The Impact of Big Data on Database Administrators from a Skills Perspective

Gloria Ifeoma Ikoli

A dissertation submitted to the University of Dublin
in partial fulfilment of the requirements for the degree of
MSc in Management of Information Systems

1st September 2014

Declaration

I declare that the work described in this dissertation is, except where otherwise stated, entirely my own work, and has not been submitted as an exercise for a degree at this or any other university. I further declare that this research has been carried out in full compliance with the ethical research requirements of the School of Computer Science and Statistics.

Signed: ___________________

Gloria Ifeoma Ikoli

1st September, 2014

Permission to lend and/or copy

I agree that the School of Computer Science and Statistics, Trinity College may lend or copy this dissertation upon request.

Signed: ____________________

Gloria Ifeoma Ikoli

1st September, 2014

Acknowledgements

First and foremost, I would like to thank the Almighty God for his love, strength and grace.

I would like to acknowledge my parents, Mr and Mrs Egri-Okwaji for directing me down this path and for always being there.

I would like to especially thank my husband; Frederick; for being an unflinching pillar of support throughout this trying time and to our children: Jason, Gwyneth and Nicole; the best children in the whole world for putting up with me at my most unbearable.

Many thanks to my siblings for their encouragement and support.

I acknowledge with gratitude the guidance given to me by my supervisor Anthony Niland, support from all other lecturers and fellow course mates who made my time here a very memorable one.

Finally, would like to thank the participants that took part in this research without which this research would not have been possible.

Abstract

The automation and digitalisation of society has given rise to the term big data. The explosive growth in data in recent years in terms of volume, complexity and speed has resulted in new technologies used to handle and analyse these large and complex forms of data. This presents a challenge to most business and IT leaders in storing and capturing these large and complex data as they require additional skills and techniques existing data management teams especially DBAs with traditional knowledge and capabilities lack.

The main objective of this dissertation is to determine the impact big data technologies have on the skills of DBAs, who have experience in managing traditional data environments as there is a lack of comprehensive academic research on the impact of big data on the skills of DBAs.

This exploratory study adopted a positivist methodology by applying a purely quantitative approach to data gathering via an online survey in an attempt to develop a big data skills framework suitable for DBAs. The findings, which should be of interest to organisations planning or currently working on big data initiatives, DBAs and academic researchers, conclude that there is a need for the DBA to upskill in order to manage these new technologies and forms a basis for future research for other emerging technology that may arise in the data management industry.

Table of Contents

List of Tables and Diagrams

List of Tables

List of Diagrams

Abbreviations

ACCA	Association of Chartered Certified Accountants
ACID	Atomicity, Consistency, Isolation, Durability
CEBR	Centre for Economics and Business Research
CIO	Chief Information Officer
DBA	Database Administrator
DBMS	Database Management Systems
ESG	Enterprise Strategy Group
HDFS	Hadoop Distributed File System
HP	Hewlett Packard
IBM	International Business Machines
IDC	International Data Corporation
IT	Information Technology
MPP	Massive Pararell Processing
ODS	Operational Data Stores
OLTP	Online Transaction Processing
RDBMS	Relational Database Management System
RFID	Radio frequency identification
SAP	Systems Applications and Products
SFIA	Skills Framework for the Information Age
SQL	Structured Query Language
TDWI	The Data Warehousing Institute
PaaS	Platform as a Service
SaaS	Software as a Service
IaaS	Infrastructure as a Service

1 Introduction

The emergence of new technologies, such as big data, comes with perceived changes to the data management industry. Data is generated from an increasing number of sources attributed to the advancements in technology. These numerous sources, the amount and frequency of data to be stored and analysed differently has given birth to the loosely defined IT concept called "Big Data". The key enablers for the growth of Big Data have been the exponential increase in the availability of data and the increase in storage capacities and processing power, to the extent that 2.5 quintilion bytes of digital data are created daily from sources, such as social media posts and purchase transaction records (IBM, 2013a).

