on Business Value.

4.7.1 Various Effects of Big Data Management on Business Value

This section sought to identify the effects of big data on the business value.

Table 5.0: Effects of Big Data Management on Business Value

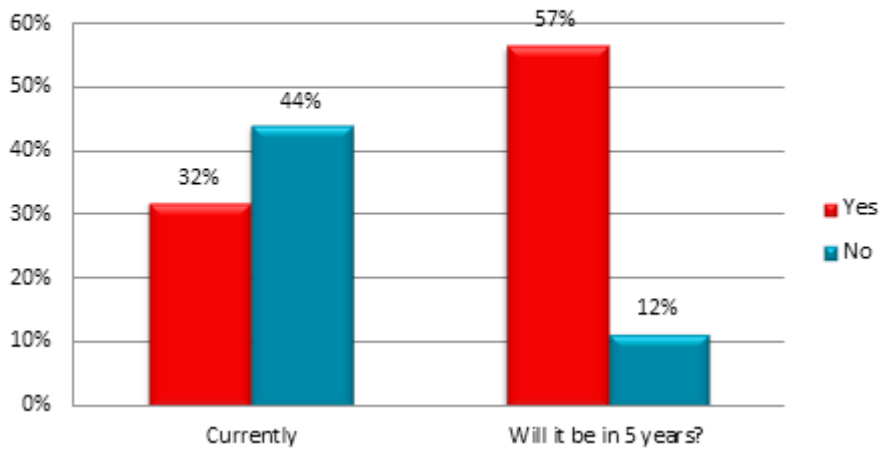
Effects	Frequency	Percent (%)
Increases ability to understand and respond to customers better	11	35.48
Identifies opportunities in the market	7	22.58
Increases ability to predict and mitigate risks	3	9.68
Improves decision making	8	25.81
Others	2	6.45
Total	**31**	**100**

35.48% indicated that big data helped them to understand and respond to their customers better, 22.58% indicated that big data helped them to identify market opportunities, 9.68% indicated that it helped them to predict and mitigate risks better, 25.81% indicated that it improved their decision making and 6.45% felt that it had other effects like reducing enterprise cost, creating speed and agility within the organization, financial management among others.

4.7.2 Big data as a Strategy

This section tried to identify whether big data was a strategic initiative in the organization.

Figure 4.5: Big data as a Strategy



When asked if big data is strategic initiative in their organization currently, or if they thought that big data would be strategic initiative in their organization within five years, 32% reported that it is currently a strategic initiative, and 57% reported that it would be in in 5 years.

CHAPTER FIVE: SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

The contents of this chapter have been divided into summary, conclusion, recommendation, limitation and future research sections.

5.2 Summary and Discussions

The study sought to evaluate the influence of big data management on business value in commercial banks in Kenya and specifically to focus on extent of big data management in commercial banks, benefits of big data management, challenges in big data management and effects of big data management on business value. The survey yielded some interesting findings, many of which are relevant to the research conducted. Survey has clearly indicated that their big data management range far beyond social media and machine-generated data to include a wide variety of traditional structured and transactional business data. Big data management has driven change in traditional data management strategies and has found a home in an expanding information ecosystem that many companies struggle to manage today.

Other types of data that the organizations were analyzing included: documents, e-mail, device data, weblogs, videos on you tube, market research, text messages, clickstreams, digital media, consumer feedback and complaints among others. Structured data remains the most common single data source. Banks with high asset value had invested heavily into their technologies thus making them also invest in big data. This is an indication that shareholders have highly invested and therefore expect high returns. Banks need to put big data management into good use due to its potential, enormous value and gain competitive advantage because after the financial crisis, they are focused on their balance sheet and have become a bit more conservative with new technologies. Banks have relied on traditional relational technologies to handle the ever-increasing data and analytics burden. It is however increasingly clear that while such technologies will continue to play an integral role, new technologies like big data management will have a more transformative role in enterprise data management. This new data management strategy leverages an array of platforms for the highest

performance possible and brings together human-sourced information, process-mediated data and machine-generated data as a complete, comprehensive business information resource. At the core of this change is a movement to align data with operational and analytic workloads, each on the best possible platform. However, the banking industry has seen significant growth through new entrants as reflected from the study. These entrants into the industry bring upon competition and this means that banks need to keep up with the trends of the business in order to stay competitive and predict the future. Banks have to develop strategies to respond to competition, to both safeguard their niches and to enlarge their market share.

