

https://dspace.org/

Big data analytics to combat cyber fraud in the South African banking sector.

Pillay, Eulene

2016
http://hdl.handle.net/10539/21921

Downloaded from a DSpace repository. https://dspace.org/

Big data analytics to combat cyber fraud in the South African banking sector

Eulene Pillay

A research report submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Business Administration

Johannesburg, 2015

ABSTRACT

Big data refers to data that is “high-volume, high-velocity and high-variety” and being able to process this unstructured data provides unparalleled opportunities and insights. Big data analytics (BDA) allows for the processing of unstructured and semi-structured data and its uses range from providing businesses with a 360 degree view of customers, to fraud detection.

This research focused on the application of BDA in cyber fraud prevention. The research aim was to identify how the use of BDA to combat cyber fraud is perceived within the South African (SA) banking sector. The intention of this research was to fill a gap that currently exists in terms of available information on how banks in SA may use BDA to combat cyber fraud.

This research made use of a qualitative research paradigm and data was collected through unstructured interviews. A triangulation strategy was used to identify the stakeholder groups and thereafter purposive sampling was applied. The data was analysed and interpreted through thematic analysis.

The data analysis revealed that cyber attacks are constantly evolving and while there is great effort being made within the banking sector to keep them under control, banks will never completely eradicate them. SA banks have introduced BDA into the fight against cyber fraud however this is still at an exploratory stage. They are also using BDA in social media and consumer analysis. There is fervour within the banks to further explore BDA in relation to the business as a whole, since they want to leverage BDA across business areas. SA banks must spend more time and effort researching and investigating BDA capabilities if they are to efficiently incorporate BDA into a cyber fraud solution.

DECLARATION

I, Eulene Pillay, declare that this research report is my own work except as indicated in the references and acknowledgements. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Eulene Pillay

Signed at ……

On the …….. day of …… 20…..

DEDICATION

This research is dedicated to:

- My darling husband Suresh Ramdial, for his everlasting support and encouragement
- My parents, for providing me with the opportunities that I am so thankful for
- My grandparents, for their guidance and support throughout my life.

ACKNOWLEDGEMENTS

I place on record my sincere thanks to:

- My supervisor, Antony Soicher for assisting me through this process
- All my research participants for their time spent and insights shared
- Investec, for the support during my MBA
- All my friends and family for their understanding and patience over the duration of the MBA.

TABLE OF CONTENTS

ABSTRACT ii

DECLARATION iii

DEDICATION iv

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER 1. INTRODUCTION 12

1.1 PURPOSE OF THE STUDY 12
1.2 CONTEXT OF THE STUDY 12
1.3 PROBLEM STATEMENT 15
1.3.1 MAIN PROBLEM 15
1.4 SIGNIFICANCE OF THE STUDY 15
1.5 DELIMITATIONS OF THE STUDY 17
1.6 DEFINITION OF TERMS 17
1.7 ASSUMPTIONS 17

CHAPTER 2. LITERATURE REVIEW 19

2.1 INTRODUCTION 19
2.2 BACKGROUND AND DEFINITION OF FRAUD 20
2.3 THE EVOLUTION OF CYBER FRAUD 22
2.4 CYBER FRAUD IN SOUTH AFRICA 23
2.4.1 ACCOUNT TAKEOVER FRAUD 25
2.5 THE CHALLENGES CYBER ATTACKS PRESENT TO ONLINE BANKING 26
2.5.1 LIMITATIONS OF TRADITIONAL FRAUD DETECTION TOOLS 30
RESEARCH QUESTION 1: 30
2.6 BDA TO COMBAT CYBER ATTACKS 30
2.6.1 FROM BUSINESS INTELLIGENCE TO BIG DATA 31
2.6.2 THE ERA OF BIG DATA 32
RESEARCH QUESTION 2: 33
2.6.3 BIG DATA ANALYTICS (BDA) 33
RESEARCH QUESTION 3: 35
RESEARCH QUESTION 4: 35
2.7 CONCLUSION OF LITERATURE REVIEW 35
2.7.1 RESEARCH QUESTION 1: 36

2.7.2 RESEARCH QUESTION 2: 36
2.7.3 RESEARCH QUESTION 3: 36
2.7.1 RESEARCH QUESTION 4: 36

CHAPTER 3. RESEARCH METHODOLOGY 37

3.1 RESEARCH METHODOLOGY /PARADIGM 37
3.2 RESEARCH DESIGN 38
3.3 POPULATION AND SAMPLE 39
3.3.1 POPULATION 39
3.3.2 SAMPLE AND SAMPLING METHOD 40
3.4 THE RESEARCH INSTRUMENT 42
3.5 PROCEDURE FOR DATA COLLECTION 43
3.6 DATA ANALYSIS AND INTERPRETATION 44
3.7 LIMITATIONS OF THE STUDY 47
3.8 VALIDITY AND RELIABILITY 47
3.8.1 EXTERNAL VALIDITY - TRANSFERABILITY 48
3.8.2 INTERNAL VALIDITY - CREDIBILITY 48
3.8.3 RELIABILITY – DEPENDABILITY AND CONFIRMABILITY 48
3.9 DEMOGRAPHIC PROFILE OF INTERVIEWEES 49

CHAPTER 4. PRESENTATION OF RESULTS 52

4.1 INTRODUCTION 52
4.2 RESULTS PERTAINING TO RESEARCH QUESTION 1 52
4.2.1 CURRENT CYBER FRAUD ACTIVITY AT SA BANKS 54
4.2.2 CYBER THREATS 55
4.2.3 CYBER FRAUD ACTIVITY IN THE FUTURE 56
4.3 RESULTS PERTAINING TO RESEARCH QUESTION 2 57
4.4 RESULTS PERTAINING TO RESEARCH QUESTION 3 59
4.4.1 CURRENT BDA ACTIVITY WITHIN THE BANKS 60
4.4.2 REASONS TO IMPLEMENT 61
4.4.3 IMPLEMENTATION CHALLENGES 62
4.4.4 OVERCOMING IMPLEMENTATION CHALLENGES 64
4.4.5 FUTURE BDA ACTIVITY WITHIN THE BANKS 65
4.5 RESULTS PERTAINING TO RESEARCH QUESTION 4 65
4.5.1 USE OF BDA FOR CYBER FRAUD PREVENTION 66
4.6 SUMMARY OF THE RESULTS 67

CHAPTER 5. DISCUSSION OF THE RESULTS 68

5.1 INTRODUCTION 68
5.2 DISCUSSION PERTAINING TO RESEARCH QUESTION 1 68
5.2.1 CURRENT CYBER FRAUD ACTIVITY AT SA BANKS 68
5.2.2 CYBER THREATS 71
5.2.3 CYBER FRAUD ACTIVITY IN THE FUTURE 73
5.2.4 CONCLUSION 76
5.3 DISCUSSION PERTAINING TO RESEARCH QUESTION 2 77

5.3.1 No consistent definition 77
5.3.2 Surrounded by hype 78
5.3.3 Gives rise to big data analytics 78
5.3.4 Requires infrastructure 79
5.3.5 Conclusion 79
5.4 Discussion pertaining to Research Question 3 80
5.4.1 Current BDA activity within the banks 80
5.4.2 Reasons to implement 83
5.4.3 Implementation challenges 84
5.4.4 Overcoming implementation challenges 86
5.4.5 Future BDA activity within the banks 87
5.4.6 Conclusion 89
5.5 Discussion pertaining to Research Question 4 90
5.5.1 Use of BDA for cyber fraud prevention 90
5.6 Conclusion 92
CHAPTER 6. CONCLUSIONS & RECOMMENDATIONS 93
6.1 Introduction 93
6.2 Conclusions of the study 93
6.3 Recommendations 97
6.3.1 Business intelligence 97
6.3.2 Fraud prevention 97
6.4 Suggestions for further research 98
6.4.1 BDA capabilities for fraud detection 98
6.4.2 Benefits and risks of using open-source technology within banks 98
6.4.3 Internet maturity vs cyber fraud activity globally 98
6.4.4 Openness to technological advancements in smaller banks vs the big four banks 99
6.4.5 Impact of legislation on BDA for social media 99
REFERENCES 100
APPENDIX A 104
Actual research instrument 104
APPENDIX B 105
Study information sheet and interview consent form 105

LIST OF TABLES

Table 1: Profile of stakeholder groups 40

Table 2: Profile of interviewees 42

Table 3: Planned vs Actual sample for the five largest SA banks 50

Table 4: Planned vs actual sample for banking industry and advisory institutions 51

Table 5: Planned vs actual sample for IT vendors 51

Table 6: Global and organising themes for Research Question 1 53

Table 7: Top five themes for current cyber fraud activity at SA banks 54

Table 8: Top five themes for cyber threats 55

Table 9: Top five themes for cyber fraud activity in the future 56

Table 10: Top five themes for big data awareness 58

Table 11: Global and organising themes for Research Question 3 60

Table 12: Top five themes for reasons to implement 62

Table 13: Top five themes for implementation challenges 63

Table 14: Top three themes for overcoming implementation challenges 64

Table 15: Global and organising themes for Research Question 4 66

Table 16: Key findings to research questions 95

LIST OF FIGURES

Figure 1: The 3V's model of big data (Zhang & Huang, 2013) 13

Figure 2: Emerging technologies and trends (EY, 2014)..... 14

Figure 3: The fraud triangle (Cressey, 1953)..... 21

Figure 4: National Cyber Security Policy Framework objectives (Wolfpack, 2013)..... 25

Figure 5: (a) Social engineering attack; (b) Drive-by download attack (Chang et al., 2013) 27

Figure 6: Phishing attack (Paganini, 2013) 29

Figure 7: The four V's of big data (EY, 2014)..... 32

Figure 8: Structure of a thematic network 46

Figure 9: Cyber fraud expertise among the three stakeholder groups 53

Figure 10: Current cyber fraud activity at SA banks..... 54

Figure 11: Cyber threats 55

Figure 12: Cyber fraud activity in the future 56

Figure 13: Big data awareness within the sample from the banks 57

Figure 14: Big data awareness..... 58

Figure 15: Current BDA activity within banks..... 61

Figure 16: Reasons to implement 62

Figure 17: Implementation challenges 63

Figure 18: Overcoming implementation challenges 64

Figure 19: Future of BDA within the banks..... 65

Figure 20: Use of BDA for cyber fraud prevention..... 66

CHAPTER 1. INTRODUCTION

1.1 Purpose of the study

The purpose of this research was to establish whether big data analytics (BDA) may be leveraged to spot trends and patterns consistent with potential cyber fraud activity within the banking sector in South Africa (SA).

1.2 Context of the study

Big data is a term that has become widely used in recent years and on the surface seems to refer to a concept that emerged in the 21st century. Digging a little deeper, however, reveals that researchers may have alluded to the concept in the late 1990s (Diebold et al., 2012). Even more interesting, is that more than 70 years ago, Rider (1944) was already trying to quantify the rate at which data will grow by the 21st century. In his book "*The scholar and the future of the research library",* Rider uses the Yale University library to illustrate this concept and suggests that Yale would contain nearly two hundred million volumes by the year 2040 and a staff complement of over five thousand people would be required to catalogue them. This suggests that the concept of big data has been around for decades even if the term may have only been coined within the last 20 years.

The most commonly accepted definition of big data comes from Gartner. Gartner defines it as being data that is "high-volume, high-velocity and high-variety" and that requires special processing in order to add value (Gartner, 2013). The basis for this definition comes from Laney (2001), who in his unpublished research discusses changes in data velocity, data variety, data volume and the inability of traditional data management tools to handle these changes. Figure 1 below, makes clear the 3 V's of big data.

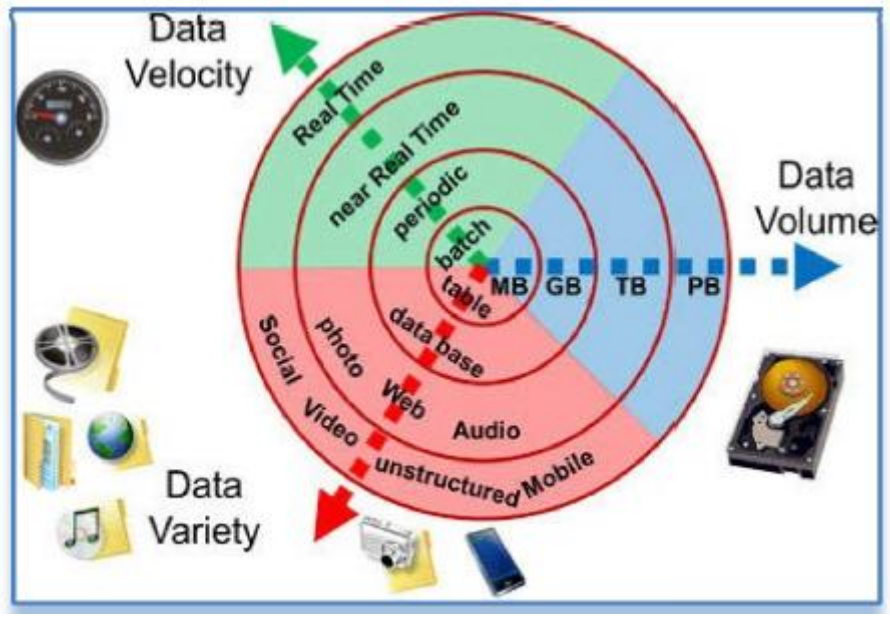

Figure 1: The 3V's model of big data (Zhang & Huang, 2013)

Today, big data is not only a universally recognised term and technological phenomenon, but also an emerging discipline (Diebold et al., 2012). Ninety per cent of data in organisations is unstructured (IBM, 2014). This means that decision making processes are based on ten per cent of data only. Therefore, being able to process unstructured data provides unparalleled competitive advantage and this brings us to big data analytics (BDA).

BDA now allows for unstructured data to be processed at high speeds. The uses of BDA vary from providing companies with a 360 degree view of customers to social media analysis and fraud detection (Chen, Chiang, & Storey, 2012). Figure 2 indicates that big data has been identified as an emerging trend and its use in business is imminent, with it being classified as being "around the corner".

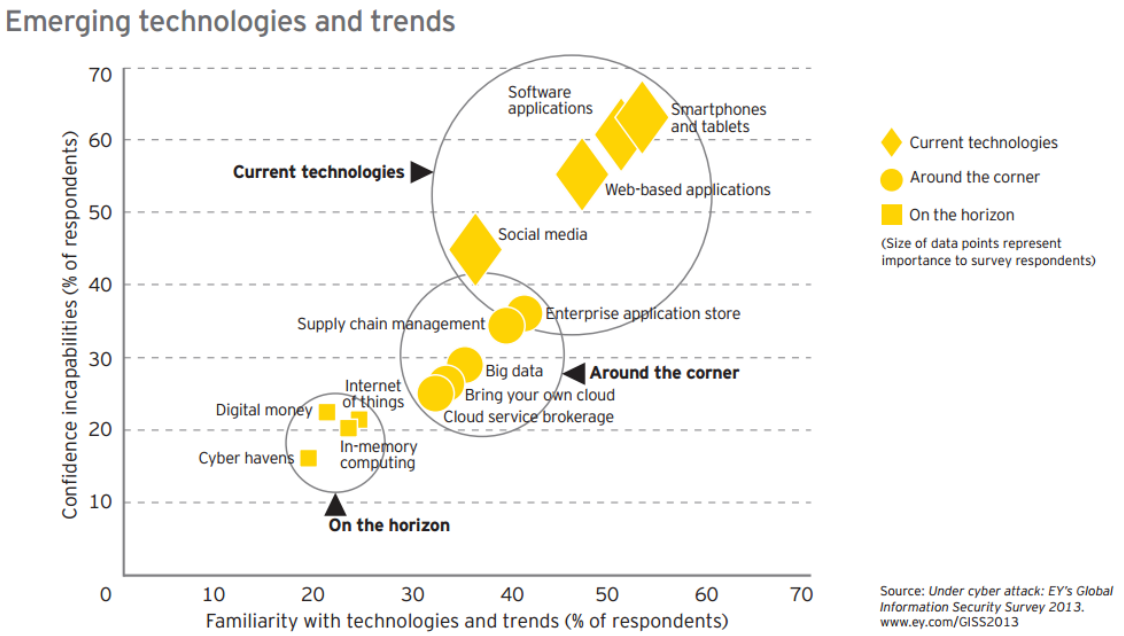

Figure 2: Emerging technologies and trends (EY, 2014)

This research focuses on the uses of BDA in cyber fraud prevention. South African law defines fraud as "the unlawful and intentional making of a misrepresentation which causes actual and or potential prejudice to another" (Van Niekerk, 2008). The intention of this research was to fill a gap that currently exists in terms of available information on how banks in SA may use BDA to detect and combat cyber fraud. While there has been work done internationally in the field, it's not always relevant or easily applied to the SA context.

The cyber world has revolutionised our lives and transformed the way we communicate, socialise and conduct business. However this paradigm shift has also brought with it many unintended consequences (Raghavan & Parthiban, 2014) . The internet facilitates global knowledge sharing, but has also become an easy entry point for exploitative technologies. Gilman (2011) describes this as deviant globalisation, where individuals utilise technical infrastructure derived from globalisation for their repugnant and exploitative means; a good example of this is cyber fraud.

In this technologically booming information age, as banks move to a larger online presence, the patterns of cyber fraud can be found in the vast amounts of

machine data. This unstructured machine data is generated through various ways eg authentication systems, firewalls, web proxies etc (Splunk, 2014). By analysing this data to detect patterns of fraud, banks may be able to act on fraud alerts in real time in order to prevent it.

Cyber fraud attacks in SA like everywhere else in the world, is on the increase. In 2013, a form of malware known as Dexter, was inserted into point-of-sale (POS) devices at certain fast-food outlets in SA. This proved to be one of the nastiest breaches of customer card data in SA's history and cost banks tens of millions of rands (Mcleod, 2013). In situations like these, it is the banks that need to take the fall for their customers and all banks are generally hit equally hard. With the Dexter virus however banks that were quicker at the roll out of chip cards were less affected. This suggests that banks need to be more proactive in the fight against fraud and need to constantly investigate more relevant ways to combat cyber fraud.

The SA banking sector consists of 17 registered banks. There are also three mutual banks, two co-operative banks and a mix of 53 local branches and representative offices of foreign banks (SARB, 2014). This research however focused on the five largest SA banking groups: Standard Bank Group, Absa Group, Firstrand Banking Group, Nedbank Group, and Investec Group.

1.3 Problem statement

1.3.1 *Main problem*

The research aim was to identify how the use of big data analytics to combat cyber fraud is perceived within the SA banking sector.

1.4 Significance of the study

Firewalls, intrusion detection and prevention systems, database encryption and various other security solutions make use of signatures and blacklists in the fight against cyber attacks (Ahn, Kim, & Chung, 2014). These security

mechanisms are ineffective as they only detect a narrow scope of threats that have been previously encountered (Mahmood & Afzal, 2013). Therefore in order to keep up with these insidious attacks, security communities need to find new solutions to detect and prevent previously unheard of attacks.

The Symantec Intelligence Report for 2012 identified South Africa as experiencing the second highest phishing attacks globally, with one in every 170.9 emails being phishing emails (eNCA, 2013). The typical result of phishing attacks is account takeover fraud which is one of the basic forms of identity theft fraud. It is where the fraudster retrieves and makes use of a victim's personal information to hijack existing accounts and conduct unauthorised transactions using them. The growth of e-commerce and the changing nature of fraud make it imperative that more advanced measures be put in place to combat fraud and more specifically fraud resulting from cyber attacks. This research aims to make known the relevance of BDA in the fight against cyber fraud.

The study may provide guidance to those concerned with combating cyber fraud within the banking sector and the broader financial services industry in South Africa. While this research focuses on the banking sector, the techniques discussed may also be applied to the insurance industry. This research therefore may be extrapolated to include other sectors within the financial services industry. The study may also be of interest to those involved in the application of BDA in solving business problems. Interested groups may therefore range from those involved in fraud prevention to those involved in business intelligence. The potential take-away may be different for each of these groups but it is hoped that they may benefit from having more insight into the use of BDA within the SA banking sector.