This exponential increase in data, according to Bean et al. (2012) has led to an increase in the frequency and granuality of data collected by organisations. These new forms of data, its volume and velocity in addition to corporate data generated by organisations, has amplified the challenge to the data management industry and has subsequently given rise to available new technologies (Peppard, 2013). According to the research firm Gartner, organisations have greater understanding of big data's potential value and with this understanding comes a "new set of hurdles spanning people, process and technology" (Gartner, 2013). In relation to technology, the difficulty in storing and managing big data with standard database management and analytical tools is highlighted in a popular definition of big data by the McKinsey Global Institute, which refers to big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (McKinsey Global Institute, 2011, p. 1).

Traditional Business Intelligence, which provides organisations with oppourtunities for fact based decision making and insights for better customer experience, have become inadequate to handle the new forms of diverse datasets generated, resulting in the growth of new analyical methods and technologies. An EMC report based on discussions with the Leadership Council for Information Advantage, an advisory group made up of global information leaders and CIOs as council members, emphasise that the high volume, unstructured data is unsuitable to be used by traditional relational database systems and has thus resulted in the demand for new technologies for data storage and analysis (EMC Corporation, 2011). Similarly, a report by TechAmerica Foundation's Federal Big Data Commission (2004) reveals that the vast amount of data collected by governments and private organisations across society, represent oppourtunuties for those organisations that seek to enhance its value, as well as poses its own set of challenges. These challenges exist because changes to hardware and software data processing techniques are necessary, thereby requiring a set of support professionals with enhanced data management skills. These skills comprise programming, statistical, mathematical, business acumen and communication skills. This rare combination is in short supply in one individual role resulting in organisations creating teams of individuals who possess one or more skills to make up the short fall (Bean et al., 2012).

Questions now arise, such as where does the Database Administrator (DBA) fit in as big data changes the way data is managed and analysed with new terms and new skills. A survey on data management conducted by Enterprise Strategy Group (ESG) shows that organisations identified a lack of skilled resources as a challenge they encounter as they attempt to manage their current database and supporting infrastructure. When asked about the specific skilled resources they lacked, almost half identified DBAs and data architects (Enterprise Strategy Group [ESG], 2011).

1.1 Relevance of the Study

This research attempts to understand the effect big data has on the skills of DBAs by identifying the skills gap encountered if the role of the DBA has to transition to effectively support the implementation of big data initiatives. It can be argued that the management of big data can become a DBA's responsibility. In some organisations, data scientists lack the expertise to store and manage the large amounts of data generated, DBAs are then saddled with the task and have to take up the responsibility, which is unfamiilar territory in terms of skills and capabilities.

This research will provide insights into what skills DBAs should have in order to successfully implement big data initiatives, as big data from a DBA's perspective in terms of skills varies from the traditional data warehousing and business intelligence framework that most organisations have in place. This research will give a rich understanding of Big Data as an emerging technology in data management and will ultimately focus on the skills DBAs need to fit in a big data team.

1.2 Research Question

The primary research question posed is:

What is the impact of big data technologies on the skills of DBAs?

In seeking to answer the research question the following questions were raised:

1. Is the DBA role crucial to the successful implementation of big data initiatives?

2. Will the DBA role need to evolve into a hybrid role in order to support big data initiatives?

3. Will DBAs need new skills to fit in a big data team?

1.3 Scope of this Research

This research primarily focuses on the specific skills needed by Database Administrators (DBAs) to fit into a big data team from the viewpoint of DBAs, other IT personnel and management, as well as from other big data stakeholders. It addresses the skills shortage in terms of storage and management of big data and new technologies that have subsequently arisen from the emergence of big data technologies.

1.4 Beneficiaries of this Research

Apart from the obvious benefit to DBAs, this research could also be of interest to organisations who seek to embark on implementing big data initiatives by pinpointing the skills needed in a big data team and how the DBA role can contribute to its successful implementation. This investigation is based on a formulated big data skills framework, which can be of benefit to organisations and academic researchers for further extensive research in this area.

1.5 Literature Research Sources

It is worth noting that the limited amount of academic and business related literature corncerning the research question. Although, white paper reports on big data by various organisatons were widely available, there was very little research evidence available from the view point of business stakeholders on the impact of big data technologies on DBA skills in particular. Given the limited academic literature on the research topic, it is hoped that this exploratory research can form the basis for further indepth research in the research area.