The study sought to find out the extent to which commercial banks were managing big data. Banks with big data management in operation shows that they already have hands-on experience with big data techniques and tools. The banks under serious planning bracket represented those that are close to or on the verge of signing contracts for big data management. Investigating banks represented those still looking at big data management requirements and technologies and are far from implement big data solutions. The category of banks without any big data management initiatives had some banks who were thinking of allocating exploratory budgets for investigation and the growth of big data. There is another category that didn't have big data management initiatives in the pipeline but were still trying to figure out how the technology fits into their strategy. This is clear that Kenyan banks are at the early stages of deploying big data management initiatives. This can tell us that the one ones who have implemented are piloting and experimenting, the ones who are planning might be developing strategy or knowledge gathering about the initiative.

On the part of big data management capabilities of an organization, there were some exciting findings which helped to know if the organization has heavily invested on its technology or otherwise. Today's less expensive technologies have opened the door for a democratization of innovation, allowing most companies to leverage big data and its opportunities at reasonably affordable rates the earlier the organizations' who have not initiatives in place realize this these better. None of the respondents rated their big data management capabilities as world class because this is a new concept in the market and even those who have are already harnessing its potential are struggling to keep with the pace. Whether their current analytics capabilities are ranked as "Less

than adequate" or as "world class," all respondents should look at big data to have a major impact on their business. Big data management initiatives have brought with it a lot of benefits and values to the organizations that have put it into use.

The true value of big data lies not just in having it, but in being able to use it for fast, fact-based decisions that lead to real business value. The demands on data are incredibly complex and people are relying on data to run their businesses. Big data management initiatives help banks to be more proactive in their decision making and less reactive, be able to deal with critical issues of risk, fraud, compliance, gain more insights in their customers, improve on their operations, gain competitive advantage, to increase their company's efficiency by using big data to reduce costs, develop new products and business models using insight gleaned from big data. From the findings it is clear that five key business areas that present low-risk opportunities for tangible performance using big data include: fraud detection, consumer sentiment, intelligent forecasting, customer profiling and target marketing, and customer service.

When we asked IT managers about their top big data challenges, we found that most were rated similarly in terms of their significance. This indicates that respondents face more than one challenge with big data. The influx of data presents many barriers to effectively analyzing it and to the creation of business insight for most decision makers. There are challenges in management of big data and if they can be addressed can help organizations to capture the full potential of big data. The range of technology challenges and the priorities set for tackling them will differ depending on the data maturity of the institution. Data governance and security, large data volume, speed, flexibility of data deployment platforms, cloud based solutions, understanding and utilization of big data, lack of sharing capacity and integration of a wider variety of data, lack of business support from top management are most of the challenges that kept on coming up from the survey.

Big data management has the potential to significantly shape the future of the banking industry. Big data management: increases ability to understand and respond to customers better; identifies opportunities in the market; identifies fraud patterns; increases ability to predict and mitigate risks; increases sales and marketing effectiveness; improves financial management; creates speed and agility within the

organization; reduces enterprise cost and helps an organization to have better, fact-based decision-making.

5.3 Conclusions

Clearly, the banking sector is beginning to build out road maps of where big data could deliver the most value within this broader set of technology investments, but many financial services firms are cautious about making broad-based investments in these new and relatively nascent big data management initiatives. The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. It is also expected that over the next five years, there will be notable and significant progress toward big data deployments within financial services in Kenya.