1.5 Delimitations of the study

- In this research the users of BDA were limited to the South African banking sector
- The area of cyber fraud is vast; this research focused specifically on account takeover fraud resulting from the most prominent cyber threats facing the SA banking sector.
- Protection of Personal Information Act (POPI) may introduce ethical issues in the use of BDA. However this research did not delve into how legislation may impact the use of BDA.

1.6 Definition of terms

Big data – refers to "high-volume, high-velocity and high-variety" data that requires special processing in order to add value (Gartner, 2013).

Big data analytics – refers to the methods employed to extract insights and value from big data (Mahmood & Afzal, 2013).

Cyber fraud – "the causing of a loss of property to another by: (a) any input, alteration, deletion or suppression of computer data, (b) any interference with the functioning of a computer system; with fraudulent or dishonest intent of procuring, without right an economic benefit for oneself or for another" (Fletcher, 2007).

1.7 Assumptions

This research made the following assumptions:

- That the interviewees within the banking sector and industry bodies were open and honest about their approaches to combating cyber fraud. This was a reasonable assumption, since the industry as a whole stands to benefit from knowledge sharing and insights in this regard. Interviewees withholding information may have resulted in this research lacking depth.

- That the five largest banking groups in SA have the significant expertise and insight to cover all the current BDA approaches to combat cyber fraud. This was a reasonable assumption as these banks have the necessary scale to be forerunners in this area. If this is untrue, this research may lack valuable insight that some of the smaller players may have been able to provide.
- That the technological service providers were honest in their responses about the capabilities of their tools and did not use this as a platform for a sales pitch. This was a reasonable assumption as their tools are likely already being used or may be put to test in the near future. If these vendors were not true in their responses, it has the potential to mislead readers of this report.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

The literature review is an iterative process that takes place throughout the research and provides insights into what is already known about the area of interest (Bryman, 2012). An efficient literature review sets the basis for advancing knowledge by closing areas where surplus research exists and revealing areas where research is required (Webster & Watson, 2002).

This literature review will address the research problem directly as there are no sub problems. The research aim is to identify how the use of big data analytics to combat cyber fraud is perceived within the SA banking sector. The literature first looked at the evolution of fraud and the theory behind it. From there it focused on cyber fraud and the challenges that this new variation of fraud presents to South Africa and then more specifically the banking sector in SA. Cyber fraud involves a vast and constantly changing realm of information. The researcher focused on the most insidious cyber threats to online banking and also looked at the limitations of the traditional anti-fraud tools in fighting this.

The researcher then looked at the rise of business intelligence and its evolution to include BDA. From there the researcher attempted to illustrate through the literature how BDA may be relevant to the fight against cyber fraud. However, as mentioned earlier the area of cyber fraud is vast and for purposes of this research it was not possible to investigate the use of BDA in relation to all potential cyber fraud activity. The researcher therefore honed in on what was identified as being one of the most common types of cyber fraud, ie account takeover fraud resulting from cyber attacks. The researcher investigated the relevance of BDA in this specific context. The researcher then established the research questions.

2.2 Background and definition of fraud

South African law defines fraud as “the unlawful and intentional making of a misrepresentation which causes actual and or potential prejudice to another” (Van Niekerk, 2008). The inference drawn from this definition is that fraud has been around for as long as dishonest people have walked the earth. In his 1776 book “*An Inquiry into the Nature and Causes of the Wealth of Nations”*, Adam Smith, one of the world’s great economists, identified fraud as a major shortcoming in the world of commerce (Dorminey, Fleming, Kranacher, & Riley Jr, 2012).

In the 1950’s Donald Cressey produced seminal pieces of research, which helped explain the motivation of individuals for committing fraud. In his book, *Other people's money; a study of the social psychology of embezzlement,* Cressey (1953) hypothesised that there must be three factors present simultaneously in order for an individual to commit fraud. These factors were pressure, opportunity and rationalisation. Pressure refers to the stress that an individual feels when confronted with a non-sharable financial problem. Opportunity refers to the access the individual has to commit the crime and rationalisation refers to the process where the individual convinces him/herself that he/she is justified in committing the act. It is an attempt to minimise the cognitive dissonance that the individual experiences internally (Cressey, 1953). This hypothesis and the three emerging factors led to the creation of the popular fraud triangle shown in Figure 3 below.

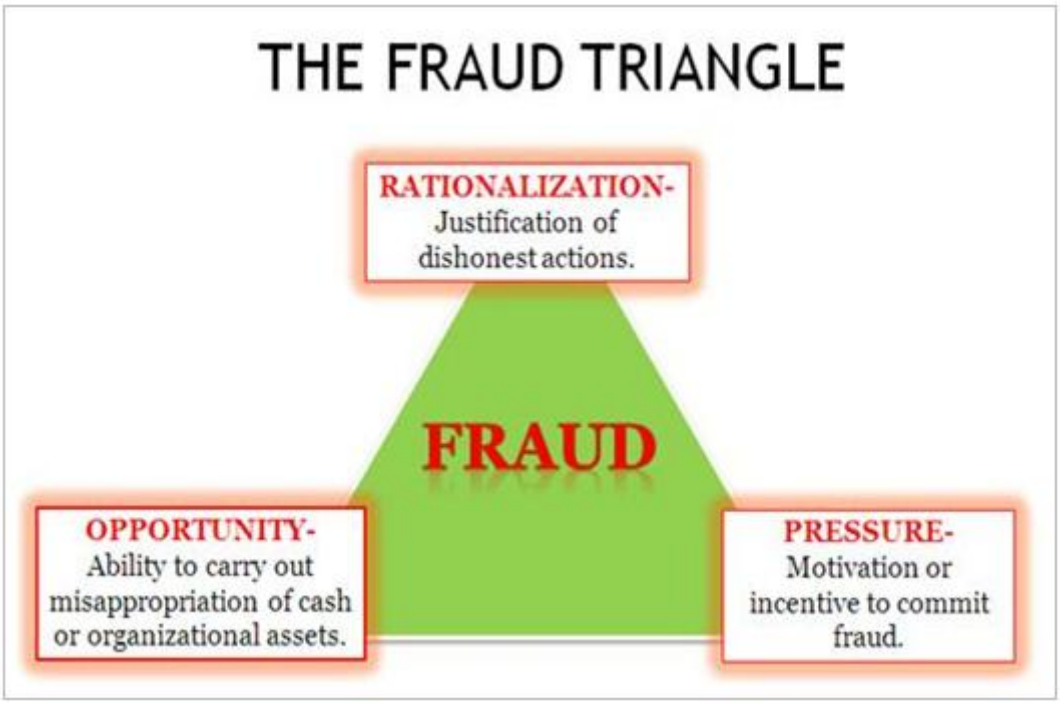

Figure 3: The fraud triangle (Cressey, 1953)

Since the emergence of the fraud triangle in the early 1950's additional models and theories were developed in an attempt to further explain the motivation of individuals to commit fraud. This includes the fraud diamond, the fraud scale, MICE and the predator versus the accidental fraudster (Dorminey et al., 2012). However the fraud triangle was still seminal in nature and the researcher did not delve into every model. Worth mentioning though is the ABC model of fraud which emerged in 2009. This model refers to the: Bad **A**pple, Bad **B**ushel and Bad **C**rop. The Bad Apple is the individual that commits fraud, the Bad Bushel refers to a group of individuals or collusive behaviour in committing fraud and the Bad Crop refers to the cultural and societal forces that enable fraud (Ramamoorti, Morrison, & Koletar, 2009). This model is relevant in explaining the increasing prevalence of fraud today.

Once practices, even unethical ones, become trendy it has the potential to place pressure on competitors to at least consider copying those practices. This is in line with the concept of the bad crop and the societal forces that enable fraud (Dorminey et al., 2012). Another enabling factor is the rapidly changing and complex business environments that result in controls continually falling out of date and becoming ineffective. We are living in the technological age, where

transactions happen rapidly and with ease while audit and monitoring functions are becoming increasingly difficult. Advancements in technology have helped us create a global village which at the same time has come with great perils. Even the nature of fraud has changed and evolved many times over since Cressey (1953) first introduced the fraud triangle.

We are now dealing with a new phenomenon called cyber fraud which is quickly spiralling out of control. Cyber fraud refers to fraudulent activity conducted using the internet and where the targets are often financial institutions such as banks. However, due to the varying nature of fraud and its mutation across the industries, there is no consistent definition of cyber fraud. Using our earlier definition of fraud as a basis, The researcher defined cyber fraud, for the purpose of this research, as “the interference with computer data or the functioning of a computer system in an attempt to cause actual and or potential prejudice to another and economic benefit without right for oneself or another” (Fletcher, 2007).

2.3 The evolution of cyber fraud

While online banking has been around since the 1980s, cases of cyber fraud only became prominent from around 2004 (Paganini, 2013). This means the rapid escalation in cyber fraud attacks has taken place over a very limited timeframe. In July 2007, the first version of the Zeus trojan was detected while being used to steal information from the Department of Transportation in the United States (Paganini, 2013). Little did we know that in the years to follow, Zeus would become one of the most dangerous cyber threats to online banking. Today Zeus is probably the most rampant varieties of malware on the web and is being used in phishing scams on a mammoth scale.

In 2010, the FBI arrested members of a syndicate believed to be using this prolific botnet-based malware Zeus. This malware was used by the syndicate to collect credentials from banks in an attempt to steal $220 million. The syndicate was apprehended but not before successfully stealing $70 million. Then in September 2011 a more resilient mutation of Zeus called GameOver emerged

(Andriesse, Rossow, Stone-Gross, Plohmann, & Bos, 2013). Today, there are more than one million GameOver Zeus (GOZ) infections worldwide with losses running into hundreds of millions of dollars. The GOZ malware generally propagates via phishing campaigns and also infects computers under its control with Cryptolocker. Cryptolocker is a ransomware that encrypts the victim's files and holds them hostage until the owner makes payment to the perpetrators. It is believed that the malware is being managed by a group of hackers in Ukraine and Russia. The FBI is partnering with ten countries in a takedown operation of GOZ (FBI, 2014).

Another prolific cyber attack on banks has been Operation Ababil which was carried out by a group known as IZZ as-din Al Qassam Cyber Fighters. As part of this operation distributed denial of service (DDoS) attacks knocked out the websites of various banking giants in the United States, including JP Morgan Chase, Wells Fargo, and Bank of America. While the motivation behind Operation Ababil was not to commit fraud, DDoS attacks are often used as a diversion tactic while attackers carry out fraudulent wire transfers (Paganini, 2013).

As we continue to see advancements in technology, so too are the threats becoming increasingly sophisticated and intelligent. Nowadays we are confronted with targeted and persistent attacks such as Advanced Persistent Threats (APTs). APTs target systems and exploit their vulnerabilities over a prolonged period of time (Sood & Enbody, 2013). The aim of the attacker is to gather as much information about the target before the attack, so that a range of vectors may be built and used against the target. Therefore APTs are dependent on data gathering. Most APT attacks generally start with a targeted phishing or what is termed a spear-phishing email.

2.4 Cyber fraud in South Africa

Internet connectivity in South Africa is on the increase. There are huge investment projects underway to bridge the digital divide for various social and economic reasons, however with this comes new risk. New internet users are

generally not security-savvy and this makes them especially vulnerable to cyber attacks (Wolfpack, 2013). This puts the country as a whole at risk, since unsuspecting users can be used as entry points for widespread compromisation. Humans have therefore been identified as being the weakest link with social engineering being a massive problem.

According to Wolfpack (2013) DDoS attacks and system paralysis were identified as being the most dangerous cyber threats affecting SA. Due to the various undersea cables linking SA to the global internet, a large scale DDoS attack could have dire consequences for the country. Even more concerning is that South Africa is still in the process of creating a national Computer Security Incident Response Team (CSIRT). Nonetheless government has recognised the threat of cybercrime and in March 2012, approved the National Cyber Security Policy Framework for South Africa (Wolfpack, 2013). While the intentions behind the policy are great, there is still some way to go before SA achieves what the policy sets out to do. Figure 4 illustrates what this policy attempts to achieve.



Figure 4: National Cyber Security Policy Framework objectives (Wolfpack, 2013)

Internet banking is the most targeted cyber service in South Africa (Wolfpack, 2013). The Symantec Intelligence Report for 2012 identified phishing as the major cyber attack facing internet banking in SA, with account takeover fraud being identified as the most prevalent fraud resulting from cyber attacks in SA (eNCA, 2013).

2.4.1 *Account takeover fraud*

Account takeover fraud occurs when fraudsters accumulate sufficient personal identity information (PII) in order to hijack a legitimate account (PWC, 2011). PII refers to a first name or initial, and surname, together with one or more

additional pieces of personal information eg identity number, account number, credit or debit card number, etc. Account takeover fraud is on the rise globally and is the fastest growing fraud in SA. It increased by 100.5 per cent from 2012 to 2013 (AFP, 2014). According to PWC Forensic Services the largest cause of account takeover fraud is phishing and this includes both phished customers and phished insiders (PWC, 2011).

2.5 The challenges cyber attacks present to online banking

Looking at the emergence of cyber fraud and relating this back to Cressey's fraud triangle, it can immediately be seen that in the age of cyber fraud all three factors (pressure, opportunity and rationalisation) are amplified. As we emerge from a global recession and still face uncertain financial markets, financial problems are rife thereby amplifying the **pressure** factor. The introduction of the cyber world has introduced a plethora of new ways and methods for would be criminals to commit fraud and also to do so with anonymity, therefore amplifying the **opportunity** factor. In cyberspace, even law abiding individuals find it easier to **rationalise** and justify unethical behaviour that they would not normally condone out in the real world.

For example, an individual may not be willing to steal a music record from a store but will rationalise downloading a pirate copy online. This indicates that all three factors are amplified and in accordance with Cressey (1953), this would mean there is increased motivation to commit cyber fraud and this may explain the prevalence of fraudulent behaviour today. The cyber attacks used to carry out cyber fraud are varied, advanced and constantly evolving. For purposes of this research, the researcher only focused on the most menacing types of cyber attacks on online banking as identified by Paganini (2013). These are discussed below.

a) Web-based malware attacks

Malware refers to malicious software that attacks parts of a computer without the owner's consent or even knowledge. This can be in the form of spyware, computer viruses, worms or any other form of unwanted

software. The most popular types of malware on the web are trojans, adware and computer worms (Chang, Venkatasubramanian, West, & Lee, 2013). These attacks are devastating to banks, eg the “man in the browser” attack where the bank domain is specified in the configuration file of a trojan (eg Torpig) and when a user visits that bank site, a form in the login page requests personal information. The stolen information is then uploaded to drop servers and in the case of Torpig this happens every 20 minutes (Cárdenas, Radosavac, Grossklags, Chuang, & Hoofnagle, 2010). Malware is often used in phishing attacks.

There are two types of web-based malware attacks:

- Social engineering – various psychological manipulations and decoys are used to trick the user into authorising the download and installation process of malware (Chang et al., 2013). This method exploits cognitive biases in the reasoning process of the human mind, indicated in Figure 5 (a) below.
- Drive-by download – these are web pages that are designed to contain malicious code that automatically triggers the download and installation of malware (Chang et al., 2013). This method exploits the vulnerabilities of the computer system, indicated in Figure 5 (b) below.

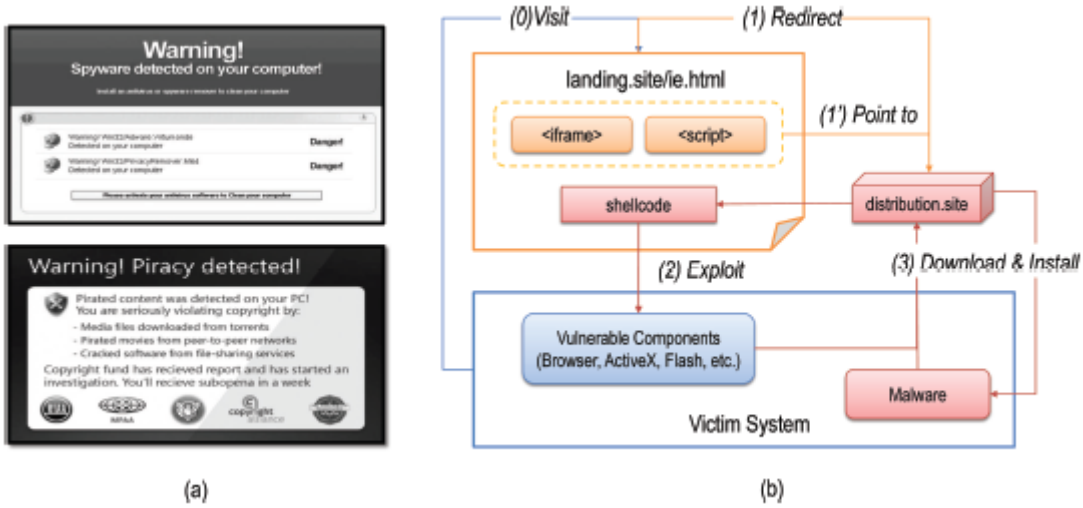

Figure 5: (a) Social engineering attack; (b) Drive-by download attack (Chang et al., 2013)

Another avenue that leads to the creation of malware is vulnerability research, which involves researchers trying to find potential vulnerabilities in security software. However, following the trend of deviant globalisation, these researchers become connected to underground groups and sell information derived from research in underground markets (Cárdenas et al., 2010). Exploit tools (also referred to as shellcode) are then developed to take advantage of these vulnerabilities and this ultimately leads to the creation of malware. This is a major threat to banks that are continually monitoring and testing their systems for vulnerabilities. This information can easily fall into the wrong hands with lethal consequences.

b) Botnet attacks

Botnets refer to all the infected or compromised computers under the control of a single authority (Cárdenas et al., 2010). Attackers install programs on compromised computers in order to keep track of their botnet. Through a botnet, attackers have various methods of carrying out fraudulent activity at a bank eg theft of client information, spamming, click fraud, phishing, as well as DDoS attacks as seen in Operation Ababil mentioned earlier. DDoS attacks refer to cases where botnets are used to flood a target with traffic, so that every point from the internet to the victim is blocked. This is often used for extortion, blackmail and even to express strong radical views (Cárdenas et al., 2010).

c) Phishing

Phishing attacks are becoming increasingly sophisticated and are mainly directed at banks. These attacks make use of social engineering techniques to trick users and steal personal information and bank account details (Paganini, 2013). This is done mainly through email and web spoofing (Mahmood & Afzal, 2013). In a phishing attack, fraudsters send e-mails and other forms of communication to victims. The content of the communication looks legitimate and requests victims to click on a link or attachment. Clicking on the link then replicates a bank's website,

where the victim unwittingly enters personal information. Fraudsters then make use of malware to steal this information (Paganini, 2013).

According to APWG Global Phishing Survey report 2013 phishers are targeting shared web hosting servers to conduct mass phishing attacks (Paganini, 2013). The fraudsters compromise a server and update its configuration to allow for phishing pages to be displayed and this allows them to exploit thousands of websites during these attacks. Figure 6 visually explains a phishing attack.

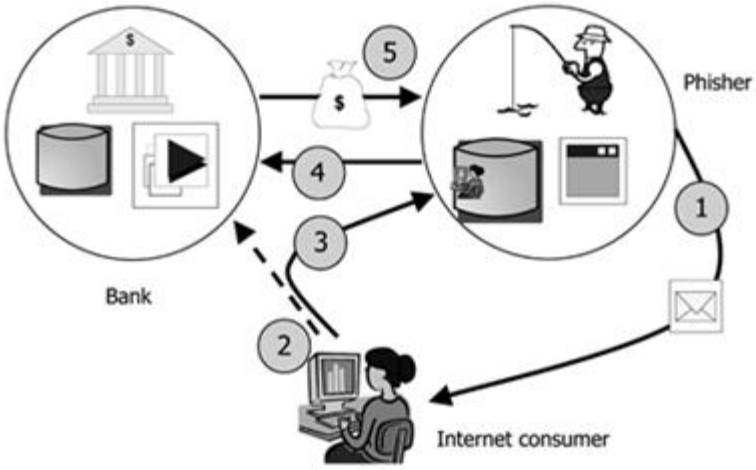

Figure 6: Phishing attack (Paganini, 2013)

d) Pharming

Pharming refers to parasitic URL naming, where perpetrators create domain names similar to websites and attempt to take advantage of the fact that users may incorrectly type the URL address. The URL then leads to phony sites where unsuspecting users are exploited (Banerjee, Barman, Faloutsos, & Bhuyan, 2008). Banking websites are often the target of pharming or what is also referred to as URL poaching.