1.6 Dissertation Roadmap

Chapter one consists of background information regarding the research question and sub-questions.The chapter concludes with the relevance of carrying out the research and its beneficiaries.

Chapter two offers a critical review of literature of big data, its definition, current trends and exponential growth, as well as related technologies.The chapter also hightlights the difference in the traditional and emerging big data technologies, ending with a review of literature concerning different skills framework for big data professionals.

Chapter three describes the methodological underpinnings of this research used to answer the research question in terms of strategy and approach. The chapter presents a big data skills framework formulated as part of this study, discusses the ethical process and considerations applied in this research and concludes with a brief summary of the main points of the chapter. The chapter concludes with a formulation of a big data skills framework, the ethical process and considerations applied in this research.

Chapter four presents the findings and analysis of data derived from the online survey, which attempts to answer the research question and objectives.

Chapter five concludes the research by highlighting the key findings, the limitations of the research and suggests future research directions pertaining to the research topic.

2 Literature Review

2.1 Introduction

The main aim of this chapter is to review available literature relating to big data and its impact on data management professionals in particular Database Adminstrators. It also intends to provide and contribute to a deeper understanding of what big data is as an emerging technology in data management.

The review is presented in sections starting with the definition of big data and its three major characteristics in section 2.2.

The current trend and exponential growth of big data, its adoption rate in Ireland and globally, as well as its benefits in general will constitute section 2.3.

Section 2.4, explores big data technologies and traditional systems highlighting the differences between them in terms of technology.

Section 2.5 presents an overview of some big data challenges faced by organisations in implementing big data initiatives.

Section 2.6 discusses the literature surrounding skill framework for big data professionals.

Section 2.7 summarises and concludes the chapter.

2.2 Big Data Definition

Big data has recently received considerable amount of attention from academia, governments, press and IT professionals to the extent it has become popular in the IT world. As a result, what constitutes the term “big data” is loosely defined, relative and varies from one organisation to another.

Despite the lack of consensus, definitions of Big Data in literature all seem to take into account its complexity, diversity and unstructured nature (O'leary, 2013). Snijders et al. (2012, p .1) offer a simple definition by describing big data as “large and complex data sets that become awkward to work with using standard statistical software”. Similarly, Ahuja and Moore (2013, p. 62) describe big data quite simply as data that is “too large, grows too fast and does not fit into traditional database systems”.

According to McKinsey Global Institute (2011, p. 1) as cited by Kim et al. (2013), "Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse"*.*
Along the same lines, Navint (2012, p. 2) defined Big Data as:

> a term that refers to data sets or combinations of data sets whose size (volume), complexity (variability), and rate of growth (velocity) make them difficult to be captured, managed, processed or analysed by conventional technologies and tools, such as relational databases and desktop statistics or visualization packages, within the time necessary to make them useful.

In a report by Global Pulse, a new initiative launched by the United Nations, big data is described as the speed and frequency with which data is created and transmitted on one hand and on the other; the growth in the variety of sources, constitutes what can be referred to as the Data Deluge or Big Data (Letouzé, 2012).

According to Zikopoulos et al. (2012) as cited by O'leary (2013), big data is characterised by "3V" words. The first "V" is volume, which is an indicator of the sheer amount or quantity of data that is produced and analysed compared to other traditional sources. Velocity is the second, which implies that data is generated at great speeds and Variety, the third "V"; this suggests that there are different forms of this emerging data which are different from traditional systems.

The most widely accepted and most popular definition is one coined in a February 2001 research report by Meta Group (now Gartner's) analyst Doug Laney. He described big data as data characterised by "three Vs", which are volume, velocity and variety (Laney, 2001).

In 2012, The IT analyst firm Gartner, updated its definition to:

> Big data is high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. (Gartner, 2012a).