Using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world. Leaders of organizations need to recognize the potential opportunity as well as the strategic threats that big data represent and should assess and then close any gap between their current IT capabilities and their data strategy and what is necessary to capture big data opportunities relevant to their enterprise. In this task, they will need to be creative and proactive in determining which pools of data they can combine to create value and how to gain access to those pools.

However, many technical and organizational challenges described in this paper must be addressed before this potential can be realized fully. Big data productivity will come as a result of giving users across the organization the power to work with diverse data sets. Achieving the vast potential of big data demands a thoughtful, holistic approach to data management, analysis and information intelligence. Across industries, organizations that get ahead of big data will create new operational efficiencies, new revenue streams, differentiated competitive advantage and entirely new business models. Business leaders should start thinking strategically about how to prepare the organizations for big data.

5.4 Recommendations

Organizations should prioritize on technology investments and use a test-and-lean mentality to determine how fast-and deep-to go when it comes to big data management initiatives. Banks should understand immediate and strategic information needs of the organizations which will help them in understanding as to why they are interested in big data and if they need to get help and what existing investments in data platforms they currently have. They should develop a roadmap and look for guidance on what technologies are the best investments based on current business strategies and existing investments.

Big data is not just about helping an organization be more successful by marketing more effectively or improving business operations. It reaches to far more socially significant issues as well. Organizations should lead in the social revolution by looking for data in new sources, going beyond traditional structured data sources. The potential for deriving new levels of value from big data management with big data analytics is there. That's why banks need to reinvent their architecture and software to satisfy the demands of big data, larger problems and more complex scenarios, and to take advantage of new technology advancements.

To get organizations on the right path when it comes to big data management initiatives, the focus should be on understanding where to start and what the business value is. Finding value from within by auditing and leveraging information that already exists in corporate data sources, understanding existing data assets can help drive more streamlined big data use cases.

5.5 Limitations

Due the sensitivity of the data that commercial banks hold, there were difficulties in getting respondents give timely and comprehensive information. In the course of the research, time was a limiting factor. In some cases it was difficult to get sufficient time to go through the questionnaires with the respondents, other times the respondents could not have focused attention and likewise due to the qualitative nature of the study, more time in the study could have been appropriate.

5.6 Future Research

By studying more cases, the results will likely be more generic, hence could improve the validity of the results. Another way to improve the level of generalization is by focusing on other markets such as health care. This is particularly valuable for organizations working in that specific market. As big data management is considered new and relatively unexplored, the same study should be done in a few years to see how banks are currently struggling with conceptualization and implementation, have succeeded (or not) in turning their big data chaos into big data opportunities.

The same design should be used in order to create comparable results. In this study, the effects of big data management on organization's value creation were studied. Value creation and competitive advantage are closely related as competitive advantage equals the difference between the value created by the company and the potential value created by its competitors. As technologies are considered doubtful regarding their added value according to the resource-based view, and data is becoming more publicly accessible, it might be argued how this difference is expressed in the big data analytics as a value creation driver topic. Furthermore, as big data management affects organization's value creation and possibly affects competitive advantage too, it also can lead to reinvented or new business models, especially to leapfrog competitors.

With big data management, due to improved technologies capable of storing and handling unstructured and multi-structured data, and the increased capacity, more data sources can be analyzed. Yet, which data sources are most valuable and how can we identify whether it is constantly improving the quality of the outcomes of the analysis? Hence, studying the value of a data source is interested as it gives organizations the ability to decide if investing in the implementation of a particular data source is worth the effort. A good starting point is Boyd & Crawford's (2011) article arguing that bigger data is not always better data when analyzing data sets classified as big data.

REFERENCES

Adhikari, S. (2012). *Time for a big data diet*. Retrieved August 12, 2012, from http://technologyspectator.com.au/emerging-tech/big-data/need-big-data-speed.

Amit, R., & Zott, C. (2001). Value creation in E-business. Strategic Management Journal, 22(6-7), 493–520. doi:10.1002/smj.187.