Of course the information collected using malware, botnets, phishing and pharming must be monetised to be worth anything to the cyber attackers. Some of the ways this can be done is through, account takeover fraud, application

fraud, money mules, etc. In order for this to happen cyber criminals require the assistance of various other fraudsters and swindlers which once again perpetuates deviant globalisation. For purposes of this research, the researcher will focus on account takeover fraud that results from cyber attacks.

2.5.1 *Limitations of traditional fraud detection tools*

Traditional anti-fraud technologies are built for structured data analysis. Their rigid architectures do not support the velocity and variety of the unstructured data created by cyber activity (Splunk, 2014). Besides being limited in the types of data they ingest, they also have scale limitations. They are unable to quickly return results on massive volumes of data. These traditional tools generally make use of data normalisation and schemas, therefore users are constrained in the types of investigations carried out and the visualisation of the data (Splunk, 2014). They also often require extensive configuration and maintenance as they typically involve multiple data stores and products. They are very costly to deploy and maintain. These traditional ant-fraud tools are therefore ill- equipped to deal with the enormous amounts of unstructured data generated through online activity.

Research question 1:

How are SA banks faring in the fight against cyber fraud?

2.6 BDA to combat cyber attacks

As cyber attackers keep up with the trends of technology to advance their cause, existing security systems are failing dismally at detecting or preventing these attacks (Ahn et al., 2014). This can be attributed to the inability of traditional fraud detection tools to detect cyber fraud, as discussed earlier, as well as the inability of current pattern matching technologies to detect never seen before attacks. Criminals are therefore now brazenly and successfully targeting large-scale systems.

2.6.1 *From business intelligence to big data*

In 1958, Hans Peter Luhn in his article, *A business intelligence system,* described how automated solutions would be developed to allow for data processing and the circulation of information within organisations. While difficult to confirm, it is believed by many to be the first reference to business intelligence (BI) in the data processing context. Luhn (1958) defined business intelligence as the ability to capture the relationships between facts so that they may be used as a guide for decision making.

In 1989, Howard Dresner proposed business intelligence as an umbrella term that refers to tools and techniques used to improve business decisions made using fact-based systems (Elena, 2011). However it was only in the 1990s that BI started to gain traction. Today, while business intelligence is core to business operations and provides functions such as data mining, reporting and analytics; the requirements of business are changing (Chen et al., 2012).

With the advent of the internet and advancements in technology our methods of communication have evolved over time. Today, we communicate via email, video conferencing, telephone conferencing, memos, blogging etc. Much of these new methods of communication give rise to either unstructured or semi-structured data (Blumberg & Atre, 2003). Structured data generally refers to data that can be stored as rows and columns and represented in a relational database, while unstructured data does not fit into a database and is often stored as text (Park & Song, 2011).

While BI accommodates for and includes both structured and unstructured data; searching and processing unstructured data is difficult and has posed somewhat of a conundrum to the information technology industry (Blumberg & Atre, 2003). Due to this, businesses have in the past ignored unstructured data and focused only on structured data which was easier to analyse and search. This suggests that business decisions based on data analytics are often uninformed, considering that structured data only accounts for approximately 15 per cent of total data (Blumberg & Atre, 2003). This puts into perspective how

much untapped insight is contained in the large volumes of unstructured data and brings us to the discipline of BDA.

2.6.2 *The era of big data*

The definition of big data is constantly evolving with IBM adding a further "v" for veracity to the definition of Gartner (2013). Veracity relates to the accuracy of the data and essentially refers to how we would want use big data rather than what it is (Zikopoulos, Parasuraman, Deutsch, Giles, & Corrigan, 2012). Figure 7 explains the four V's of big data in more detail. For purposes of this research, big data is defined as large volumes of structured and unstructured data that can be aggregated, stored, and analysed at high speeds to provide insights.

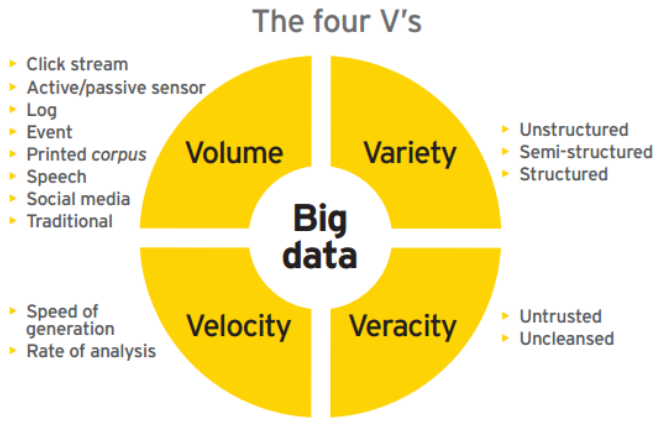

Figure 7: The four V's of big data (EY, 2014)

Nonetheless, it is not the volume, variety or velocity of data that makes big data revolutionary, but rather that we can now tap into those volumes for insights like never before (Bollier & Firestone, 2010). With big data tools, we can now aggregate data across disciplines as disparate as geology and finance. We can link data sets to have an all encompassing view of events (Bollier & Firestone, 2010). In fact in addition to forecasting, real-time processing of big data also allows for nowcasting. Nowcasting refers to predicting the present and is a term used by meteorologists for predicting short-term weather conditions (Bollier & Firestone, 2010). With the emergence of big data, the term is now being used beyond weather. With real time processing of big data we can nowcast the incidence of flu in different parts of the world at a specific point in time or alert ourselves to fraudulent activity as it happens (Bollier & Firestone, 2010).

Research question 2:

How is big data perceived within the SA banking sector?

2.6.3 *Big data analytics (BDA)*

Following on from big data is BDA which focuses on the extraction of insights from big data (Mahmood & Afzal, 2013). BDA is the new buzz and is being sold as a solution to various business problems and is already actively being used in providing businesses with a 360 degree view of customers and in social media analysis (Chen et al., 2012). More recently BDA is being punted in the fight against cyber fraud. The motivation behind this is the assumption that patterns and behaviours exist in big data. In terms of cyber fraud this means the ability to separate the normal from the abnormal and more specifically the lawful from the suspicious and malicious (Mahmood & Afzal, 2013).

Data, when used for security purposes can be derived from both active and passive data sources. Examples of passive sources include internet protocol (IP) location, global positioning system (GPS) date, location of log in, etc. Active data sources can include one-time passwords, digital certificates, biometric data etc. BDA has the ability to draw on both active and passive sources of data to create an holistic view of web traffic and to identify any abnormalities in user behaviour (Mahmood & Afzal, 2013). This provides the opportunity for users to be singled out and quarantined through preventative techniques if need be. Real-time searches and correlations can be performed across data sets to connect disparate events to identify fraud as it occurs (Splunk, 2014).

Using BDA banks can create baselines of normal activity and then apply analytical techniques to detect anomalies or outliers that indicate potential fraudulent activity. Such anomalies can be found in historical data as well as real-time data (Splunk, 2014). Through machine learning techniques, the system can automatically process the data to look for such anomalies and based on the findings, block client accounts without any human intervention. BDA through machine learning may also detect never before seen patterns of fraud.

BDA in cyber fraud prevention focuses on identifying and studying both familiar and unknown attack patterns in order to:

- identify hidden threats,
- predict future attacks, and
- track down perpetrators.

Mahmood and Afzal (2013) identified the following ways in which BDA can achieve this:

a) *Network traffic*
   By looking for abnormal traffic patterns, thereby identifying and predicting sources and destinations that are suspicious

b) *Web transactions*
   Identifying or predicting abnormal user patterns

c) *Network servers*
   Identifying and predicting abnormal server manipulation

d) *Network source*
   Identifying and predicting abnormal machine usage patterns, by looking at the type of data the machine sends, receives and processes

e) *User credentials*
   By identifying abnormal, user-specific behaviour eg unusual transactions and/or access times

While there is information available internationally on the use of BDA in cyber fraud prevention, there is almost no information on its relevance and application to the SA context and more specifically within the banking sector in SA.

Research question 3:

How is the use of big data analytics perceived within the SA banking sector?

Research question 4:

How can SA banks use big data analytics in the fight against cyber fraud?

2.7 Conclusion of Literature Review

The theory behind both the fraud triangle and the ABC model of fraud suggests that present conditions will exacerbate the prevalence of cyber fraud. The literature indicates that a common way for fraudsters to monetise information stolen through cyber attacks is through account takeover fraud. The main cause of account takeover fraud is phishing. Banks need more relevant ways of securing and protecting their client information.

With the emergence of BDA, banks have the capability to apply predictive, behavioural and security analytics on vast amounts of data. It seems that the logical next step for banks in the fight against cyber fraud would be to make use of the large amounts of structured and unstructured data at their fingertips, by leveraging BDA.

In order to combat cyber fraud, it is important that the players within the banking industry share information about major incidents and trends. Banks, industry bodies and vendors of technology need to work collaboratively to build defence mechanisms (Paganini, 2013). The literature suggests that BDA can be a powerful tool in the fight against cyber fraud. There is limited information available on the use of BDA within South Africa and even less information available publicly about its use within the banking sector. Through the following research questions, the researcher will attempt to test perceptions on the use of BDA within the SA banking sector as well as explore the use of BDA in the fight against cyber fraud.

To summarise the research questions to be answered are:

2.7.1 *Research question 1:*

How are SA banks faring in the fight against cyber fraud?

2.7.2 *Research question 2:*

How is big data perceived within the SA banking sector?

2.7.3 *Research question 3:*

How is the use of big data analytics perceived within the SA banking sector?

2.7.1 *Research question 4:*

How can SA banks use big data analytics in the fight against cyber fraud?

CHAPTER 3. RESEARCH METHODOLOGY

3.1 Research methodology /paradigm

This research makes use of a qualitative research paradigm due to the nature of the research problem. Bryman (2012) describes qualitative research as emphasising words in the collection and analysis of data and as being constructionist, interpretivist and inductivist. According to Bluhm, Harman, Lee, and Mitchell (2011), sometimes the state of literature related to a topic may require further examination in order to identify new phenomena and causal mechanisms through an inductive qualitative approach. On the other hand, research related to an established phenomenon generally addresses issues of prevalence and calibration which is consistent with a deductive quantitative research (Bluhm et al., 2011). Big data analytics is a relatively new phenomenon and its use in combating cyber fraud has not been thoroughly explored. This research further explored this phenomenon and tested perceptions through interpretive inquiry, which could only have been done through a qualitative abductive approach.

A qualitative approach is like exploring a maze that has multiple paths and entry points which you enter without a map. It may lead you to dead ends that force you to backtrack but you forge ahead never really knowing whether any path will lead you to a place of importance (Grinnell Jr & Unrau, 2010). Similarly, when the researcher embarked on this exploratory study, it was to test perceptions through an unstructured interviewing technique, not knowing what direction the data gathering process may take.

Qualitative research has a high degree of flexibility and requires limited structure. Considering the exploratory nature of this research, the limited structure of the qualitative approach, allowed the researcher to thoroughly probe the perspectives of the research interviewees through the unstructured interviewing technique.

3.2 Research Design

A cross-sectional research design is defined as one in which quantitative data is collected on multiple cases, at a single point in time in order to detect patterns of association on two or more variables (Bryman, 2012). Therefore the cross-sectional design falls squarely in the realm of quantitative research. However, qualitative research also follows a cross-sectional design when the research instrument is an unstructured interview.

Therefore although this research is based on a qualitative research paradigm, it makes use of a cross-sectional design and fulfils the above definition in the following ways:

a) An unstructured interview was employed as the research instrument and qualitative content analysis was applied on the interview documents
b) Experts from each of the five banks, from IT vendors and from industry and advisory bodies within the banking sector were interviewed
c) These interviews took place over a specific period of time, limited to three months
d) This research attempted to find patterns of association between BDA and cyber fraud prevention.

The researcher identified the following limitations with the cross-sectional research design:

- When used with an unstructured interview, it is not easily replicated, and therefore may lack external reliability (Bryman, 2012)
- It only provides a snapshot of a context at a point in time.

Using a cross-sectional design has the following advantages:

- Takes limited time to conduct
- Is relatively inexpensive.

3.3 Population and sample

3.3.1 *Population*

Denzin (1970) states that triangulation entails the use of dissimilar comparison groups as data sources to investigate commonalities across settings. According to Thurmond (2001) triangulation caters for multiple perspectives, helps decrease bias and increases validity as well as interpretive potential of research. To give a more complete view of BDA and cyber fraud prevention in the banking sector, the researcher used triangulation of data sources. The sub-population for this research was made up of the following stakeholder groups:

1) The five largest South African banks by market capitalisation
2) Banking industry and advisory institutions
3) IT vendors of BDA tools.

The literature review revealed that while BDA in fraud prevention is relatively new, there have already been strides made with the implementation of BDA in social media analytics and in consumer profiling in South Africa. So to keep with the triangulation strategy, experts interviewed in the area of BDA were not limited to those operating in the fraud space but was extended to include BDA experts in social media and consumer profiling. The following table identifies the stakeholder groups that made up the sub-population.

Table 1: Profile of stakeholder groups

Stakeholder group	Area of expertise
Five largest South African banks by market capitalisation	• Fraud • Cyber security • BDA in cyber fraud prevention • BDA in social media • BDA in consumer profiling
Banking industry and advisory institutions	• Cyber fraud • Fraud analytics
IT vendors	• BDA in cyber fraud prevention • BDA in social media • BDA in consumer profiling

3.3.2 *Sample and sampling method*

From the sub-population a sample was drawn based on generic purposive sampling. According to Bryman (2012), when using generic purposive sampling, the researcher identifies appropriate contexts that address the research questions and then samples from those contexts. Since the use of BDA in fraud prevention is still relatively new to the world and more so to South Africa, it was decided to use a triangulation strategy as mentioned earlier. Therefore experts with BDA experience in the following contexts were included in the sample:

- Fraud prevention
- Social media analytics
- Consumer profiling.

The sample for this research included:

1) Banking professionals working at any of the five largest banks in South Africa, in the fields of:
    - Fraud
    - Cybersecurity
    - BDA in fraud, social media and consumer profiling.
2) IT experts in the field of BDA with experience in fraud, social media and consumer profiling. IT vendors included in the sample were chosen based on the applicability of their BDA tools to the banking industry.
3) Banking industry and advisory institutions dealing specifically with fraud-related activity within the banking sector.

The reason for using general purposive sampling was so that the researcher could sample interviewees in a strategic way. This allowed the researcher to ensure that those interviewed were relevant to the research questions while differing from one another in terms of key characteristics pertinent to the research questions (Bryman, 2012). The advantage of this was a varied and yet relevant sample. However, a disadvantage of using purposive sampling is that it is a non-probability sampling approach meaning that the results in this report cannot be extrapolated to the population (Bryman, 2012). Another drawback to purposive sampling is that there is subjectivity on the part of the researcher in selecting the sample. The researcher tried to overcome this by the above-mentioned triangulation strategy.

A sample of 12 participants was interviewed and the researcher found this number sufficient as data saturation had been achieved. According to Bryman (2012) there is little guidance to establish or recognise when saturation is achieved, however the researcher was sufficiently satisfied that no new information pertaining to the research questions was being revealed ie there was information redundancy. Table 2 profiles the 12 interviewees that make up the sample.

Table 2: Profile of interviewees

Company	Area of expertise	Sample
South African Banks		
Standard Bank	Cybersecurity	1
Standard Bank	BDA in social media	1
Investec	Cybersecurity & BDA in fraud prevention	1
Investec	Financial crime	1
Investec	Fraud	1
Nedbank	Cybersecurity & BDA in fraud prevention	1
FNB	BDA in consumer profiling	1
Banking industry and advisory institutions		
SABRIC	Cyber fraud	1
SABRIC	Fraud analytics	1
Wolfpack Information Risk	Cyber fraud	1
IT Vendors		
RSA	BDA in cybersecurity	1
Ubiquity	BDA in social media and consumer profiling	1
Total sample		12

3.4 The research instrument

The research instrument used was an unstructured interview. This instrument was chosen due to its flexibility and its assistance to the researcher during this exploratory research (Bryman, 2012). Through the unstructured interview the researcher had the opportunity to react to responses from the interviewees by either probing or changing direction if interest was sparked (Gideon, 2012). In a totally unstructured interview, there may be just one question that an interviewer asks and the interviewer may make use of an aide-mémoire, purely to prompt the him/her to cover certain points (Bryman, 2012). The researcher in this process used a totally unstructured interview with just one question.

The advantages of making use of an unstructured interview were:

- The interviews could be presented on short notice without extensive planning (Gideon, 2012)
- The flexibility of the interview allowed the researcher to react to what was happening in the field
- The conversational style of the interview allowed interviewees to speak without restriction thereby providing valuable insight to the researcher (Gideon, 2012).

Some of the shortcomings associated with this instrument were:

- The same question had to be asked in a similar way to all interviewees, if reliable quantitative analysis of results was to take place
- The unplanned nature of questions may have resulted in the researcher unwittingly introducing bias into the interview process.

In order to overcome the negatives, the researcher made use of an aide-mémoire (refer to Appendix A), to ensure that questions were asked in a similar way to all interviewees. The researcher also made use of triangulation of data sources in order to overcome bias introduced in the interview process.

3.5 Procedure for data collection

Interviews are so attractive due to the flexibility they offer and the richness of the information that emanates (Bryman, 2012). For this research, interviewees were identified through various professional networking platforms and initial contact was made via email. Once contact was made and a response was elicited the researcher sent further information regarding the research topic and participation requirements to the potential interviewee (refer to Appendix B). This was done via email, to give the potential interviewee an opportunity to peruse the information at leisure and to feel comfortable enough to decline participation, if desired. Once consent to participate was received via email, a face-to-face interview was arranged at the convenience of the interviewee. Interviews took place at the place of work of the interviewees.

Before the interview commenced written consent, signed and dated, was received from the interviewee (refer to Appendix B). The written consent included permission to voice record the interview. Each interview commenced with the participant being asked to articulate their current roles and responsibilities within their organisation. From there, the interviewee was asked one question (refer to Appendix B) in line with the unstructured interviewing technique. The researcher watched and listened attentively and responded to comments that sparked an interest and required further probing. The researcher went to great lengths to ensure that questions being asked were open-ended.

The researcher made use of an aide-mémoire to ensure questions were phrased in a similar manner across interviews. The researcher also took notes during the interview process in an attempt to capture details that the audio could not reveal such as body language and facial expression of interviewees. These details were necessary for the researcher to accurately capture the perspective of the interviewees (Bryman, 2012). Interviews lasted between 40 and 60 minutes.

After the interview process the recorded interviews were transcribed to produce documents as sources of data. This allowed for further examination of the interviewees responses (Bryman, 2012). Both the interview recordings and transcripts were treated with confidentiality and stored for safekeeping.

3.6 Data analysis and interpretation

Unlike with quantitative data analysis, rules for the analysis of qualitative data are generally scarce and often unknown to many. Qualitative data analysis lacks the unambiguous processing rules of quantitative data analysis, however there are still broad guidelines that may be used (Bryman, 2012). Based on these guidelines, the researcher made use of thematic analysis and the interview transcripts served as sources of data.

Attride-Stirling (2001) provides a technique for carrying out thematic analysis whereby the analysis makes use of thematic networks. Thematic analysis helps unearth the themes in the text and thematic networks help facilitate the

structuring and presentation of those themes (Attride-Stirling, 2001). The researcher chose this technique as it made transparent the steps involved in moving from text to interpretation. Also, the research instrument used was an unstructured interview and the thematic networks provided structure in interpreting the results. Themes identified were based on whether it captured something significant in relation to the research questions and were not necessarily based on quantifiable measures (Braun & Clarke, 2006).

Thematic networks provided structure in the following way:

1. Basic themes

   Text data from the transcripts were coded and given very low level themes, referred to as “basic themes”. However these basic themes offered little value on their own and needed to be viewed in conjunction with other basic themes (Attride-Stirling, 2001). This led to “organising themes”.

2. Organising themes

   These represented clusters of similar basic themes which were more revealing of the actual textual data (Attride-Stirling, 2001).

3. Global themes

   From organising themes the researcher produced “global themes” which can be described as final tenets. Global themes are groups of organising themes that present a position or assertion (Attride-Stirling, 2001). Global themes were then explored, described and interpreted.