International Data Corporation (2011, p. 6) has a similar definition of big data as:

> a <u>new</u> generation of technologies and architectures, designed to <u>economically</u> extract <u>value</u> from very large <u>volumes</u> of a wide <u>variety</u> of data, by enabling high-<u>velocity</u> capture, discovery, and/or analysis.

However Gualtieri (2012), Forrester's Analyst in his blog defines big data as:

> the frontier of a firm's ability to store, process, and access (SPA) all the data it needs to operate effectively, make decisions, reduce risks, and serve customers.

This definition does not only take into account the measures of data in terms of the size of the datasets, which the 3Vs encompass, it also acknowledges that there are difficulties managing big data in terms of storage, processing and access.

2.2.1 The Characteristics of Big Data

The "Three V" model is used to describe the rate at which the large volume of data generated has evolved. Looking at big data in terms of just one characteristic is inadequate as each "V" has its own implications and adds meaning to the paradigm Big Data.

1. Volume:

   Volume is a relative term because data volumes vary from one organisation to another ranging from multiple petabytes in large organisations to tens of terabytes for small or midsize organisations. These volumes then become difficult to manage, hence the term "Big Data" (Sheppard, 2011).

   According to Harvard Business Review (2012), 2.5 Exabytes of data were created by the Internet's use each day in 2012 alone. This gives an idea of the volume of data organisations work with in a single dataset. Walmart, an American multinational retail corporation, for example, is estimated to generate more than 2.5 petabytes of data hourly from customer tranactions alone.

   As data volumes increase, there is a need for greater sophistication of technologies to harness the benefits that can be derived from it (Centre for Economics and Business Research, 2013).

2. Variety:

   Big data is derived from many sources and in different formats. Generally, data can be divided into structured, semi-structured and unstructured data. According to Gartner (2011), unstructured data represents 85% of data.
   Structured data: Structured data is described as data grouped into rows and

columns, which makes it easy to query and obtain information for an organisation's operational requirements (Centre for Economics and Business Research, 2013). Traditional data management systems store data in a structured format but the need to store unstructured data has brought about new systems to store and analyse such data.

Semi-structured data: This data is a mix of structured and unstructured data.

Unstructured data: The unstructured nature of big data derived from sources, such as Internet searches, weblogs, radio frequency Id (RFIDs) to mention a few, drives its complexity and most often is combined with structured data; that is, data from a more conventional source, such as data from a relational database or Customer Relationship Management System (Navint, 2012). With the emergence of new sources of information, which are ubiquitous and unstructured by nature, as well as the decline in memory, processing, storage and cheap computing in general, a new era has dawned whereby relational databases are inadequate to cope with the storage and processing of large volumes of data. (Harvard Business Review, 2012).

3. Velocity:

This characteristic of big data is fuelled by the ubiquitous nature of real-time data systems and networks (Navint, 2012). The rate or speed by which data is generated is as important as the other characteristics as this provides rapid vision that can in turn provide competitive advantage and prove to be beneficial to organisations (Harvard Business Review, 2012).

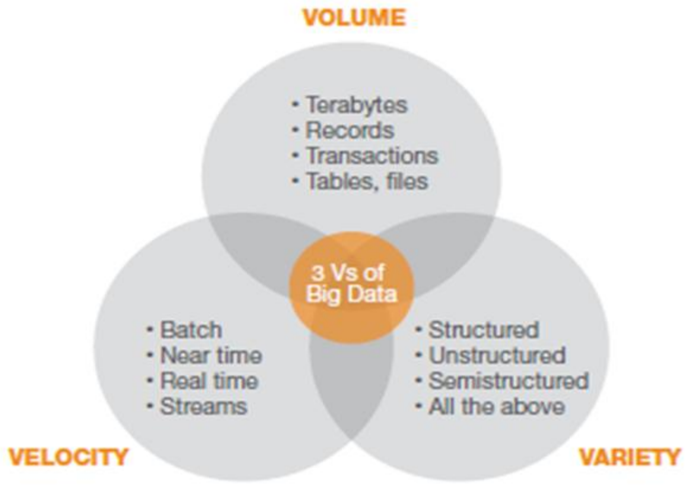


Figure 2.1 – The '3V' model of big data (The Data Warehousing Institute 2011)

There are two other attributes of Big Data that have recently emerged. Zikopoulos et al. (2013) in a more recent publication stipulate Veracity, which refers to the reliability and accuracy of data; and value, which refers to the quality of data and its commercial value. These are the fourth and fifth characteristics of Big data. Volume is by far the most emphasised characteristic by data management professionals and business leaders with minimal consideration for the other two major quantifiable characteristics: variety and velocity (Gartner, 2011).