Babbin J., Kleiman D., Carter E, Faircloth J.,Burnett M. and Gutierrez E. (2006). *Security Log Management: Identifying Patterns in the Chaos*. Rockland, Canada: Syngress Publishing, Inc.

Bloem, J., Doorn, M. V., Duivestein, S., Manen, T. V., & Ommeren, E. V. (2012). *Creating clarity with Big Data*. Retrieved August, 2012, from www.sogeti.com/vint/bigdata.

Bomarius F. (2009). *Product-Focused Software Process Improvement.* 10Th International Conference, PROFES 2009, Oulu, Finland, June 15-17, 2009.

Borkar, V., Carey, M., & Li, C. (2012). *Inside "Big Data Management": Ogres, Onions, or Parfaits?* ACM EDBT/ICDT Joint Conference, Berlin, Germany.

Boyd, D., & Crawford, K. (2011). *Six Provocations for Big Data*. SSRN eLibrary.

Bryant, R., Katz, R., & Lazowska, E. (2008). *Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society*. http://www.cra.org/ccc/initiatives.

Burgel, R. (2012, July 14). *Interview Red Data*. (N. Mouthaan, Ed.). Almere.

CEBR (2012). *Data equity: unlocking the value of big data*. Centre for Economics and Business Research White Paper, 4, 7-26.

Central Bank of Kenya (2013). *Prudential Guidelines*. Nairobi, Kenya: Government Printers.

Central Bank of Kenya (2013). *Risk Management Guidelines*. Nairobi, Kenya: Government Printers.

Corporation Essvale. (2011). *Business Knowledge for Information Technology in Global Retail Banking.* 1 Newton Place, London: Essvale Corporation Limited.

Cotte, V. 2012. *Big Data Seats CIOs at the Boardroom Table.* SAS. Retrieved May 8, 2012, from http://blogs.sas.com/content/anz/2012/05/08/big-data-seats-cios-at-theboardroom-table/ .

Dumbill E. (2012). *Planning for Big Data*. O'Reilly Media, Inc.

Eyler R. (2009). *Money and Banking: An International Text*.

Freedman A. D. (2005). *Statistical Models: Theory and Practice.* Cambridge University Press.

Gadd D., Karstedt S. and Messner S. (2011). *The SAGE Handbook of Criminological Research Methods*. Pekin Street, Singapore: SAGE Publications Inc.

Galunic, C., & Rodan, S. (1998). *Resource recombinations in the firm: knowledge structures and the potential for Schumpeterian innovation*. Strategic Management Journal, 19(12), 1193–1201.

Gravetter F. and Wallnau L. (2007*). Statistics for the Behavioral Sciences*. USA: Vicki Knight.

Goel S. (2012). *Financial Services.* Patparganj, Delhi: Asoke K Ghosh, PHI Learning Private limited.

Golfarelli, M., & Rizzi, S. (2009). *Data warehouse design: modern principles and methodologies.* Columbus: McGraw-Hill.

Gordon, A. (2005). *Privacy and ubiquitous network societies*. Workshop on ITU Ubiquitous Network Societies, 6-15.

Gordon K. (2007). *Principles of Data Management: Facilitating Information Sharing*. North Star Avenue, Swindon: British Informatics Society Limited.

Government of Kenya. 2008. *Banking Act 488*. Nairobi, Kenya: Government Printers.

Government of Kenya. 2008. *Finance Act 2008*. Nairobi, Kenya: Government Printers.

Hurwitz J, Nugent A, Halper F and Kaufman M. (2013). *Big Data for Dummies*.

IDC. (2009). *As the economy contracts, the digital universe expands*. IDC White Paper, 5, 12-18.

Inmon H, W. (1997). *Building the Data warehouse*. Indianapolis, Indiana: Wiley Publishing, Inc.

International Monetary Fund (2012). *Basel Core Principles for Effective Banking Supervision: Detailed Assessment of Compliance.* Washington, D.C: Andrews McMeel Publishing.