Figure 8 below is a graphical representation of a thematic network.

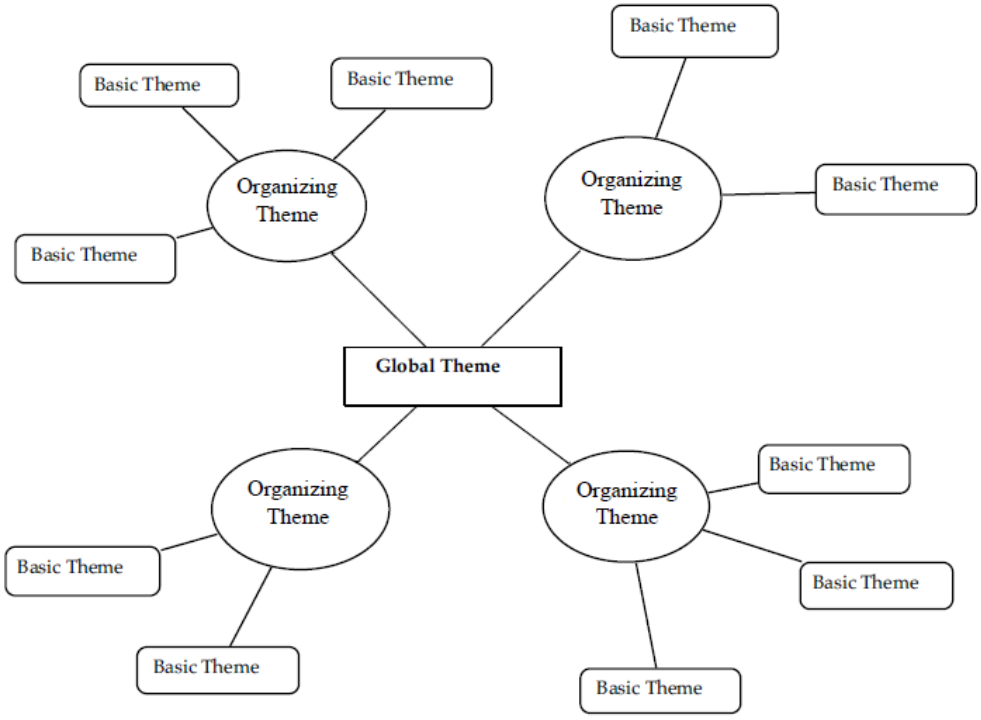

Figure 8: Structure of a thematic network

In accordance with Attride-Stirling (2001), the researcher made use of the following steps for conducting the thematic analysis and then creating the thematic networks:

- Step 1 – Coded the material

  Using a coding framework, the researcher dissected the text into segments

- Step 2 – Identified the themes

  Themes were generated from the coded text segments and these themes were further refined

- Step 3 – Constructed the thematic networks

  This was done by:

  - Arranging themes
  - Selecting basic themes
  - Arranging basic themes into organising themes
  - Deducing global themes

- Step 4 – Described and explored the networks
- Step 5 – Summarised the networks
- Step 6 – Interpreted the patterns.

3.7 Limitations of the study

- The chosen research method was solely dependent on gaining interviews with key experts
- The exploratory nature of the research meant it was largely dependent on participant feedback, perceptions and insight. Therefore providing insight on the research problem was largely dependent on the level of knowledge the interviewees had in BDA, which is a relatively new concept
- Cyber attacks and the related fraud have a massive reputational effect on a bank, so although no stats were requested, interviewees from the banks may have downplayed the threat of cyber attacks somewhat.

3.8 Validity and reliability

According to (Morse, Barrett, Mayan, Olson, & Spiers, 2008) research validity and reliability can be translated as being research transferability, credibility, dependability and confirmability. In order to address these issues, the researcher made use of verification strategies in accordance with Morse et al. (2008). Verification involves, checking, making sure, confirming and being certain. This assisted the researcher to identify when to proceed, stop or change the research process to achieve validity and reliability (Morse et al., 2008).

3.8.1 *External validity - transferability*

External validity refers to the extent to which findings can be extrapolated across social settings, ie transferability (Bryman, 2012). While this study focused specifically on the banking context, it dealt with a problem faced by multiple industries today, ie cyber fraud. This research may easily be extrapolated to include other sectors, especially within the financial services industry. It may also be valid to those sectors interested in the application of BDA in solving business problems.

3.8.2 *Internal validity - credibility*

Internal validity refers to whether the researcher is measuring what he/she says they will, ie credibility (Bryman, 2012). The research interviewees were purposively selected in order to address the research problem and were experts in the area. It is therefore expected that interviewing these participants accurately addressed the research problem. Interviews were audio recorded to ensure that the researcher drew legitimate conclusions from the data analysed.

3.8.3 *Reliability – dependability and confirmability*

According to (Bryman, 2012), external reliability refers to the extent to which a study can be replicated and internal reliability refers to consistency of observations, when there is more than one observer present. This was a cross-sectional study in the field of information technology with just one observer. It is highly improbable that this may be replicated in the future due to the following reasons:

- The researcher made use of an unstructured interview and purposive sampling
- Cyber threats and attacks are constantly evolving and what is relevant today may not be in the future
- The study involved the use of technology, which is also rapidly changing

- The research problem was relatively unexplored at this point in time, but there may well be established answers to the problem in the future, which changes the context.

In order to address this, the researcher can at best:

- Deliver a highly detailed report of findings
- Transcribe the interview process.

3.9 Demographic profile of interviewees

The researcher chose to use triangulation of data sources for purposes of this research. Based on that, three stakeholder groups were identified as being relevant to the research problem and they were:

1) The five largest South African banks by market capitalisation
2) Banking industry and advisory institutions dealing with fraud-related activity
3) IT vendors with expertise in BDA.

Interviewees within these three stakeholder groups were selected based on their ability to answer the research questions. The researcher judged suitability based on the interviewees current work profile and professional experience within the companies they work for.

Individuals in senior management and with a current work profile related to any of the following were identified as being suitable interviewees for purposes of this research:

- Cybersecurity
- Cyber fraud
- BDA in fraud
- BDA in social media
- BDA in consumer profiling.

Factors such as age, gender, race and educational background were of no relevance to the selection process. Table 3 differentiates between the planned and actual characteristics of the sample. For purposes of anonymity, interviewees' job titles were not used, instead interviewees were profiled by their area of expertise only.

Table 3: Planned vs Actual sample for the five largest SA banks

	Planned		Actual	
Banks	**Area of Expertise**	**Sample**	**Area of Expertise**	**Sample**
Absa	Cyber fraud	1		0
FNB	BDA	1	Analytics	1
Investec	Cybersecurity	1	Cybersecurity & BDA	1
	BDA	1	Forensics	1
			Fraud	1
Nedbank	Cyber fraud	1	Cyber fraud & BDA	1
	BDA	1		
Standard Bank	Cybersecurity	1	Financial crime control	1
	BDA	1	BDA	1

The planned sample from the banks differed slightly to the actual sample. One reason for this is that the researcher did not have prior information as to how the areas of expertise identified as being suitable for this research were defined into job profiles at the banks. For example at Nedbank, the researcher initially was under the assumption that BDA in fraud prevention and cyber fraud expertise may reside in two different job profiles. However this was not the case and interviewing a single respondent at Nedbank covered both areas of expertise. The same applied to Investec.

Two additional interviewees currently working at Investec were identified as being relevant due to their experience, at an industry level in the fields of fraud and forensics. After 12 interviewees, it was decided that interviewing a respondent at Absa was not required as data saturation was achieved.

Table 4 differentiates between the planned and actual characteristics of the sample for banking industry and advisory institutions.

Table 4: Planned vs actual sample for banking industry and advisory institutions

	Planned		Actual	
Banking industry and advisory institutions	**Area of Expertise**	**Sample**	**Area of Expertise**	**Sample**
SABRIC	Cyber fraud	1	Cyber fraud	1
	Fraud analytics	1	Fraud analytics	1
Wolfpack Information Risk			Cyber fraud	1

With regard to the banking industry and advisory institutions, SABRIC was initially identified to be interviewed and that remained unchanged. In addition, Wolfpack Information Risk, who has done a lot of advisory work in the sector, was identified as being a worthy respondent.

Table 5 differentiates between the planned and actual characteristics of the sample for IT vendors.

Table 5: Planned vs actual sample for IT vendors

	Planned		**Actual**	
IT vendors	**Area of Expertise**	**Sample**	**Area of Expertise**	**Sample**
RSA	BDA in fraud prevention	1	BDA in fraud prevention	1
Ubiquity	BDA in consumer profiling &social media	1	BDA in consumer profiling &social media	1

The planned and actual sample of the IT vendors remained unchanged.

CHAPTER 4. PRESENTATION OF RESULTS

4.1 Introduction

The research aim was to identify how the use of BDA to combat cyber fraud is perceived within the SA banking sector. An unstructured interview was used to probe and gain insight from each of the interviewees and this section presents those insights in the form of findings to each of the research questions. Results to each of the questions are presented in the form of thematic networks. These networks illustrate all the themes that emerged through the thematic analysis. The researcher also identified the themes that emerged prominently across respondents for each of the questions. This provides some guidance into where there's strong sentiment exhibited in relation to the questions. Due to the triangulation strategy used, not all questions were relevant to all interviewees and the results are presented accordingly.

4.2 Results pertaining to Research Question 1

The first research question was how are SA banks faring in the fight against cyber fraud? Responses for this question were received from across the three stakeholder groups. Due to the triangulation strategy used in the sample selection process, not all interviewees in the sample were sufficiently versed in the area of cyber fraud. Therefore 25 per cent of interviewees chose not to discuss cyber fraud as part of their interview process.

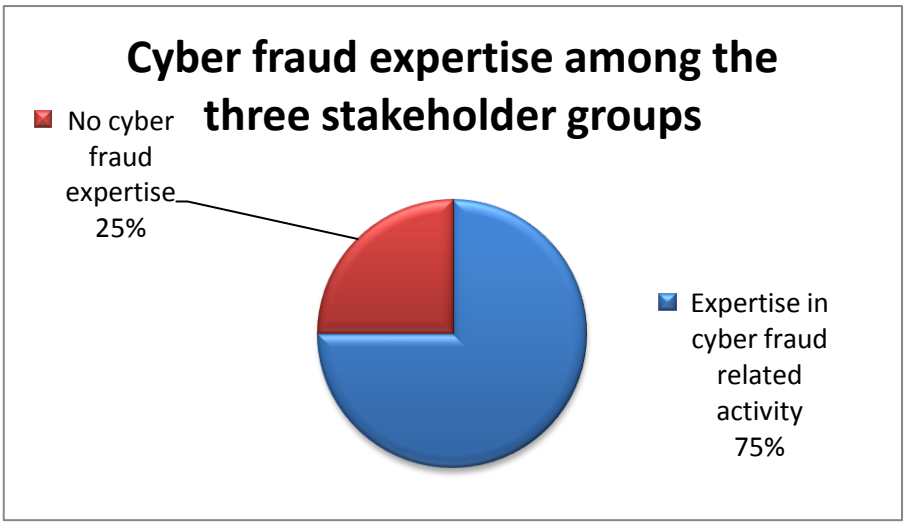

Figure 9: Cyber fraud expertise among the three stakeholder groups

The 75 per cent of interviewees who shared perceptions on cyber fraud make up 100 per cent of interviewees for this research question. The following table includes the global and organising themes that emerged.

Table 6: Global and organising themes for Research Question 1

Global themes	Organising themes
Current cyber fraud activity at SA banks	• Stats not very high • Legal issues • Globalisation • High risk • Fight requires collaborated effort • Less technologically advanced
Cyber threats	• Malware • Constantly evolving • Phishing • Consumers are most targeted
Cyber fraud activity in the future	• Banks starting to challenge payouts • Constantly evolving • Cybercrime better legislated • Will require more spend from banks • Biggest concerns

The next sections look at the results for each of the global themes deduced.

4.2.1 *Current cyber fraud activity at SA banks*

This section includes responses from interviewees regarding the current issues the banking sector faces in terms of cyber fraud activity. The top five themes that emerged appear in the following table:

Table 7: Top five themes for current cyber fraud activity at SA banks

Top five basic themes	Mentioned by
Not properly legislated (Cybercrime)	67% of interviewees
Risk of data theft	67% of interviewees
Work together as an industry	56% of interviewees
Losses don't always materialise in cyber attack itself	56% of interviewees
Risk of merchants being breached	44% of interviewees

Figure 10 provides an overview of all the themes identified:

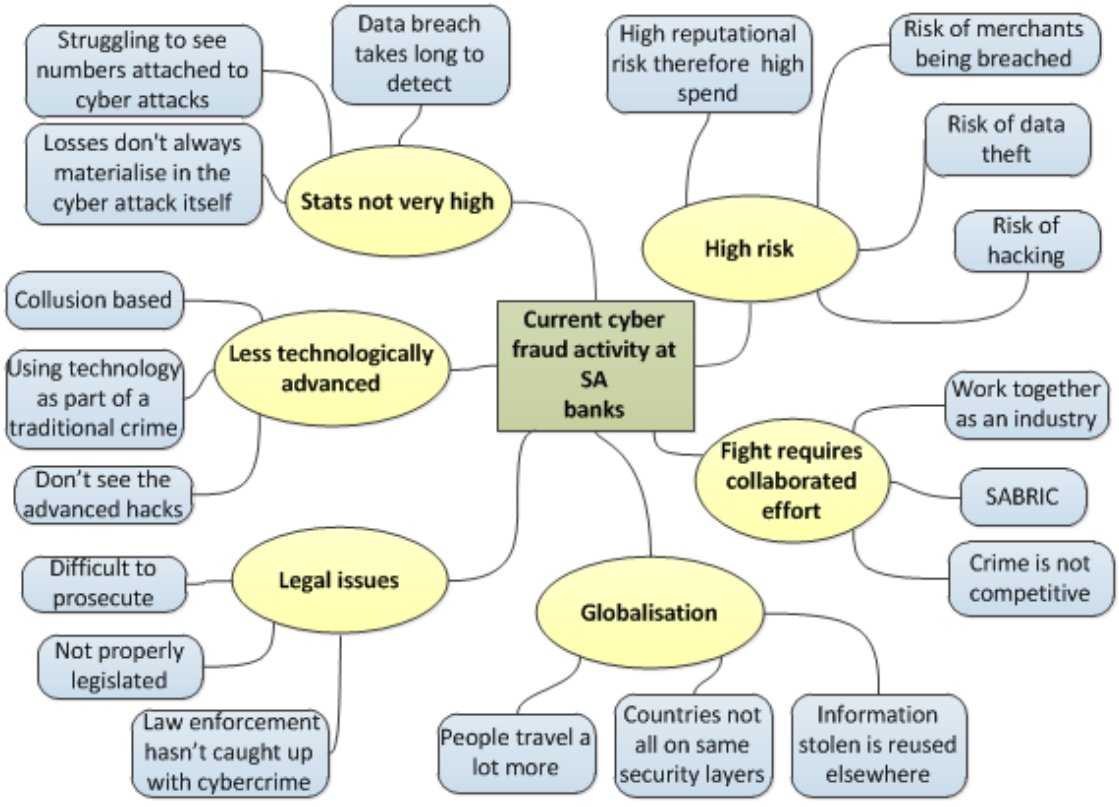

Figure 10: Current cyber fraud activity at SA banks

4.2.2 *Cyber threats*

One of the global themes that emerged referred to the cyber threats faced by the industry. The top five basic themes relating to cyber threats appear in the following table:

Table 8: Top five themes for cyber threats

Top five themes	Mentioned by
Big concern (Malware)	78% of interviewees
Prevalent among the big four banks (Phishing)	67% of interviewees
Cyclical (Phishing)	67% of interviewees
User behavior must change	67% of interviewees

Figure 11 provides an overview of all the themes identified:

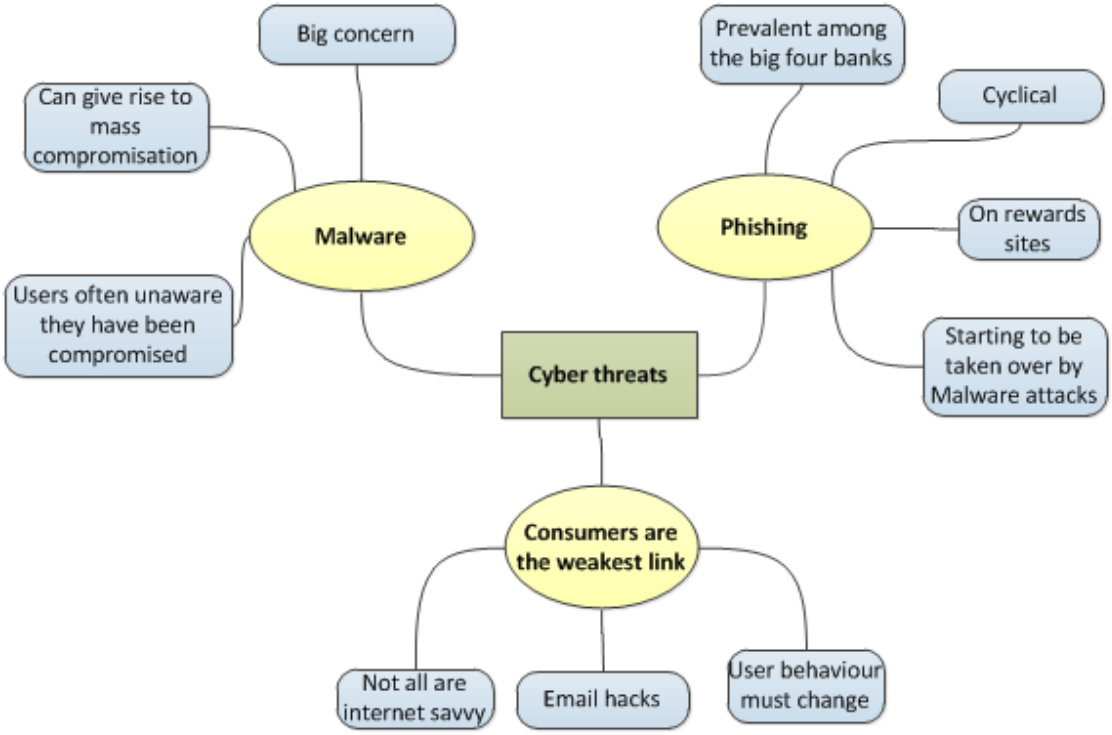

Figure 11: Cyber threats

4.2.3 *Cyber fraud activity in the future*

Interviewees went on to discuss what they perceive to happen in the future with regard to cyber fraud activity. Five organising themes emerged and the top five basic themes are displayed in the table below:

Table 9: Top five themes for cyber fraud activity in the future

Top five basic themes	Mentioned by
Cyber attacks constantly evolving	100% of interviewees
Don't know where the next threat is	100% of interviewees
Client data stolen or leaked	67% of interviewees
Mass compromisation	56% of interviewees
Cost to banks increasing	44% of interviewees

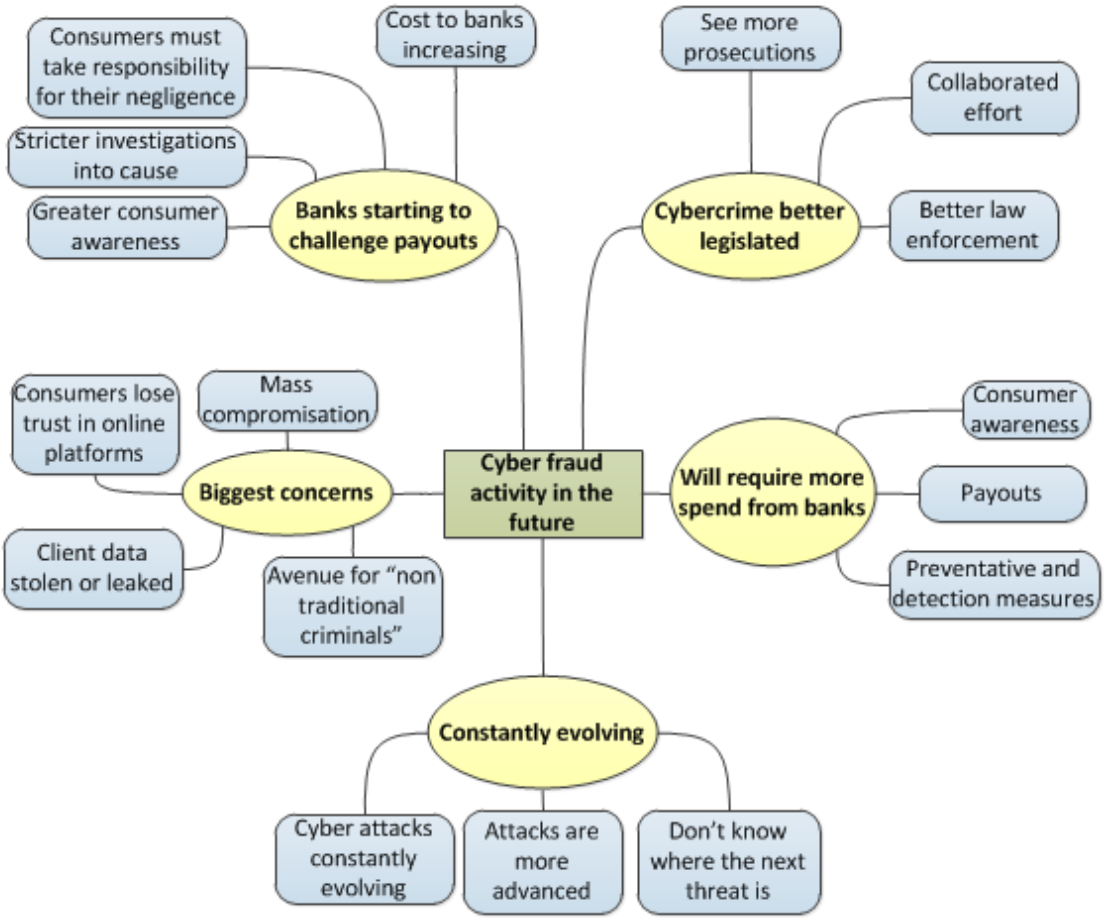

Figure 12: Cyber fraud activity in the future

4.3 Results pertaining to Research Question 2

The second research question was how is big data perceived within the SA banking sector? Since this question was specific to the banks, it only related to one stakeholder group, ie the five largest banks in SA. As previously mentioned, interviewees were selected for their expertise in the areas of BDA, cyber fraud and cyber security. Therefore while 100 per cent of interviewees from the banks were familiar with the concept of big data, not all were sufficiently versed in the area in order to share their perceptions. Fourteen per cent of interviewees chose not to discuss big data and big data analytics as part of their interview process.