2.3 Current Trend and the Exponential Growth of Big Data

Consumer Data, about two decades ago, was difficult to obtain as companies had to pay data collection and survey companies. Today, the reverse is the case, as data is in abundance and keeping up with the amount that is generated is an enormous challenge most companies face (Economist Intelligence Unit, 2011). According to Gartner (2011), a google search on the term big data in February 2011 resulted in 2.9 million hits with the term's popularity emerging in 2009.

With the advent of the digital age comes an exponential growth in the amount of data that is generated from different sources ranging from social media to online purchase transactions (Ahuja and Moore, 2013). The growth of data in the digital universe continues at an exponential/quantum rate exacerbated by mobile devices and social media. This growth presents a challenge and an opportunity to unlock its economic value (e-skills UK, 2013).

Another key driver to this exponential growth of data is the “Internet of Everything”. This is the rate at which new digital devices are connected to the Internet so much so that there are now more than 10 billion wireless devices in the market, with an estimate of over 30 billion expected by 2020 (ABI Research, 2013).

The vast amount of data and the frequency at which data has become uncontrolable, like information are produced (Figure 2.2) has led to this data to be referred to as “Big Data” (Daas et al., 2012). The astronomical growth in global data traffic is expected to reach a staggering 6.6 zettabytes by the end of 2016 (Tole, 2013).

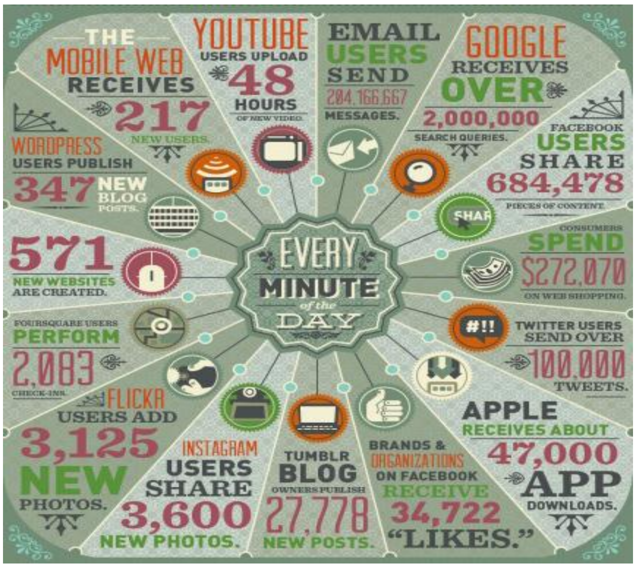


Figure 2.2 – Sources of Big Data (James, 2012)

According to IBM (2013a), 2.5 quintillion bytes of data are created daily, so much so that 90 per cent of the world's data was created two years ago. It is expected that by the end of 2020, data would have increased to a staggering 40 exabytes (International Data Corporation, 2012).