Jacobsen R and Misner S. and Hitachi Consulting. (2006). *Microsoft SQL Server 2005 Analysis Services Step by Step*. Microsoft Press.

Jarillo, J. C. (1988). *On strategic networks*. Strategic Management Journal, 9(1), 31–41.

King, G., & Powell, E. (2008). *How not to lie without statistics.* Working paper: Harvard university. Retrieved August 12, 2012, from http://gking.harvard.edu/gking/files/nolie.pdf.

Krishnan K. (2013). *Data Warehousing in the Age of Big Data*.

Lacy K., Diamond S. and Ferrara J. (2012). *Social CRM for Dummies*. Hoboken, New Jersey: John Wiley and Sons Inc.

Lehdonvirta, V., & Ernkvist, M. (2011). *Converting the virtual economy into development potential: knowledge map of the virtual economy*. InfoDev/World Bank White Paper, 1, 5-17.

Lemieux V. (2012*). Financial Analysis and Risk Management: Data Governance, Analytics and Life Cycle Management*. Berlin, Germany, Springer Berlin – Heidelberg.

Liu B. (2010). *Sentiment analysis and subjectivity. Handbook of Natural Language Processing*, Second Edition, (editors: N. Indurkhya and F. J. Damerau).

Loshin D. (2010). *Master Data Management*. Morgan Kaufmann Publishers.

Macmillan Publishers (Feb 1, 2006). *Anti-Money Laundering and Know Your Customer*: *Know Your Banking III India Limited.*

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity [Kindle edition]*. McKinsey Global Institute. Retrieved June 11, 2012 from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Minelli M., Chambers M., and Dhiraj A. (2012). *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*.

Mohanty S., Jagadeesh M. and Srivatsa H. (2013). *Big Data Imperatives: Enterprise Big Data Warehouse, BI Implementations and Analytics*.

Morley D and Parker C. (12th ed) (2009). *Understanding computers: today and tomorrow*. Boston, USA: Course Technology Cengage Learning.

Moran, P. (1997). Value creation by firms.

Mouthaan N. (2012). *Business Information Systems*, University of Amsterdam.

Mruthyunjayappa S., (2011*) Building a business case for the Kenyan Banking Sector*. Retrieved Feb 18, 2011 from http://aitecafrica.com/news/view/137.

Mugenda O. and Mugenda A. (1999). *Research Methods: Quantitative and Qualitative Approaches*. Nairobi: African Centre for Technology Studies.

Mugenda O. and Mugenda A. (2003). *Research Methods: Quantitative and Qualitative Approaches*. Nairobi: African Centre for Technology Studies.

Ohlhorst, J. F. (2012). *Big Data Analytics: Turning Big Data into Big Money*. Cary, North Carolina: SAS Institute Inc.

Oracle Corporation (Jun 2012). *White Paper on Financial Services Data Management: Big Data Technology in Financial Services*. Retrieved from www.oracle.com.

Pagani M. (2005). *Mobile and Wireless Systems beyond 3G: Managing New Business Opportunities.*

Porter, M. E., & Millar, V. E. (1985). *How information gives you competitive advantage*. Harvard Business Review, 63(4), 149–160.

Rahman H. and Ramos I. (2013). *Ethical Data Mining Applications for Socio-Economic Development.* Idea Group Inc (IGI).

Ramachandran S and Malladi S. (2013). *Leveraging unstructured data for improved decision making: A retail banking perspective*. Bangalore, India: Infosys Limited.

Russell A M. (2011). *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites*. Sebastopol, Carolina: O'Reilly Media, Inc.

Russom, P. (2011). Big Data Analytics. TDWI Research, 4th Quarter, 2011.

Sammer E. (2012). *Hadoop Operations*. O'Reilly Media, Inc.

Schönberger M V. and Cukier K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work and Think.* Houghton Mifflin Harcourt.

Slack, E. (2012). *What is big data?* Retrieved August 5, 2012, from http://www.storage-switzerland.com/Articles/Entries/2012/8/3_What_is_Big_Data.html.