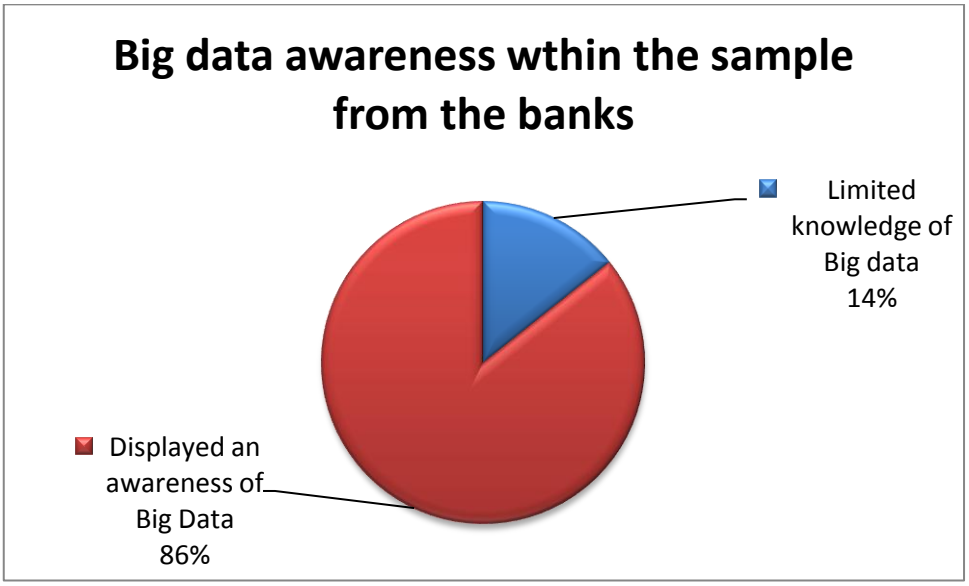

Figure 13: Big data awareness within the sample from the banks

Therefore 86 per cent of interviewees discussed their perceptions around big data. Based on their responses 16 basic themes were identified. The themes that were most commonly identified across interviewees are seen in table 10. The percentage is calculated by considering only those that shared perceptions regarding big data, therefore in the following table the 86 per cent of interviewees mentioned above, make up 100 per cent of interviewees below.

Table 10: Top five themes for big data awareness

Top five basic themes	Mentioned by
Makes sense of unstructured data	80% of interviewees
Requires non-traditional technology	80% of interviewees
Buzzword	80% of interviewees
Not a new concept	60% of interviewees
Various definitions	60% of interviewees

The researcher then grouped the 16 basic themes into four organising themes that better explained the underlying themes. The global theme was then deduced as being “Big data awareness”.

The following thematic network displays all the basic, organising and global themes identified through the thematic content analysis, for this research question.

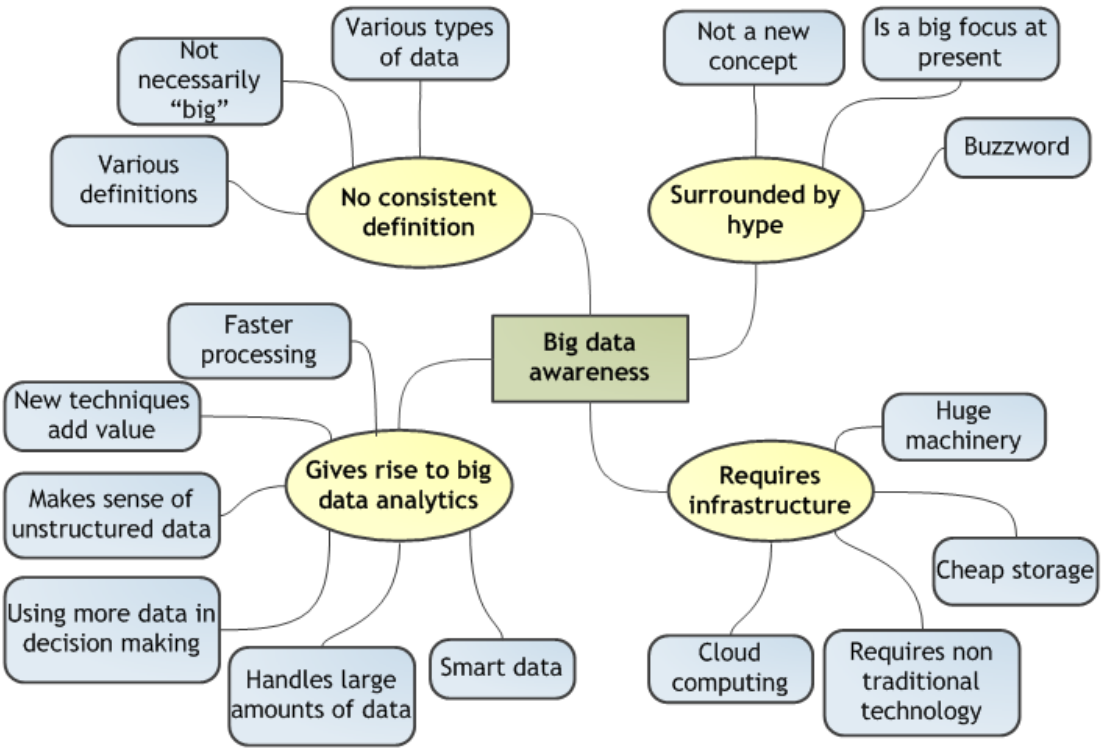

Figure 14: Big data awareness

4.4 Results pertaining to Research Question 3

The third research question was how is the use of big data analytics perceived within the SA banking sector? Responses associated with this research question were received from just one stakeholder group ie the big five banks. Once again as with the previous research question, only 86 per cent of interviewees within the bank provided responses on BDA. The 86 per cent makes up 100 per cent of interviewees for purposes of the results presented for this research question.

Five major areas of interest (global themes) emerged during the interviews. These are listed in table 11 below, together with the organising themes that they are each divided into.

Table 11: Global and organising themes for Research Question 3

Global themes	Organising themes
Current BDA activity within the banks	• Social media analysis • Consumer analysis • Fraud prevention • Cybersecurity
Reasons to implement	• Competitive advantage • Not just hype • Return on investment • Traditional technologies are limiting
Implementation challenges	• Skills shortage • Culture • Infrastructure layout • Starting from scratch
Overcoming implementation challenges	• IT vendors • Proof of concept projects • On the job learning • Infrastructure
Future BDA activity within the banks	• Social media analysis • Consumer analysis • Fraud prevention • Cybersecurity

The next sections looks at the results for each of the global themes deduced.

4.4.1 *Current BDA activity within the banks*

When looking at the current use of BDA, 100 per cent of interviewees indicated that there was some sort of BDA activity within the banks they worked for. From the responses received, 16 basic themes were identified. These basic themes

related to four distinct business areas. Interviewees each spoke about their area of expertise and therefore the researcher could not rate the themes according to prevalence, since unique themes emerged within each area. The theme network below provides an overview of the themes that emerged during the interviews.

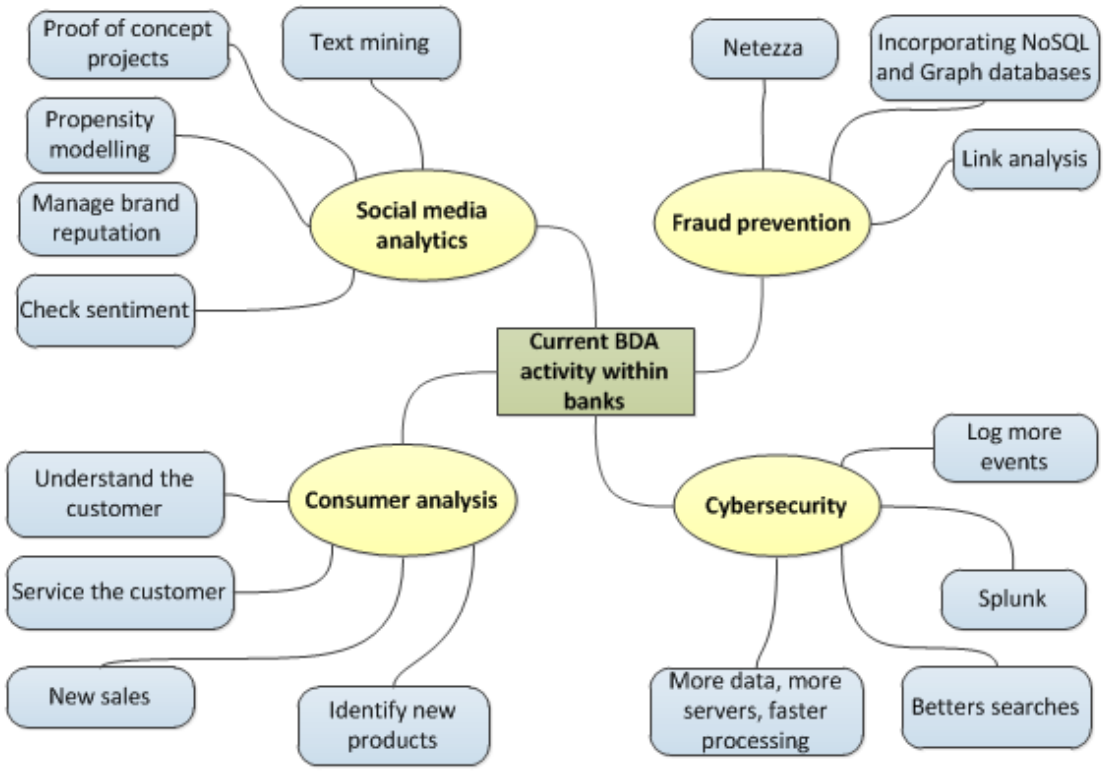

Figure 15: Current BDA activity within banks

4.4.2 *Reasons to implement*

This section includes responses from interviewees around the motivating factors to implement BDA within banks. Since interviewees already had experience with BDA, the responses received were based on what interviewees were aware of based on experience. Responses received were generic and based on BDA capabilities. Fourteen basic themes emerged and the following table shows the common themes identified.

Table 12: Top five themes for reasons to implement

Top five basic themes	Mentioned by
Processing power insufficient (traditional technologies)	100% of interviewees
Banks have experienced ROI on implementations	80% of interviewees
Banks have experienced successful implementations	80% of interviewees
Increased market share	40% of interviewees
Better service clients	40% of interviewees

The theme network below provides more details around the themes that emerged.

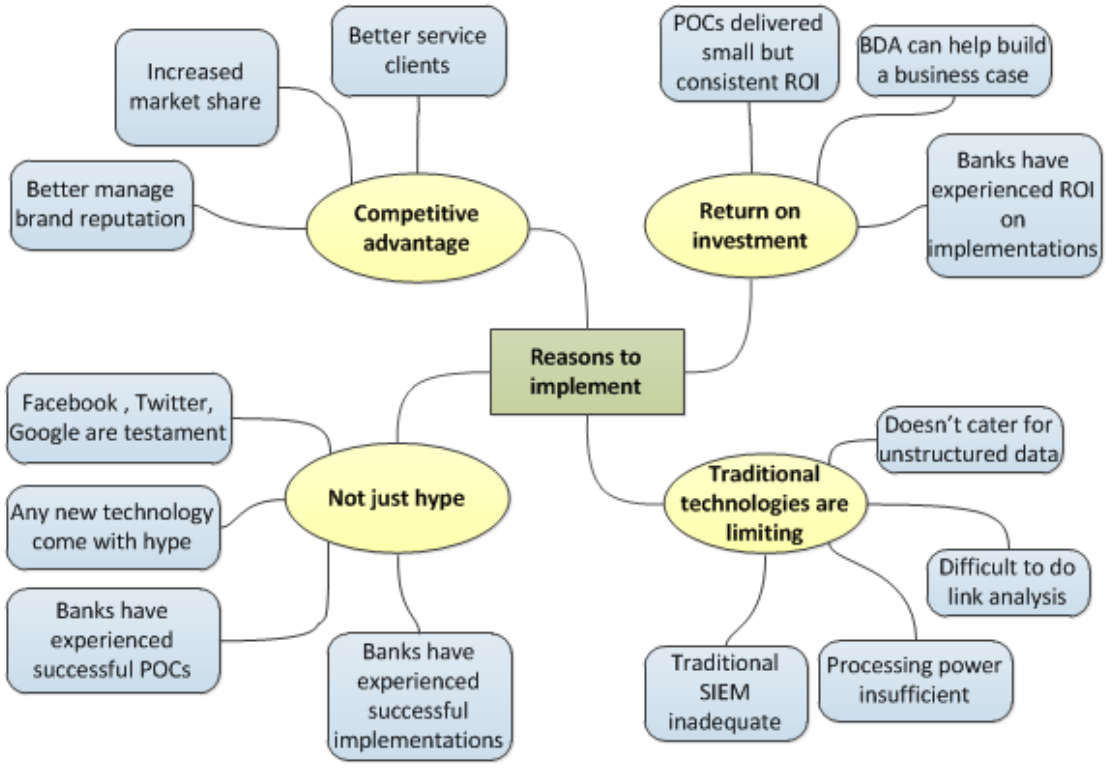

Figure 16: Reasons to implement

4.4.3 *Implementation challenges*

In this section interviewees shared information around what they perceived to be the barriers to BDA implementations within banks. Responses were generic and not specific to business units. Here the most common basic theme

resonated by all interviewees was skills shortages. The table below lists the most identified themes by interviewees.

Table 13: Top five themes for implementation challenges

Top five basic themes	Mentioned by
Need different skills in one individual	100% of interviewees
Change is slow at the big banks	60% of interviewees
Quantitative skills are limited	40% of interviewees
Expensive to start up	40% of interviewees
Return on investment may be slow	40% of interviewees

The rest of the themes identified appear in the network below:

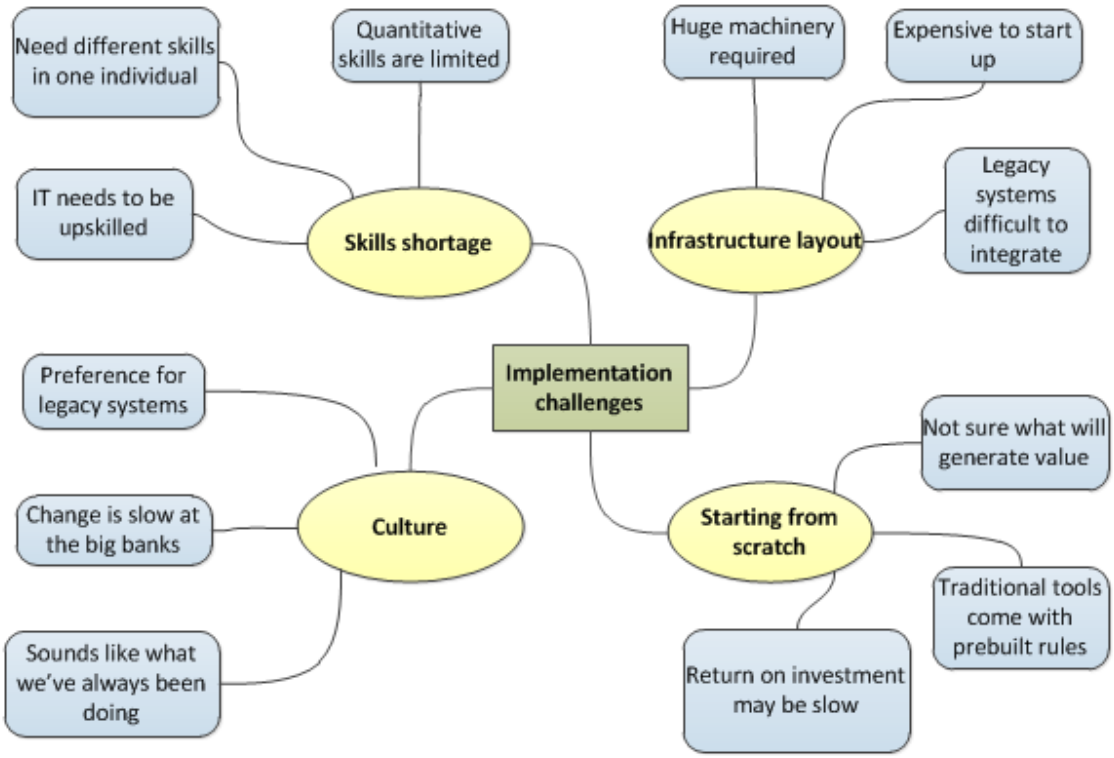

Figure 17: Implementation challenges

4.4.4 *Overcoming implementation challenges*

After identifying implementation challenges, interviewees shared perceptions on how these challenges could be overcome and this section presents the results. The top three themes that emerged appear in the following table:

Table 14: Top three themes for overcoming implementation challenges

Top three basic themes	Mentioned by
Grow talent	100% of interviewees
Start small	60% of interviewees
Experiment	60% of interviewees

The following theme network provides the themes mentioned in more detail:

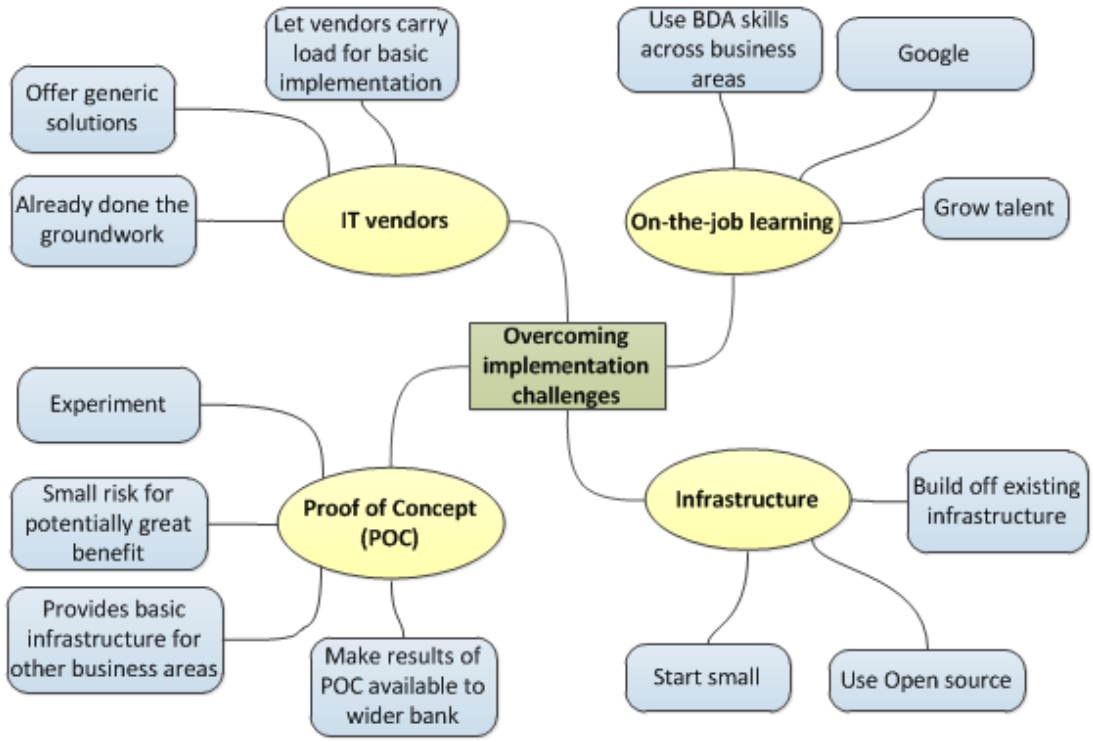

Figure 18: Overcoming implementation challenges

4.4.5 *Future BDA activity within the banks*

This section discusses where the future of BDA lies within banks. Interviewees provided responses in relation to their areas of expertise. Therefore the grouping of themes naturally emerged as business areas. The researcher was unable to identify the most prevalent themes as themes identified were generally unique. However it did emerge that interviewees felt BDA for social media could be leveraged across business areas. The following theme network reflects this and the other themes that emerged.