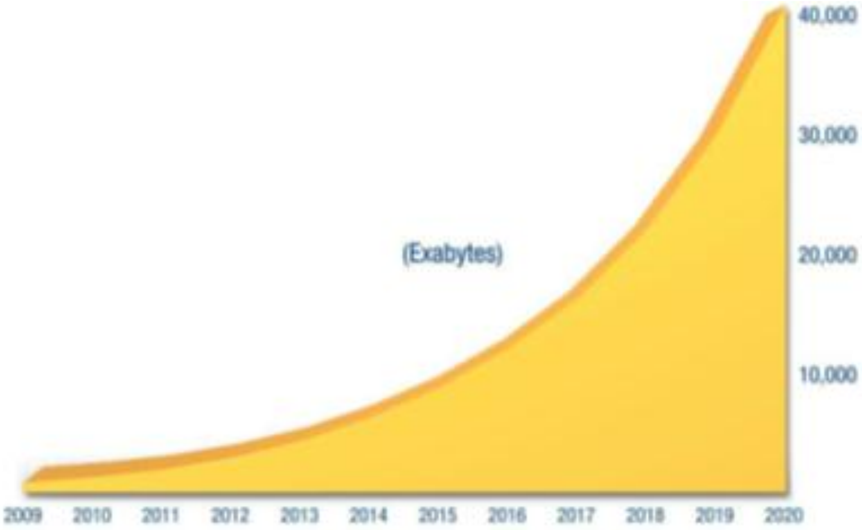


Figure 2.3 – Growth in the Digital Universe from 2010 to the end of 2020 (International Data Corporation, 2012)

With this astronomical growth in data patterns, there is a necessity for large scale infrastructure that balances high performance with cost, density and energy efficiency in order to facilitate the economically correct handling of colossal data storage requirements and processing (Villars et al., 2011).

David Cearley, Gartner's vice president as cited by TMC News (2012) points out the key consideration of organisations should be the realisation of their inablility to store the vast amount of data generated today and acknowledge the emergence of new techniques to manage the exceptional data, as well as the development of new skills for the effective use and management of these technologies.

2.3.1 Adoption Rate and Benefits of Big Data

Ularu et al. (2012) in their study explain that if Big data is implemented properly, there will be a better view of the business, and efficiencies in different departments would be created. Big data can become a strategic planning and commercial tool if properly analysed and mapped to granular decision making practices and procedures (ACCA, 2013).

In a 2013 report conducted by The Centre for Economics and Business Research (Cebr), big data technologies in Ireland are in the early stages of adoption and there are different rates of adoption by industry (Centre for Economics and Business Research, 2013). Table 2.1 gives a breakdown of the 2012 adoption rates in Ireland by industry. A similar research published by e-skills UK on behalf of SAS UK & Ireland showed the growth of big data implementation by sector and identified the IT firms as more likely to have implemented big data than other sectors, such as the manufacturing and public sectors.

According to e-skills UK (2013), the adoption rate of big data among large organisations in the UK is 14 percent, with another 7 percent at various stages of implementation. This equates to about 4,600 businesses with big data already implemented by the end of 2013. The report suggests that the main motivator for its adoption among the organisations interviewed was the large amount of data they had to deal with. Data variation in the form of structured and unstructured data was the second main reason for adoption leaving data velocity, a core characteristic of big data, as the third motivator for adoption. Despite this, the adoption rate amongst larger organisations according to the report will more than double between 2012 and 2017, resulting in a growth rate in employment of big data specialist of 243 percent over the period.

Table 2.1 – 2012 Big Data Adoption Rates in Ireland by Industry (Cebr, 2013)

Industry	2012 Big Data Analytics Adoption
Telecoms	31%
Transportation & Logistics	31%
Financial Intermediation	30%
Manufacturing	30%
Insurance	30%
Central Government	29%
Energy & Utilities	29%
Other Activities	29%
Retail & Wholesale Trade	27%
Real Estate & Professional Sevices	26%
Healthcare	26%
Irish economy	29%

The potential benefits of big data are wide ranging when big data is efficiently analysed. Efficiency improvements, such as increased sales, better quality customer service and improved products/services materialise from gaining insights into the overall business, customers and products (Navint, 2012). There are enormous benefits and untapped opportunities that exist in big data but many organisations in most sectors find it hard to handle and exploit these opportunities (Hagström and Manocha, 2013).

The first ever research on the impact of big data on the Irish economy was conducted by The Centre for Economics and Business Research (Cebr) in 2013. The report discloses that many Irish organisations are reaping the benefits of big data and that it would be worth 27 billion euro to the Irish economy by 2017. The report also focused on the impacts big data will have on employment. In the same vein, a report published by e-skills UK acknowledges that the rewards of adopting big data in an organisation can be huge but depends on the ability to acquire qualified staff, as well as retain and develop them (e-skills UK, 2013).