Sundbo J. (1998). *The Theory of Innovation: Entrepreneurs, Technology and Strategy*. Massachusetts, USA: Edward Elgar Publishing, Inc.

Swoyer, S. (2012, July 24). *Big Data -- Why the 3Vs Just Don't Make Sense*. Retrieved August 1, 2012, from http://tdwi.org/Articles/2012/07/24/Big-Data-4th-V.aspx?Page=1.

Tarantino A. (May 16, 2008*). Governance, Risk, and Compliance Handbook: Technology, Finance, Environmental, and International Guidance and Best Practices.* River Street, Hoboken, New Jersey: John Wiley and Sons Inc.

Versace M. and Massey K. (Sep 2012). *The Case for Big Data in the Financial Services Industry*.

Westphal C. (2008*). Data Mining for Intelligence, Fraud and Criminal Detection: Advanced Analytics and Information Sharing Technologies*. Park Avenue South, New York: Taylor and Francis Group.

Wernerfelt, B. (1984). *A resource-based view of the firm*. Strategic Management Journal, 5(2), 171–180.

Zikopoulos P., Eaton C., Deutsch T., Deroos D., Lapis G. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw Hill Professional.

APPENDIXIES

APPENDIX ONE

RE: LETTER OF INTRODUCTION

I am Winnie Mbaluka, undertaking Masters of Business Administration course at the University of Nairobi.

I am undertaking survey on big data management and business value in the commercial banking sector in Kenya, leading to the production of my master's project report.

I would be obliged if you volunteered to assist in this project, by consenting to completing a questionnaire which covers certain aspects of the topic.

Any information provided will be treated in confidence and none of the participants will be individually identifiable in the resulting report.

Thanks in advance for supporting my efforts in achieving the course expectations

Best Regards,

Winnie Mbaluka

APPENDIX TWO

RESEARCH QUESTIONNAIRE

This is part of an educational study that is being carried out on the adoption of big data technology and its business value in Kenyan commercial banks. Your bank has been selected for this study. The information obtained will be treated as confidential and will be used for the study and nothing else. Your cooperation and assistance will be highly appreciated. Thank you in advance for taking time to fill in the questionnaire

Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using traditional database management tools or traditional data processing applications.

SECTION A: ABOUT YOUR ORGANIZATION

1. Please indicate your gender. _____

2. What functional area of IT do you work in?

- Database Administration
- IT security
- Operations and Support
- Systems Administration
- Network Administration

3. How long has your organization been in business? _____(Years)

4. What is the size of the bank in terms of total assets value in Kenya Shillings?

- Above 20 billion
- Between 5 and 20 billion
- Less than 5 billion

5. How many branches does the bank have in Kenya? _____

6. Is there any big data management initiative in your organization?

- Yes, initiatives in operations
- Serious Planning

 No initiatives underway

SECTION B: THE EXTENT TO WHICH YOUR ORGANIZATION IS MANAGING BIG DATA.

7. On a scale of 1-5 how would you describe your awareness of and the visibility you have on big data products currently on the market (Write a number)?
   _____

8. How would you rate your current organization's big data management capabilities?

	Less than Adequate	Adequate	More than Adequate	World class	Don't know
Current big data management capabilities					

9. To what extent are the below data sources utilized in big management in your organization?

	No Extent at All	Little Extent	Moderate Extent	Great Extent	Very Great Extent
Systems and applications logs					
Business transactions					
Telephone conversation					
Social media and Online surveys					
Others (Please indicate)					

10. Please indicate the big data management techniques and technologies being used in your organization

a) Big data management techniques

Visualization		Cluster analysis		Spatial analysis			
Text analytics		Crowd sourcing		Statistics			
In-memory analytics		Data mining		Supervised learning			
Predictive analysis		Data fusion and integration		Simulation			
Graph analytics		Ensemble learning		Unsupervised learning			
Mobile business analytics		Genetic algorithms		Time series analysis			
Video analytics		Machine learning		Social network analysis			