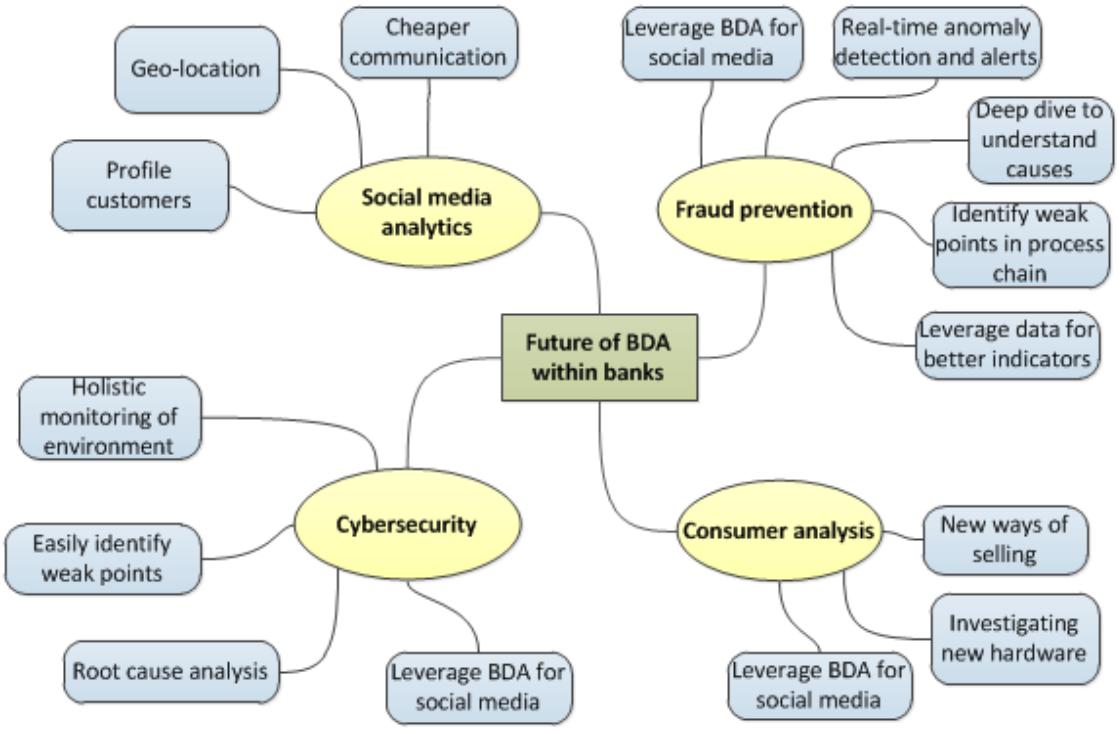

Figure 19: Future of BDA within the banks

4.5 Results pertaining to Research Question 4

The fourth research question was how can SA banks use big data analytics in the fight against cyber fraud? This was addressed in part through each of the other research questions as well as directly through this question. Responses for this question were received from all three stakeholder groups. Since this question ties together the different areas of expertise of those in the sample, interviewees were able to contribute and share perceptions in some form or

another. Therefore 100 per cent of the sample contributed responses to this question.

4.5.1 *Use of BDA for cyber fraud prevention*

The global and organising themes that emerged are represented in the table below.

Table 15: Global and organising themes for Research Question 4

Global themes	Organising themes
Use of BDA for cyber fraud prevention	• Has been introduced to the banks for fraud prevention • BDA capabilities for fraud prevention not thoroughly explored • Go to experts • Eager to look at what's to come

The following theme network shows all the themes that emerged.

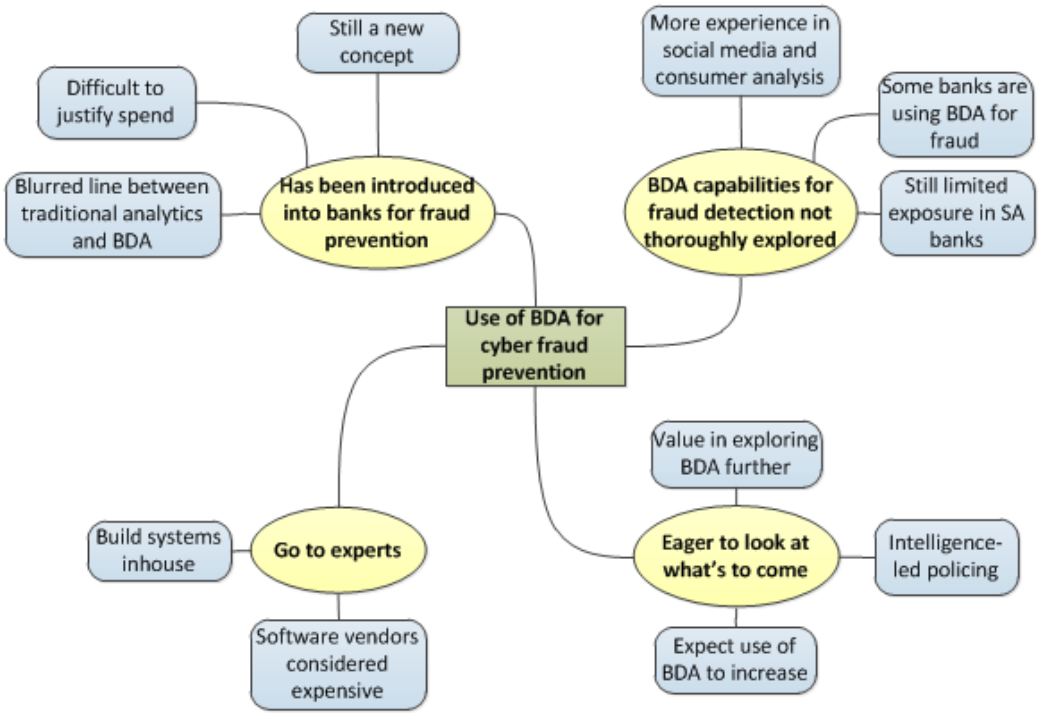

Figure 20: Use of BDA for cyber fraud prevention

4.6 Summary of the results

As mentioned in the introduction to this chapter, the research aim was to identify how the use of big data analytics to combat cyber fraud is perceived within the SA banking sector. A triangulation strategy was used and interviewees provided perceptions and insights based on their areas of expertise. This meant that not all research questions were relevant to all interviewees. Interviewees chose to refrain from commenting on topics outside their areas of expertise. The triangulation strategy resulted in a comprehensive look at BDA and cyber fraud within the banking sector as well as the use of BDA in cyber fraud prevention. A total of 10 global themes, 42 organising themes and 143 basic themes emerged. There were 12 prominent basic themes that were each mentioned by more than 75 per cent of interviewees.

Each of the following themes around BDA was mentioned by more than 75 per cent of interviewees:

1. Processing power insufficient (traditional technologies)
2. Need different skills in one individual
3. Grow talent
4. Makes sense of unstructured data
5. Requires non-traditional technology
6. Buzzword
7. Banks have experienced ROI on implementations
8. Banks have experienced successful implementations.

Each of the following themes around cyber fraud was mentioned by more than 75 per cent of interviewees:

1. Always coming up with new attacks
2. Cyber activity constantly evolving
3. Don’t know where the next threat is
4. Big concern (Malware).

The next chapter discusses the results that were presented in this chapter.

CHAPTER 5. DISCUSSION OF THE RESULTS

5.1 Introduction

In this chapter, the results presented in chapter 4, will be discussed and explained. The demographics are not material to the research problems and were already discussed as part of the Research Methodology chapter. The discussion of results will follow the same layout as the presentation of results in the previous chapter, ie per research question.

5.2 Discussion pertaining to Research Question 1

The first research question was how are SA banks faring in the fight against cyber fraud? The purpose of this question was to understand what are the biggest challenges and threats that banks face with regard to cyber fraud. Also to ascertain what steps banks are taking to protect themselves against these vicious attacks. Interviewees also discussed cyber attacks since these are the precursors of a cyber fraud case. Three global themes emerged with numerous basic themes. The key themes will be discussed in more detail below.

5.2.1 *Current cyber fraud activity at SA banks*

Six major themes emerged around current cyber fraud activity at the banks. These are discussed below.

a) Fight requires collaborated effort

All stakeholder groups agreed that the fight against cyber fraud requires a combined effort with a lot of information sharing. Fifty six per cent of interviewees mentioned that banks as an industry need to work together and that SABRIC has been an enabler of that. Through discussions, it emerged that SABRIC initially started 13 years ago for the main purpose of sharing information on an industry level. This in accordance with Paganini (2013) who mentioned that to be successful in the fight

against cyber attacks banks, governance bodies and vendors of technology must work together to build defence mechanisms.

b) Stats not very high

Interviewees generally mentioned that the stats around cyber fraud aren't very high. Fifty six per cent of interviewees mentioned that the losses don't always materialise in the cyber attack itself and that data breaches often take a long time to detect. Therefore when the losses do arise, it is sometimes difficult to link it back to a specific cyber attack. However the literature, contrary to the view expressed by interviewees, suggested that stats around cyber fraud, especially account takeover fraud are increasing. It may be that due to reputational risk, interviewees were cautious of discussing actual cyber fraud stats and the researcher chose not to probe this aspect further.

c) Legal issues

A theme that came across strongly across stakeholder groups, was the legalities around prosecution of cybercrimes and more specifically cyber fraud. Sixty seven per cent of interviewees mentioned that lack of appropriate legislation made it difficult to prosecute cybercrimes. Interviewees felt that although banks take the necessary steps to identify and alert police to fraudsters, the legal system seems to be failing at the prosecution of these criminals. Banks are finding that fraudsters are apprehended, released and then go back to doing the very same activity that they were initially apprehended for.

Therefore the sentiment among interviewees was that law enforcement has not caught up with cybercrime. One interviewee said: *"For the South African police to go cross border and arrest somebody takes months. For a criminal to get into his car and drive across a border takes a couple of hours."* Interviewees from the banks did however place a lot of confidence in the work SABRIC is doing together with the police to apprehend criminals. SABRIC's operational division is responsible for liaising with banks and the police and assisting them both when it comes

to organised crime matters. SABRIC makes use of information gathered on an industry level to guide them during their investigations.

d) Globalisation

With the advent of globalisation, we now have global communities. Interviewees mentioned that with people travelling a lot more, data breaches have increased exponentially and have become increasingly difficult to track. Information gets stolen in one country and gets reused elsewhere. Interviewees mentioned that this is a problem as not all countries are on the same security layers. One example quoted: *"You know in South Africa we have the cards that are chip enabled, 3D secure and all those types of things. In the States, your card is not EMV Compliant. So I basically just need your details, I don't need the actual card. I can make my own card with your details and then use it."*

This interviewee mentioned that banks are seeing information being stolen and then reused in markets where the method of getting money out is not secured. This makes prevention and prosecution difficult and is in accordance with Gilman, Goldhammer, and Weber (2011) who stated that through deviant globalisation individuals utilise technical infrastructure derived from globalisation for exploitative means.

e) High risk

Interviewees mentioned that cyber fraud poses great risks to banks. It has massive reputational risk associated with it as it can cause clients to lose trust in a bank's online platforms. Interviewees also spoke about cyber attacks and the constant risk banks face of having their systems hacked. Sixty seven per cent of interviewees mentioned that a major risk in the cyber world was the risk of data theft since losing client information is seen as the ultimate breach within a bank.

Forty four per cent of interviewees mentioned they also have to contend with risks outside of the bank that have a major impact, eg the risk of merchants being breached. The stolen data is ultimately used in a

fraudulent attack against the bank. Sony, Anthem and Target were all quoted as examples of this. This is similar to the 2013 incident, mentioned in the literature review, where malware was inserted into point-of-sale (POS) devices at certain fast-food outlets in SA (Mcleod, 2013). Interviewees mentioned that banks rely on SABRIC's risk information division to make banks aware of what is happening on an industry level.

f) Less technologically advanced

It was revealed that cyber attacks in South Africa are more collusion based and criminal syndicates approach individuals for assistance in the attack. They are less technologically advanced than in other countries. One interviewee said the following: *"So it's just using technology as part of a traditional crime as opposed to some of the more advanced hacks like what you're just seeing the Kapersky report is bringing out where these guys have been hidden in hundreds of banks for over a year and half now. Where they've been watching and slowly infiltrating the networks and stealing information that they've used now to cyphen off hundreds of millions of dollars. So we're not seeing that degree of technical sophistication on a mass scale. It happens here and there".*

However there was agreement that in the future we can expect to see some of the more sophisticated attacks as mentioned by Sood and Enbody (2013).

5.2.2 *Cyber threats*

Cyber fraud is accomplished through cyber attacks. Therefore when discussing cyber fraud, interviewees unsurprisingly spent a lot of time discussing cyber threats and cyber attacks. Four major themes emerged and these are discussed below.

a) Phishing

Interviewees concurred that phishing is the most prevalent cyber attack against the big four banks with 67 per cent of interviewees mentioning that attacks are generally cyclical in nature. This is in accordance with the literature where the Symantec Intelligence Report for 2012 identified South Africa as the second-most phished country globally (eNCA, 2013). Interviewees mentioned that perpetrators generally focus on one bank at a time and as soon as that bank steps up their security they move onto another bank only to return to the initial bank a few months later. They felt that the only way to reduce successful phishing attacks was through educating consumers thereby changing user behaviour.

Interviewees confirmed that phishing has for a period of years been prevalent within the banking sector and that attacks are not only directed at the online banking site but at the banks' rewards sites as well. They did also mention that as attacks evolve, they are now starting to see an increase in malware attacks and that very soon these attacks may become more prevalent than phishing.

b) Consumers are the weakest link

All three stakeholder groups strongly felt that the weakest link in cyber fraud attacks are the consumers. One interviewee felt the reason for this may be attributed to how internet savvy the South African population are. He made the following comment: *"I think an interesting twist on it would be to look at the different maturity levels of the internet in different countries. In my mind there's going to be a strong correlation between them. If your client base is internet savvy, they are less likely to fall for these types of incidents. Whereas in South Africa I think the majority of customers are not as internet savvy as you want and hence they are more susceptible to these things".*

A further concern mentioned was email hacks. Perpetrators hack the web-based email accounts of customers and then attempt to retrieve personal information from these accounts. The fraudsters also takeover

these accounts and contact banks posing as the customer in an attempt to fraudulently access banking services. Once again, the cause for this was attributed to lack of awareness on the part of consumers, where they make use of weak passwords and use the same password across applications. Sixty seven per cent of interviewees felt that user behavior must change.

c) Malware

There is a general concern across the sector with regard to malware attacks with 78 per cent of interviewees identifying this as their biggest concern when it comes to cyber attacks. This was cited as a big concern since users may become affected without any careless behavior ie these attacks do not require users to click on a link or enter special information. Users are compromised and are often completely unaware of it. Banks are concerned about these attacks as they have the potential to lead to mass compromisation of bank systems.

5.2.3 *Cyber fraud activity in the future*

In this section interviewees revealed what they perceived cyber fraud activity within the banks to look like in the future.

a) Banks starting to challenge payouts

According to some interviewees, banks are starting to lose large sums of money to cyber fraud. Access to social media platforms means that it is easy for begrudged customers to slander banks and cause great reputational damage. Therefore, although many cyber fraud cases are due to reckless behavior on the part of bank customers, banks often make refunds to clients to avoid reputational damage. However, according to interviewees this is changing.

While in some countries the law stipulates that banks need to refund customers, there is no such law in South Africa and the banks are starting to challenge these payouts. According to interviewees, the

process of investigating these crimes and proving that the client was at fault is onerous and expensive, but banks also are unable to blindly make payouts to clients. The expectation among interviewees is that in the future banks will be far more stringent with payouts.

b) Constantly evolving

The strongest sentiment that emerged around cyber attacks was that they are constantly evolving and that completely eradicating these attacks is therefore next to impossible. One hundred per cent of interviewees across stakeholder groups felt this way. It also emerged that SA banks are starting to see DDoS attacks. These attacks, although they affected the rest of the world, were previously unheard of in SA.

According to Cárdenas et al. (2010) DDoS attacks are often used for extortion, blackmail and even to express strong radical views. When probing one of the interviews from the advisory group around the DDoS attacks, he said the following: "*From a threat environment, we don't have as many political enemies as the US, Israel or China for example. So it could also even be they are just testing our sites before attacking other countries as well. So you never know... it's difficult to actually prove where the attacks actually came from and what the motive was, unless you have inside information."*

c) Cybercrime better legislated

One of the interviewees in the advisory stakeholder group explained that although government has thus far been slow to react to the threat of cybercrime, this is expected to change going forward. According to him, the importance is seen at middle management level and there's action being taken but what we need to see in the future is something being done at a strategic level. He had the following to say: *"You'll see Obama in the state of the nation talking extensively about this subject to the country. Here it's not even mentioned. So at a minister level, it's not perceived to be as high a threat as it actually is. So it's top management*

that needs to do something about it". He did however feel that this will receive attention in the future.

d) Will require more spend from banks

According to an interviewee in the banking industry and advisory stakeholder group, there is an enormous amount of money that goes into dealing with cyber fraud activity in the banking sector. There are exorbitant costs that go into prevention and detection, consumer awareness, investigating crimes and then finally payouts. For instance an institution like SABRIC, which was set up by the banks, exists solely for preventing and investigating fraud. Apart from this each bank has their own fraud and forensics teams which, according to the interviewees, are made up of thousands of people. The interviewees felt that the consumers are unaware of what the banks spend on cyber fraud and continue to expect more, which is unsustainable for the future.

e) Biggest concerns

The biggest concerns mentioned among interviewees from the banks were that the evolving attacks could ultimately lead to mass compromisation of systems and that client data may be stolen or leaked. One interviewee mentioned that currently within the big four banks, a large portion of their clients do not make use of any online banking at all and this translates into more ATMs and brick and mortar. He felt for as long as these problems or perceptions are out there about the safety of online banking, the channel will never achieve its full potential.

Another concern that emerged was that cyber fraud is attractive not only to criminals but also to those who normally would not engage in criminal activity. It is considered anonymous and is an avenue for individuals in financial difficulty to access money. This is in accordance with the fraud triangle and the factors that motivate fraudulent behaviour (Cressey, 1953). The concern is that the anonymity of the online platform will perpetuate fraudulent behaviour among a growing number of people.