Optimising big data can deliver business value to stakeholders in an organisation. As explained in a paper by ISACA (2013), this can be achieved with a comprehensive governance and manaement strategy to manage the risks and reap the benefits of collecting, storing and analysing the information from big data. The paper went further to say that enterprises can reap significant rewards by mastering big data management because when there is an alignment between data management processes and the enterprise's strategy, financial benefits can be realised. It can in fact have a positive impact in the following areas: Product development, Market development, Operational efficiency, Customer experience and loyalty, as well as Market demand predictions.

There is a tremendous amount of benefit organisations stand to gain in terms of transforming their operations, improving service to their customers and innovating in the respective markets through the implementation of big data initiatives and programs. However, this can only be achieved by implementing robust principles and not just implementing the initiatives based on latest trends or empty vendor claims (Mithas et al., 2013).

The Data Warehousing Institute (2011) in a report identifies fraud detection, better social influence marketing, better customer intelligence and better planning and forecasting as the benefits organisations derive from big data.

Table 2.2 **–** Proportion of Businesses indicating benefit of Big Data (Cebr, 2013)

Big Data Analytics Benefit	Businesses Reporting Benefit (%)
Better social influencer marketing	61%
More accurate business insights	45%
Segmentation of customer base	41%
Identifying sales and market opportunities	38%
Automated decisions for real-time processes	37%
Detection of fraud	33%
Quantification risks	30%
Better planning and forecasting	29%
Identifying cost drivers	29%

2.4 Big Data Technologies

The development of new big data tools and technologies which rely on Massive Parallel processing (MPP) have emerged and are constantly developing (Navint, 2012).
Below is an overview of the important technologies in a big data infrastructure.

2.4.1 The Hadoop Framework And NoSQL Databases

Big data applications are challenged by the use of complex data. This challenge goes beyond the regular managing and processing of these data, which has led to an open source project called Apache Hadoop (Wu et al., 2014).

Fingar (2011) considers Apache Hadoop the most popular open-source software with the most market traction in recent years. Fingar goes on to describe Hadoop's functional ability as being able to:

> process large caches of data by breaking them into smaller, more accessible batches and distributing them to multiple servers to analyse. It's like cutting your food into smaller pieces for easier consumption. Hadoop then processes queries and delivers the requested results in far less time than old-school analytics software—most often minutes instead of hours or days.

According to Lo (2013), Hadoop has been a game changer in the support of processing large amounts of data as its computational model takes high intensitity data processes and spreads them across its cluster, which is an endless number of servers. The Hadoop framework is an open source software library, administered by Apache Software Foundation (ASF), whose main attribute is the distributed processing of large amounts of datasets in a clustered environment, whilst offering local storage and computation (The Data Warehousing Institute [TDWI], 2013b).

According to Chandran (2013), the HDFS (Hadoop Distributed File System) offers a high degree of scalability and availability for storing data on affordable hardware. Despite this, Hadoop is not a relational database engine even with its SQL execution engine, it is unable to maintain complex relationships between tables and manage user access permissions, as well as concurrent access. In that regard, the traditional relational database management systems (RDBMSs) offer a faster and richer environment. The most widely used open source big data technology from a software perspective is Hadoop. However, it has its limitations in terms of transaction processing: data sharing and mixing data formats are more unstable compared to traditional RDBMSs (Villars et al., 2011).