Audio analytics		Natural language processing		SaaS-based business analytics	
A/B testing		Network analysis		Signal processing	
Association rule learning		Pattern recognition		Sentiment analysis	
Classification		Predictive modeling		Complex event processing	
Stream processing		Regression			

b) Big data management technology

Business intelligence		Hive Hadoop		Massively parallel-processing	
Hadoop		MongoDB		Nosql	
HDFS		Pig Apache		Cassandra	
Mapreduce		Riak		Data federation technology	
Mashup		CouchDB		Phase change memory	
Non-relational database		SCOPE		Quantum Computing	

SECTION C: BENEFITS OF USING BIG DATA MANAGEMENT.

11. To what extend has big data management helped your organization achieve the benefits below.

	No Extent at All	Little Extent	Moderate Extent	Great Extent	Very Great Extent
Reduced risk and fraud					
Real time decisions					
New product innovations					
More efficient operations					
Competitive advantage					
Improved customer experience					
Others (Please indicate)					

SECTION D: CHALLENGES IN MANAGEMENT OF BIG DATA.

12. Please indicate the extent to which the below challenges inhibit the management of big data in your organization.

	No Extent at All	Little Extent	Moderate Extent	Great Extent	Very Great Extent
Privacy and Security					
Access and information sharing					
Characteristics of big data (volume, variety, velocity)					
Analytical challenges of new data sources					
Others (Please indicate)					

SECTION E: EFFECTS OF BIG DATA MANAGEMENT ON BUSINESS VALUE

13. In a scale of 1-5, please rank how the below results of big data management has affected the organizations business value.

	1	2	3	4	5
Increases ability to understand and respond to customers better					
Identifies opportunities in the market					
Increases ability to predict and mitigate risks					
Improves decision making					
Others (Please indicate)					

14. Do you agree that big data should be a key strategic priority for IT in your organization now and in 5 years?

- Yes
- No

APPENDIX THREE

List of Commercial Banks in Kenya

1. African Banking Corporation Ltd.
2. Bank of Africa Kenya Ltd.
3. Bank of Baroda (K) Ltd.
4. Bank of India.
5. Barclays Bank of Kenya Ltd.
6. CFC Stanbic Bank Ltd.
7. Charterhouse Bank Ltd
8. Chase Bank (K) Ltd.
9. Citibank N.A Kenya.
10. Commercial Bank of Africa Ltd.
11. Consolidated Bank of Kenya Ltd.
12. Co-operative Bank of Kenya Ltd.
13. Credit Bank Ltd.
14. Development Bank of Kenya Ltd.
15. Diamond Trust Bank (K) Ltd.
16. Dubai Bank Kenya Ltd.
17. Eco bank Kenya Ltd
18. Equatorial Commercial Bank Ltd.
19. Equity Bank Ltd.
20. Family Bank Ltd
21. Fidelity Commercial Bank Ltd.
22. Fina Bank Ltd.
23. First community Bank Limited.
24. Giro Commercial Bank Ltd.
25. Guardian Bank Ltd.
26. Gulf African Bank Limited.
27. Habib Bank A.G Zurich.
28. Habib Bank Ltd.
29. Imperial Bank Ltd.
30. I and M Bank Ltd.
31. Jamii Bora Bank Ltd.
32. Kenya Commercial Bank Ltd.
33. K-Rep Bank Ltd.
34. Middle East Bank (K) Ltd.
35. National Bank of Kenya Ltd.
36. NIC Bank Ltd.
37. Oriental Commercial Bank Ltd.
38. Paramount Universal Bank Ltd.
39. Prime Bank Ltd.
40. Standard Chartered Bank (K) Ltd.
41. Trans-National Bank Ltd.
42. Victoria Commercial Bank Ltd
43. UBA Kenya Bank Ltd.

Mortgage Finance Provider in Kenya.

1. Housing Finance Company of Kenya.