Within the advisory stakeholder group, there was again the concern that banks customers may lose complete faith in the online platforms if cyber fraud activity continues. Legislation was also a concern with an interviewee mentioning that government should have recognised cybercrime as a threat eight to ten years ago. He mentioned that even though government is trying to get traction now, they are still way behind the curve. He felt that although SABRIC and the banks are doing a lot, the sector still requires government support and SA is very vulnerable.

5.2.4 *Conclusion*

There was consensus that cyber threats are rife and constantly evolving. They present a continued threat to the online platforms and have the potential to completely shut them down. It was mentioned that because losses don't always materialise in the cyber attack itself, it is difficult to tie a fraudulent case back to an attack. The researcher was therefore unable to accurately investigate the correlation of cyber attacks and account takeover fraud. Costs associated with cyber fraud are exorbitant and only expected to increase in the future. Banks are therefore starting to take a hard line when it comes to payouts.

Although it seems like cyber fraud activity within the sector is far from over, there were also many positives that emerged through discussions. There is an expectation that cybercrime will be better legislated and therefore better investigated and prosecuted. Also, the work that SABRIC is doing has been commended within the sector. It emerged that the sector has one of the largest security communities, which includes independent advisors. SABRIC's model is unique and is the first globally. There are individuals from various countries coming to SA to study SABRIC's model, some of which include Nigeria, Namibia, Australia, China and England. This is testament that the model works.

However, starting a body like SABRIC, requires a certain amount of maturity from the banks which not all countries have. In some countries, banks use cyber fraud stats of a competitive bank to lure away customers, this does not happen in SA. The SA banks came together and started SABRIC because they

agreed that crime will never be competitive. SABRIC also has good cross-border relations, especially within Africa which helps with investigation of crimes. It seems that cyber fraud will be here for some time to come. There are grave concerns but there is also good progress being made.

5.3 Discussion pertaining to Research Question 2

The second research question was how is big data perceived within the SA banking sector? The purpose of this question was to confirm whether there was a common understanding of the concept of big data within the banking sector. Since an unstructured interview was used, interviewees were asked to generally discuss what big data meant to each of them. Interviewees were allowed to speak freely and were not constrained or limited in any way. Based on their responses, certain key themes emerged which are discussed below.

5.3.1 *No consistent definition*

According to the literature review, there is no consistent definition of big data. There's the widely accepted definition which refers to the three V's, volume, variety and velocity (Gartner, 2013). However according to Zikopoulos et al. (2012) this definition is constantly evolving with even IBM adding an additional V for veracity. In accordance with the literature, 60 per cent of interviewees mentioned that there were various definitions of big data.

Interviewees felt that often there's confusion around the "big" in big data, but that people are becoming more aware that it does not only refer to large volumes of data. Interviewees felt that having no consistent definition caused confusion as to whether big data was anything new or if it was the same thing they've always done, just rebranded. Sixty per cent of interviewees felt big data is not a new concept. This is in accordance with the literature where Rider (1944) mentioned the concept decades ago even though the term big data was coined more recently. However in contrast to what some interviewees felt, the literature suggests big data is still very different to what data analysts have

always done in that it caters for unstructured data processing (Bollier & Firestone, 2010).

5.3.2 *Surrounded by hype*

Some interviewees felt that big data has been getting a lot of attention in recent years and is a big focus at present. Eighty per cent of interviewees stated big data is a buzzword at the moment. Interviewees felt that much of the hype is because it's still relatively new and not thoroughly explored. The general perception was that hype may be putting a negative slant on big data in that individuals within the bank may believe its value is overly exaggerated. None of this emerged as part of the literature review possibly because this refers to peoples' feelings toward the concept and this varies greatly from business to business and even country to country.

5.3.3 *Gives rise to big data analytics*

Big data is often described as being a concept and according to the literature the value of big data lies in the associated analytics, ie big data analytics (Mahmood & Afzal, 2013). Interviewees shared this sentiment and volunteered this information without prompting. When discussing the value of big data, 80 per cent of interviewees stated that the value was in the analytics associated with making sense of unstructured data. Interviewees were sufficiently versed to differentiate between big data as a concept and big data analytics as the tool.

Two other key themes that emerged were that BDA allows for faster processing and handles large amounts of data. Interviewees rated the faster processing as being one of the key success factors of BDA. This is in accordance with (Bollier & Firestone, 2010). Interviewees unanimously confirmed that BDA makes the processing of mass volumes of data at fast speeds possible and that this is becoming increasingly important in businesses today. One interviewee made an important and relevant point worth quoting: *"The amount of data is increasing and there is so much on fraud detection. You never know what piece of data is*

going to be very important to you in terms of picking up a given fraud. It means in the end you have to work through huge amounts of data in order to get to what actually give you the necessary results."

5.3.4 *Requires infrastructure*

During discussions, interviewees stated that in order to leverage big data, businesses would need to invest in the appropriate infrastructure. Eighty per cent of interviewees mentioned that big data implementations require non-traditional technology. The literature is also clear on this point, with Bollier and Firestone (2010) mentioning the need for specific BDA tools in order to leverage big data. Interviewees insinuated this may be a deterrent and might be the reason we do not see as many implementations as we should. The following comment was made by one of the interviewees: *"I believe it's a national stat. We have 30 to 40 per cent of SA businesses, maybe not bank related, but SA businesses that use big data. So in South Africa there are a lot of businesses that do not end up using big data. That's the remaining 60 to 70 per cent."*

5.3.5 *Conclusion*

The interviewees were in agreement with what big data referred to and their explanations were often a combination of the three V's of Gartner and the 4 V's of IBM. While the interviewees themselves were very familiar with the concept of big data and understood it to be the same as described by the literature, they did mention that there is often confusion by those outside of the data analytics field. This can be a concern when trying to justify spend on big data technology to senior management. This is especially critical in the fraud space where, as one interviewee put it *"Obviously the systems that keep the financial systems going are the main things in the bank. We never have that type of budget. For something to make a good business case, we have to be very very cost conscious"*. This confusion might have eroded some of the value in big data, especially among those who are not familiar with data analytics.

Nonetheless the interviewees were all in agreement that BDA is where the true value lies and that it is this value that businesses want to leverage. However interviewees made it known that using BDA requires more than just buy-in from management, it requires infrastructure.

5.4 Discussion pertaining to Research Question 3

The third research question was how is the use of big data analytics perceived within the SA banking sector? The researcher was unsure of whether banks were making use of BDA and the purpose of this question was to get insights irrespective of whether interviewees at the banks had practical experience or not. Interviewees once again were allowed to share their thoughts and feelings unhindered. Interviewees were asked to discuss only what they felt comfortable sharing and were encouraged to be as candid as possible. Many themes emerged from responses to this question and it's not possible to discuss all of them, so only the main themes will be discussed.

5.4.1 *Current BDA activity within the banks*

Through discussions with interviewees, it emerged that banks are currently making use of BDA in the areas of social media analysis, consumer analysis, fraud prevention and cybersecurity. This is in accordance with the uses of BDA as mentioned by Chen et al. (2012)

a) **Social media and consumer analysis**

   One of the interviewees mentioned that the approach at the bank she worked for was to start small and test BDA value through proof of concept (POC) projects. If there's value generated they implement and build up the infrastructure. If there's no value generated, they are still happy to have gone through the POC since although they may have lost a little bit of money, the potential benefit could have been huge. She felt that the bank had embraced BDA and were comfortable to experiment with it.

She acknowledged that not all projects would be successful but there were some that were already starting to show success. Much of the projects carried out at this bank were related to text mining and propensity modelling with an ultimate view of better servicing clients. The projects were therefore in the social media and consumer space but the interviewee was confident that this is the groundwork required for other business areas to build off of and that value and insights can be shared throughout the bank.

b) Fraud prevention

Another interviewee discussed BDA specifically within the fraud space at the bank he works for. He mentioned with ever increasing banking transactions, the more you want to do with these transactions, the more you have to look at non-traditional forms of technology. He discussed link analysis, which can be used as a way of detecting fraud by tracking any client links to fraudsters. Links can be identified through a shared telephone number or residential address etc and finding these links between potential fraudsters would take just milliseconds. This then allows the bank to be wary of applications for example being submitted by these potential fraudsters. This sort of analysis however can only take place on non-traditional databases such as NOSql databases and more specifically graph databases. This is difficult to do in a traditional relational database where permutations become exponentially larger very quickly.

The interviewee mentioned that they are experimenting at this stage with graph databases and together with this are making use of an IBM big data technology known as Netezza. He mentioned the secret to Netezza is the actual hardware that this runs on and that it's significantly faster than anything traditional. From discussions with this interviewee around their current implementation for fraud prevention, it was clear that the solution was tailored and included different components, one of which was BDA. The common theme however was that in order to address the

complexities faced by businesses today, solutions are increasingly being found in non-traditional forms of technology such as BDA. Once again this echoes the sentiments of Bollier and Firestone (2010).

c) Cybersecurity

Discussions with an interviewee who is currently working on a BDA implementation for cybersecurity revealed that BDA may just be revolutionising the way security within banks is done. Traditionally bank IT security and cybersecurity was maintained through software products and services referred to as Security information and event management (SIEM). However according to the interviewee SIEM has numerous shortcomings such as being unable to process large volumes of data efficiently, being unable to efficiently store data for later retrieval and having limited search capability. The interviewee mentioned that BDA technology such as Splunk which uses map reduce technology alleviates the issues experienced with traditional SIEM.

This bank has implemented Splunk as part of their IT and cybersecurity solution. The interviewee mentioned that the benefit of using this technology was that it enabled the bank to monitor activity across servers. This is critical because with the increase in data, you require to monitor more servers and with Splunk, the more servers you add the faster the searches become. The researcher also referenced Splunk (2014) as part of the literature review.

The following comment made by the interviewee puts the importance of this technology into perspective: *“The solution also provides an archiving type functionality and users can still search through archived or frozen data. The idea in security terms these days is that you should in theory at least be in a position where you can log everything, like literally every single piece of information. Because in traditional SIEMs, you will always have to make a choice, umm maybe I shouldn't log this, maybe I*

shouldn't log that, and that's not where you want to be. What if that piece of information needed was in that log that you decided to exclude".

5.4.2 *Reasons to implement*

Another major theme that emerged during discussions around current BDA activity were the reasons banks chose to look at BDA technology and why they would continue to do so. Four main themes emerged as reasons to implement BDA within banks. Each of these major themes is discussed in more detail below.

a) Traditional technologies are limiting

Through discussions it surfaced that the driving force behind current BDA implementations is the inadequacies of traditional technologies and this was in accordance with Splunk (2014). One hundred per cent of interviewees mentioned that the processing power of traditional technologies is insufficient. This was a critical factor across business areas and the reason for this can be attributed to the massive amounts of data we now see in the business environment. This is in accordance with Blumberg and Atre (2003) who mentioned that with the advent of the world wide web and with advancements in technology, our methods of conducting business has evolved over time. We are therefore dealing with lots more data in a variety of formats.

b) Return on investment

Interviewees identified the value generated through current BDA activity as being a major reason for continued BDA use within banks. Eighty per cent of interviewees mentioned that the banks they work for have experienced successful implementations that delivered positive return on investment. Interviewees did however mention that returns can be slow but that they are consistent over time.

c) Not just hype

Interviewees also confirmed that while there is a lot of hype associated with BDA, this is common with any new technology. One interviewee mentioned that with companies like Facebook, Twitter and Google relying on BDA technology to drive their own data, it is testament to the value in it.

d) Competitive advantage

Interviewees also mentioned that BDA was a source of competitive advantage with certain banks having experienced increased market share through BDA activity. This was mainly through implementations in the consumer analytics space which allowed the banks to better service and interact with their clients. This echoes the sentiments of Chen et al. (2012) who discussed the unparalleled competitive advantage processing unstructured data provides.

Worth mentioning is a comment made by one of the interviewees regarding the use of BDA in actually building a business case. According to the interviewee, BDA can be used to help banks identify a problem, define the size of the losses related to the problem and identify a potential solution. The solution may not be a big data solution, but BDA can help identify where the bank needs to spend money on controls so that there's no wasteful expenditure.

5.4.3 *Implementation challenges*

While interviewees were very vocal about the benefits and value generated from BDA implementations, they did mention that there are definitely challenges associated with introducing BDA into the banks. These challenges are specific to the South African banking context and therefore not easily tied back to the literature. Four major themes emerged and are discussed below.

a) Skills shortage

The biggest challenge, which was echoed by 100 per cent of interviewees, was that it is difficult to find employees with BDA skills. There are currently no formal BDA qualifications and this is likely to be the case for some time to come. Interviewees felt that this field changes so rapidly that information quickly falls out of date and learning needs to be a continuous process. The consensus was that for the moment for employees to flourish in a BDA role, the employee would need to have a specific combination of skill sets and this is hard to come by. As one interviewee put it *"You need to have a mix of statistical and mathematical skills like an actuarial type person and that needs to be well rounded with someone that actually understands business because it's pointless if you have access to this information but you can't communicate the business need for it in the first place. So trying to find someone with those three things is quite difficult."*

Interviewees mentioned that due to the general shortage of quantitative skills, individuals with these skills are placed in traditional quantitative positions eg credit-modeling, risk modeling. The migration of quantitative skills into the non-traditional spaces like BDA is very slow. Another challenge that surfaced was that once BDA technology is introduced into the banks, the IT department is expected to support the system and this requires upskilling on their part. This is sometimes difficult and time consuming.

b) Culture

A further challenge identified was culture, with 60 per cent of interviewees confirming that change at the big banks happens slowly. Banks generally have legacy systems dating back decades and still prefer these systems. The large banks are generally conservative with their systems and resistant to change. One interviewee compared the culture of the larger banks to that of Capitec and made the following comment: *"Maybe you should speak to Capitec because they're young*

and their systems are all very young and they're far more adventurous when it comes to trying out new ways of doing things"

c) Starting from scratch

According to an interviewee in the cybersecurity field at one of the banks, traditional technologies often already have prebuilt rules, alerts and dashboards and show quick return on investment. With BDA on the other hand, you are often starting with a clean page and building as you go along. This is sometimes challenging and also makes banks apprehensive. However, he did mention that this has the added benefit of the bank being able to customise the solution to its environment. While traditional technologies may show quicker return on investment, solutions are often generic and don't adequately address issues.

d) Infrastructure layout

BDA requires investment in infrastructure, which according to interviewees, makes a BDA implementation expensive. Banks have various legacy systems and need to consider integration and any potential impact an implementation may have on these systems.

5.4.4 *Overcoming implementation challenges*

After discussing the challenges banks face with BDA implementations, interviewees went on to discuss potential solutions to address these challenges. This is once again very specific to the SA banking context and cannot always be tied back to the literature.

a) Proof of concept (POC) projects

One interviewee suggested that one of the ways of introducing BDA into the banks is through POCs. She had been involved with BDA POCs and shared her experiences. She mentioned that it requires minimal investment and then allows the bank to only implement projects that were successful and generated value. The bank can then build off the

infrastructure used for the POC. This offers the bank the opportunity to experiment with minimal investment in infrastructure. The infrastructure can be used for POCs across business areas.

b) On the job learning

One hundred per cent of interviewees mentioned that one way to overcome skills shortages was to grow talent within the bank. It was suggested that banks make use of online resources as training material and that employees with a quantitative background be moulded through training and on-the-job learning. Interviewees mentioned that the bank can then leverage these skills by utilising these employees across business areas as the core skills are the same.

c) Infrastructure

One interviewee suggested open source technology as a solution to expensive infrastructure. In his response he said: *"If banks were less scared of open source solutions… they out there, you can play around and experiment with them and lots of them are free. Open source technology is definitely the way."* Another suggestion was that banks should refrain from big bang implementations and rather start small and build off existing infrastructure where possible.

d) IT vendors

It was also suggested that banks make use of IT vendors where possible especially for generic implementations and thereafter the bank can carry out any customisation in-house for competitive advantage. Interviewees felt that IT vendors have already done much of the groundwork with BDA and it makes sense to leverage off them where possible.

5.4.5 *Future BDA activity within the banks*

The general consensus among interviewees was that the banks want to do more of what they are currently doing with BDA. Interviewees each provided

responses in terms of their areas of expertise and once again the organising themes emerged as the four different business areas, ie social media analytics, consumer analytics, fraud prevention and cybersecurity. Interviewees just mentioned the key areas that they would like to explore with BDA activity in the future. As this refers to future activity in various business areas, not just fraud prevention, it could be a source of competitive advantage. Therefore only key points will be discussed at a high level.

a) Social media analytics

Through BDA banks want to be able to leverage the latest social media activity wherever possible eg geo-location. They want to leverage platforms to find cheaper and more effective ways of interacting with customers. They want to leverage BDA in social media to better profile customers in order to better service them. The Protection of Personal Information Act (POPI) was not seen to be a threat in this regard as social media data is considered public information. However, interviewees did mention it's worth keeping an eye on POPI.

b) Consumer analytics

Here again banks want to use BDA to leverage social media activity to better market products to customers. Banks are also keen to leverage new technology to help them in this regard.

c) Fraud prevention

Banks want to once again use BDA to leverage social media in the fight against fraud. They want to be at a stage where they have fully implemented real-time anomaly detection using BDA. They would one day like to be in a position to use machine learning techniques to identify fraudulent activity thereby limiting human intervention where possible. They want to be able to leverage the massive amounts of data at their disposal to create better indicators for fraud prevention.

d) Cybersecurity

Through BDA, banks want to be able to conduct holistic monitoring in the cybersecurity field. Here again banks want to be able to leverage social media data. Banks want to be able to easily identify weak points in their infrastructure. They also want to be able to go deeper into security failures to better understand and prevent them in the future.

5.4.6 *Conclusion*

From the responses received in relation to this research question, it seems that banks have embraced BDA. Interviewees speaking on behalf of each of the banks they work for did admit that implementations were still in their infancy stages and that a lot more effort needs to go into researching BDA capabilities and its relevance to the banks in South Africa. Spend thus far has been cautious and controlled and banks have generally opted to look at proof of concept projects and small implementations rather than any big bang approach.

From discussions with interviewees it seems that solutions using BDA has already been a game changer in the fields of social media analytics, consumer analytics, fraud and cybersecurity. There was also some very encouraging and inspiring feedback with one interviewee mentioning that South African banks are probably some of the early adopters when it comes to BDA. The interviewee mentioned this with specific reference to Europe and said the following: *"We were invited to actually go speak there and present some of the work we've done. So I think in that sense, its brand new to all the banks and all the world. There's no one that has cracked big data completely and it's just quite cool that we're already in the game."*

The triangulation strategy proved valuable and from discussions it emerged that:

- Banks want to leverage BDA across the business areas and social media data is proving to be key to consumer analytics, fraud prevention and

cybersecurity. Text mining for example, conducted on social media data can be used in the fraud department as well.

- Due to skills shortages banks want to be able to use BDA resources across the business areas
- To reduce investment in infrastructure, the different business areas want to leverage infrastructure and proof of concept projects across business areas.

It was evident that BDA implementations seem to be considered most efficient, when they generate value across business areas. This is in accordance with Bollier and Firestone (2010) who discussed using BDA to aggregate data across disciplines and have an all encompassing view of events. Implementing BDA in this way may be of benefit to areas such as fraud that are unable to easily justify investment in BDA. Since an area like consumer analysis may have an easier time justifying the spend but fraud may still leverage off the infrastructure.

5.5 Discussion pertaining to Research Question 4

The fourth research question was how can SA banks use big data analytics in the fight against cyber fraud? This question was answered somewhat through research questions 2 and 3, but the researcher felt that it was important to have this question addressed directly as well.

5.5.1 *Use of BDA for cyber fraud prevention*

Discussions for this question were built on the comments mentioned in the other research questions. Five major themes emerged and these are discussed below.

a) **Has been introduced into the banks for fraud detection**

Some interviewees confirmed that BDA for fraud prevention has been introduced into the banks they work for and discussed their experience with it. Discussions revealed that there are various levels of how it's being used but generally the use of BDA was still at an exploratory stage.

Most banks are piloting BDA in areas such as social media and consumer analysis, but there is a strong expectation that after gaining traction in these areas, the use of BDA will extend to fraud and cybersecurity.

Interviewees mentioned that BDA is still a fairly new concept and not well understood by the masses. This makes justifying spend on it difficult especially in the area of fraud prevention which is not a profit generating activity at a bank and is generally lower down on the pecking order. The general perception was that as BDA gained more traction and was better understood, it will be more readily applied within the sector and more specifically for cyber fraud prevention.

b) BDA capabilities for fraud detection not thoroughly explored

Through discussions the researcher gathered that BDA capabilities for fraud detection were still relatively unexplored in the SA banking sector. The BDA capabilities mentioned by Mahmood and Afzal (2013) seemed unfamiliar to most interviewees. Interviewees generally associated BDA with social media and consumer analytics and SA does seem to be making good headway with the use of BDA in those fields. As mentioned before, BDA for fraud detection is still at an exploratory stage and therefore BDA for cyber fraud detection even more so.

c) Go to experts

When interviewees were probed on who they felt were the go to experts on BDA for cyber fraud prevention, the majority of interviewees stated that they would prefer to conduct BDA activity in-house. The general perception was that software vendors were too expensive. Interviewees felt that a solution would need to be built based on the specific requirements of the bank and this was best done in-house.

d) Eager to look at what's to come

Even though BDA for cyber fraud prevention has not been thoroughly explored within the banking sector, interviewees seemed positive about

the value in exploring this further. Even in the industry and advisory stakeholder group, interviewees felt there was benefit in exploring BDA further with one interviewee mentioning e-policing and intelligence-led policing using BDA. With reference to fraud, this refers to investigators mitigating risk and reducing financial crime by collectively analysing all information at their disposal.

5.6 Conclusion

BDA for fraud prevention is being used at the banks however it is still at an exploratory stage and there's even less known about BDA for cyber fraud prevention. Banks prefer to carry out BDA implementations in-house, therefore more research needs to be done within the sector on how to best leverage BDA capabilities for cyber fraud prevention. The sentiment is that once BDA is better understood and gains more traction its use will be more widespread. Institutions like SABRIC also see the value of BDA and feel that it can be leveraged in investigations. The general perception is that in the years to come we will definitely see more in-depth, customised implementations of BDA for cyber fraud prevention.

CHAPTER 6. CONCLUSIONS & RECOMMENDATIONS

6.1 Introduction

This chapter presents the conclusions to the research and recommendations based on the presentation and discussion of results in chapters 4 and 5. The conclusions are discussed per research question for ease of reference.

6.2 Conclusions of the study

With the increase in and evolution of cyber-related fraud in the banking sector and with the much talked about global big data phenomenon, the aim of this research was to identify how the use of big data analytics to combat cyber fraud is perceived within the SA banking sector. There was very little literature available on the use of BDA in the SA banking sector, lesser still in its use in fraud detection and almost nothing on how it may be applied to cyber fraud detection.

This study therefore posed four research questions to three stakeholder groups in order to gather perceptions on the use of BDA in cyber fraud prevention. The first research question looked at how SA banks are faring in the fight against cyber fraud. There was consensus that cyber attacks continued to pose a threat to the online platforms and have the potential to completely annihilate them. This was in line with (Paganini, 2013). Costs associated with cyber fraud are exorbitant and banks are starting to take a hard line when it comes to payouts.

When discussing big data, due to the lack of information available, the researcher had to start at the very beginning by identifying how big data is understood within the banking sector. This was addressed through research question 2: how is big data perceived within the SA banking sector? Interviewees' responses regarding their interpretation of big data generally followed the explanations put forward in the literature, with interviewees also mentioning the ever changing definition of big data as identified by Zikopoulos et al. (2012). It did emerge though that big data is not always well understood

by individuals outside of the data analytics field and this was identified as being a concern when trying to justify spend on big data technology.

Interviewees were taken one step further with research question 3: how is the use of big data analytics perceived within the SA banking sector? Results confirmed that banks are using BDA and although most implementations were at preliminary stages, some were already starting to show return on investment. The SA banking sector seems to have more experience with BDA in social media and consumer analytics than fraud prevention. There was no specific mention of how BDA may technically be applied to prevent fraud. While this may be to protect competitive advantage, it did emerge that overall a lot more effort needs to go into researching BDA capabilities and its relevance to the banks in SA. While Mahmood and Afzal (2013) went into great detail discussing BDA capabilities in cyber fraud prevention, this knowledge was lacking in discussions with interviewees.

Interviewees suggested that BDA implementations are most valuable when they generate value across business areas and throughout the bank. Spend thus far has been guarded and banks are weary of making use of software vendors as they are perceived as being too expensive. This would mean that research on BDA capabilities must be done within the banks themselves, if any meaningful large-scale implementations are to be done. Skills shortages were identified as a challenge with banks deciding its best to grow talent in-house.

The fourth research question followed: how can SA banks use big data analytics in the fight against cyber fraud? Results showed that BDA for fraud prevention is still at an exploratory stage within banks and there's even less known about BDA for cyber fraud prevention. The general perception is that once BDA is better understood, we will definitely see more in-depth, customised implementations of BDA for cyber fraud prevention.

The key findings to the research questions are summarised in table 16, below.

Table 16: Key findings to research questions

Research question	Key findings
How are SA banks faring in the fight against cyber fraud?	• Losses don't always materialise in the cyber attack itself and it is difficult to link a fraudulent transaction back to a specific cyber attack • Data breaches take a while to detect and due to globalisation, information stolen in SA can be monetised elsewhere. Countries are not all on the same security layers • Cyber fraud poses a high reputational risk to banks • Law enforcement hasn't caught up with cybercrime • The fight against cyber fraud requires collaborated effort, with banks and law enforcement working together • Attacks in SA are less technologically advanced • Malware is starting to take over from Phishing as being the most prominent type of cyber attack • The cyber attacks are constantly evolving and its next to impossible to completely eradicate them • Consumers are the weakest link and therefore the most targeted • Prevention, detection, and consumer awareness will require more spend from banks. Banks are starting to challenge payouts as the increased spend is unsustainable • Biggest concerns are mass compromisation of bank systems, theft of client data and customers losing trust in online platforms.
How is big data perceived within the SA banking sector	• There is no consistent definition of big data often leading to confusion among those outside of the data analytics field • It is surrounded by hype and this has the potential to erode value especially if senior management feel the benefits are over exaggerated • The value of big data is in the associated analytics, ie BDA • Requires infrastructure.

How is the use of big data analytics perceived within the SA banking sector?	• BDA use within banks is still at an exploratory stage. Banks generally have more experience with BDA in social media and consumer analysis. The use of BDA in fraud prevention and cybersecurity must be better investigated. More research needs to be done with regard to BDA capabilities in general • Implementation of BDA within banks is being seen as a source of competitive advantage • Implementations have already shown return on investment and this is expected to continue into the future • Traditional technologies are limiting and don't cater for the business needs of today • Lack of quantitative skills is a key challenge in BDA implementations. However, there is hope that through on-the job learning and training, skills can be developed in-house • The conservative culture of the large banks is a drawback when it comes to innovation and emerging trends like BDA • Due to the infrastructure layout, BDA implementations are perceived to be expensive, however interviewees identified POCs as a solution to big bang implementations • Banks prefer to leverage BDA capabilities and implementations across business areas.
How can SA banks use big data analytics in the fight against cyber fraud?	• BDA for fraud prevention has been introduced into the banks. • A lot more research needs to go into BDA capabilities, in order to benefit from implementations • Banks prefer to carry out BDA implementations in-house, and are not keen to leverage off software vendors. This means the BDA capabilities need to be researched and investigated within the banks themselves • There is a general positive attitude towards BDA with banks eager to further explore this in the future.

6.3 Recommendations

The researcher mentioned in chapter one of this report that the study may provide guidance to those concerned with combating cyber fraud within the banking sector, as well as certain other sectors in the financial services industry such as the insurance sector. The study was also identified as being of interest to those interested in the application of BDA in solving business problems. Interested groups therefore include those involved in fraud prevention as well as business intelligence. Therefore, recommendations are made to these two groups of people within the financial services industry:

6.3.1 *Business intelligence*

It was clear from discussions that banks want BDA to be leveraged in such a way that it generates value across the business. This means that business areas cannot work in isolation with POCs and implementations. Banks also want to leverage BDA skills across business areas. It would therefore make sense for the business intelligence team to play an enabling role for BDA activity. While BDA is still a new phenomenon and is being explored, it may make financial sense for this to be housed centrally within businesses. It is recommended that the Business Intelligence team take the lead with research, POCs, implementations and skills development for BDA. This will allow for skills, insights and experience to be leveraged across the business.

6.3.2 *Fraud prevention*

Through cyber activity, methods of conducting fraud have mutated many times over and continue to do so, with frightening possibilities. It is incredibly important that those tasked with fraud prevention, leverage technological advancements in the fight against cyber fraud. It is vital to their success that they recognise that with the continued use of traditional tools they are simply taking a knife to a gunfight. Fraud prevention teams need to investigate

solutions that use non-traditional technology such as BDA. While BDA alone may not be the solution, it definitely has a part to play in the overall implementation of a fraud prevention solution.

6.4 Suggestions for further research

Through this research process various key points surfaced. While these points digressed somewhat from the current topic, they were points worth looking into for future research. These are listed below.

6.4.1 *BDA capabilities for fraud detection*

While BDA is known to be valuable in the fight against cyber fraud, through discussions it emerged that interviewees were not always aware of BDA capabilities and how they should be leveraged. Researching this topic could prove valuable to banks and businesses in general.

6.4.2 *Benefits and risks of using open-source technology within banks*

It was mentioned that banks are afraid of making use of open-source technology and that this might actually be unnecessarily holding them back. Worth researching is whether the benefits of using open-source technology, outweigh any perceived risks.

6.4.3 *Internet maturity vs cyber fraud activity globally*

During an interview, a suggestion was made that internet savvy customers might be more aware of cyber threats, making them less susceptible to cyber fraud attacks. This is worth researching by comparing the internet maturity among countries and cyber fraud attacks on their consumers.

6.4.4 *Openness to technological advancements in smaller banks vs the big four banks*

It emerged that the big four banks are technologically conservative and slow to change. Capitec, a smaller, younger and more vibrant bank was identified as possibly being more aggressive in their technological investments. This might be worth looking into, together with the reasons for this, as well as the associated benefits and risks.

6.4.5 *Impact of legislation on BDA for social media*

Social media data was identified as being important across business areas, especially in fraud detection. It is seen as a key source of information that the businesses are basing competitive advantage on. Although POPI was identified as being irrelevant to this source of public information, these social media platforms have various terms and conditions of their own. It is worth further investigating what this means for businesses and any implications thereof.

REFERENCES

AFP. (2014). Don't get caught in a phishing trap. Retrieved from www.banking.org.za website: http://www.banking.org.za/index.php/media-events/new-noteworthy/dont-get-caught-in-phishing-trap

Ahn, S.-H., Kim, N.-U., & Chung, T.-M. (2014). *Big data analysis system concept for detecting unknown attacks.* Paper presented at the Advanced Communication Technology (ICACT), 2014 16th International Conference (pp. 269-272). IEEE.

Andriesse, D., Rossow, C., Stone-Gross, B., Plohmann, D., & Bos, H. (2013, 22-24 Oct. 2013). *Highly resilient peer-to-peer botnets are here: An analysis of Gameover Zeus.* Paper presented at the Malicious and Unwanted Software: "The Americas" (MALWARE), 2013 8th International Conference (pp. 116-123). IEEE.

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative research, 1*(3), 385-405.

Banerjee, A., Barman, D., Faloutsos, M., & Bhuyan, L. N. (2008). *Cyber-fraud is one typo away.* Paper presented at the INFOCOM 2008. The 27th Conference on Computer Communications. IEEE.

Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in management: a decade of progress. *Journal of Management Studies, 48*(8), 1866-1891.

Blumberg, R., & Atre, S. (2003). The problem with unstructured data. *DM REVIEW, 13*, 42-49.

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data*: Aspen Institute, Communications and Society Program Washington, DC, USA.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Bryman, A. (2012). *Social research methods*: Oxford university press.

Cárdenas, A. A., Radosavac, S., Grossklags, J., Chuang, J., & Hoofnagle, C. (2010). *An economic map of cybercrime.* Paper presented at the The 37th Research Conference on Communication, Information and Internet Policy (TPRC). George Mason University Law School, Arlington, VA.

Chang, J., Venkatasubramanian, K. K., West, A. G., & Lee, I. (2013). Analyzing and defending against web-based malware. *ACM Computing Surveys (CSUR), 45*(4), 49.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly, 36*(4), 1165-1188.

Cressey, D. R. (1953). Other people's money; a study of the social psychology of embezzlement.

Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*: Transaction publishers.

Diebold, F. X., Cheng, X., Diebold, S., Foster, D., Halperin, M., Lohr, S., . . . Pospiech, M. (2012). A Personal Perspective on the Origin (s) and Development of “Big Data”: The Phenomenon, the Term, and the Discipline∗.

Dorminey, J., Fleming, A. S., Kranacher, M.-J., & Riley Jr, R. A. (2012). The evolution of fraud theory. *Issues in Accounting Education, 27*(2), 555-579.

Elena, C. (2011). Business intelligence. *Journal of Knowledge Management, Economics and Information Technology, 1*(2).

eNCA. (2013). SA a key target of phishing scams. Retrieved from www.enca.com website: http://www.enca.com/money/sa-key-target-phishing-scams

EY. (2014). Big data, changing the way businesses compete and operate. Retrieved from www.ey.com website: http://www.ey.com/Publication/vwLUAssets/EY_-_Big_data:_changing_the_way_businesses_operate/$FILE/EY-Insights-on-GRC-Big-data.pdf

FBI. (2014). GameOver Zeus (GOZ) Malware and Botnet Architecture. http://www.fbi.gov/news/stories/2014/june/gameover-zeus-botnet-disrupted/documents/gameover-zeus-and-cryptolocker-poster-pdf

Fletcher, N. (2007). Challenges for regulating financial fraud in cyberspace. *Journal of Financial Crime, 14*(2), 190-207.

Gartner. (2013). Big Data. Retrieved 30 June 2014, from http://www.gartner.com/it-glossary/big-data/

Gideon, L. (2012). *Handbook of survey methodology for the social sciences*: Springer.

Gilman, N., Goldhammer, J., & Weber, S. (2011). *Deviant globalization: Black market economy in the 21st century*: Bloomsbury Publishing.

Grinnell Jr, R. M., & Unrau, Y. A. (2010). *Social work research and evaluation: Foundations of evidence-based practice*: Oxford University Press.

IBM. (2014). Apply new analytics tools to reveal new opportunities. Retrieved 01 July 2014, from http://www.ibm.com/smarterplanet/us/en/business_analytics/article/it_business_intelligence.html

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*.

Luhn, H. P. (1958). A business intelligence system. *IBM Journal of Research and Development, 2*(4), 314-319.

Mahmood, T., & Afzal, U. (2013). *College of Computing & Information Sciences, Karachi Institute of Economics & Technology, Pakistan.* Paper presented at the Information Assurance (NCIA), 2013 2nd National Conference (pp. 129-134). IEEE.

Mcleod, D. (2013). SA banks in massive data breach. Retrieved from Mail & Guardian website: http://mg.co.za/article/2013-10-15-sa-banks-in-massive-data-breach

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2008). Verification strategies for establishing reliability and validity in qualitative research. *International journal of qualitative methods, 1*(2), 13-22.

Paganini, P. (2013, 5 November 2013). Modern Online Banking Cyber Crime. Retrieved 17 August 2014, from http://resources.infosecinstitute.com/modern-online-banking-cyber-crime/

Park, B.-K., & Song, I.-Y. (2011). *Toward total business intelligence incorporating structured and unstructured data.* Paper presented at the Proceedings of the 2nd International Workshop on Business intelligence and the Web.

PWC. (2011). Account takeover fact sheet. Retrieved from www.pwc.com website: http://www.pwc.com/us/en/forensic-services/assets/account-takeover-fraud-assessment-fact-sheet.pdf

Raghavan, A., & Parthiban, L. (2014). The effect of cybercrime on a Bank's finances.

Ramamoorti, S., Morrison, D., & Koletar, J. W. (2009). BRINGING FREUD TO FRAUD.

Rider, F. (1944). The Scholar and the Future ofthe Research Library: New York: Hadham Press.

SARB. (2014). Selected SA Banking Sector Trends. Retrieved 05 July, 2014, from https://www.resbank.co.za/publications/detail-item-view/pages/publications.aspx?sarbweb=3b6aa07d-92ab-441f-b7bf-bb7dfb1bedb4&sarblist=21b5222e-7125-4e55-bb65-56fd3333371e&sarbitem=6158

Sood, A. K., & Enbody, R. J. (2013). Targeted cyberattacks: a superset of advanced persistent threats. *IEEE security & privacy, 11*(1), 54-61.

Splunk. (2014). Splunk for Fraud Detection. http://www.splunk.com/web_assets/pdfs/secure/Splunk_for_Fraud.pdf

Thurmond, V. A. (2001). The Point of Triangulation. *Journal of Nursing Scholarship, 33*(3), 253-258. doi: 10.1111/j.1547-5069.2001.00253.x

Van Niekerk, J. (2008). The criminal prosecution of insurance fraud: an encouraging recent decision: case comments. *SA Mercantile Law Journal= SA Tydskrif vir Handelsreg, 20*(2), 280-290.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly, 26*(2), 3.

Wolfpack. (2013). The South African cyber threat barometer. Retrieved from https://www.wolfpackrisk.com website: https://www.wolfpackrisk.com/SA%202012%20Cyber%20Threat%20Barometer_Medium_res.pdf

Zhang, J., & Huang, M. L. (2013). *5Ws model for Big Data analysis and visualization.* Paper presented at the Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference (pp. 1021-1028). IEEE.

Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., & Corrigan, D. (2012). *Harness the Power of Big Data The IBM Big Data Platform*: McGraw Hill Professional.

APPENDIX A

Actual research instrument

The researcher made use of an unstructured interview with one question and an aide memoire.

Question:

How is BDA changing your approach to cyber fraud prevention?

Aide Memoire:

BDA

- common understanding of big data
- Perceived value
- Perceived barriers to success
- Go to experts
- Investment vs business case expectations

Cyber fraud

- Cyber threats
- Consequential types of fraud
- Use of BDA
- Biggest concerns

APPENDIX B

Study information sheet and interview consent form



The Graduate School of Business Administration

2 St David's Place, Parktown, Johannesburg, South Africa

P O Box 98, Wits, 2050

Telephone: +27 11 717 3600, Facsimile: + 27 11 717 3514

Website: www.wbs.ac.za

31 March 2015

Study Information Sheet

This is an exploratory study and the research aim is to identify how the use of big data analytics to combat cyber fraud is perceived within the SA banking sector. Due to the exploratory nature of this study, the research instrument used will be an unstructured interview. This instrument was chosen due to its flexibility.

It is envisaged that the Research Report studies will be published in the Wits internal journal for further study by other students who might be interested in carrying through this study.

The purpose of the MBA Thesis is for students to develop a strong belief about a particular topic or subject, officially declare that belief on that subject within the beginning of the thesis, describe the process by which is intend to prove this belief, carry out that process, and finally describe the results of the process and write a conclusions..

Your assistance will be truly appreciated and you are welcome to question the researcher about the study in order to get a better understanding of the research being carried out.

Yours sincerely

Eulene Pillay

MBA Student

Wits Business School

Tel: 082 922 1758

Email: eulenep@gmail.com

Wits Business School
Sculpting global leaders

The Graduate School of Business Administration

2 St David's Place, Parktown, Johannesburg, South Africa

P O Box 98, Wits, 2050

Telephone: +27 11 717 3600, Facsimile: + 27 11 717 3514

Website: www.wbs.ac.za

Interview Consent Form

- I, the undersigned, have read and understood the Study Information Sheet provided.
- I have been given the opportunity to ask questions about the Study.
- I understand that taking part in the Study will include being interviewed and audio recorded.
- I have been given adequate time to consider my decision and I agree to take part in the Study.
- I understand that my words may be quoted in publications, reports, web pages and other research outputs but my name will not be used.
- I agree to assign the copyright I hold in any material related to this project to Eulene Pillay.
- I understand that I can withdraw from the Study at any time and I will not be asked any questions about why I no longer want to take part.

Name of Participant:

Signature of Participant: _____ Date: _____

Researcher Signature: _____ Date: _____