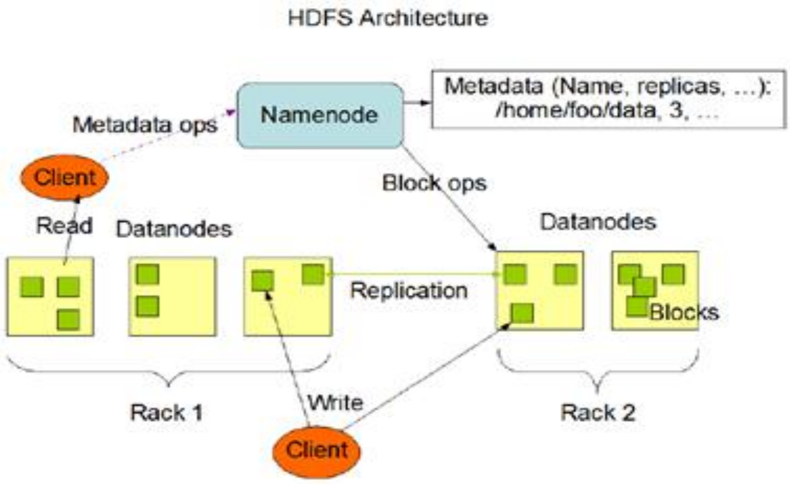


Figure 2.4 – HDFS Architecture (Borthakur, 2008)

Hadoop has four outstanding characteristics which according to IBM (2013b) are:

1. Scalability – New nodes can be added as needed without the need to change data formats, how data is loaded, how jobs are written etc.

2. Cost effective – Apache Hadoop brings parallel computing to commodity hardware, which results in a significant decline in the cost of storage which in turn makes data modelling more affordable.

3. Flexibility – Because Hadoop is schema-less, which makes it easier to absorb any type of data (structured or unstructured and from any number of sources), data from multiple sources can be joined and aggregated in indiscriminate ways enabling more indepth analysis of data than any one system can offer.

4. Fault tolerant– Hadoop is extremely fault tolerant. This is so because as a node goes down, it would automatically assign the system to another node, thereby enabling continous processing without any downtime whatsoever.

2.4.2 NoSQL Databases

As earlier stated, big data has brought with it many challenges. To deal with these challenges new technologies have emerged and grouped under the word “NoSQL” which is an acronym for “Not Only SQL”. This database management system can scale to very large datasets making it different from traditional RDBMSs, in terms of architecture and data model. Simmonds (2013) is of the opinion that Big data comprises of structured and unstructured data. The preferable storage method is using NoSQL storage types as a complement and not as a replacement to relational databases. A distinguishing feature attributed to NoSQL databases is the ability to scale horizontally, which is replicating data across many servers, allowing for large read/write transactions per second (Cattell, 2011) Cattell (2011) stipulates that NoSQL databases have six distinguishing features which are:

1. The capacity to horizontally scale “simple operation” throughput over many servers.
2. The ability to replicate and to distribute (partition) data over many servers.
3. A simple call level interface or protocol (in contrast to a SQL binding).
4. A weaker concurrency model than the ACID transactions of most relational (SQL) database systems.

5. Efficient use of distributed indexes and RAM for data storage.
6. The ability to dynamically add new attributes to data records.

Cattell (2011) also states in his paper that there are four different types of NoSQL databases that are categorised into scalable relational systems, document stores, key-value stores and extensible record stores. NoSQL databases are increasingly used by organisations involved in collecting large amounts of unstructured data. The databases are distributed, non-relational database management systems, not built on tables and do not use SQL to query the database (Moniruzzaman and Hossain, 2013). The NoSQL databases can scale to millions of active users as opposed to traditional databases, such as Oracle, SQL and MySQL databases, which have been predominant since the 1980s. NoSQL databases are used by Google and Amazon who have one of the biggest data warehouses in the world (Lo, 2013).

2.4.3 An Overview of RDBMS

Traditional relational Database Management Systems (DBMS), such as Oracle, DB2 and SQL server have an SQL interface, predefined schemas and horizontal scaling features (Cattell, 2011). The architectures of these systems are highly structured and follow a relational model introduced in 1970 by IBM’s E.F Codd, which is the basis for many popular database systems currently in use, such as Oracle and SQL Server (Codd, 1970). The need for this structure in data becomes a huge obstacle for large volumes and diversified data,and is therefore not suitable for Big Data analytics (Lo, 2013).

RDBMSs have several advantages that include familiarity with SQL commands compared with commands used by NoSQL databases, availability of third-party tools for easier administration in terms of report generation, making it possible to store data in different schemas thereby providing a simpler data structure (Cattell, 2011). According to Chandran (2013), RDBMSs offer a richer environment and much faster performance in terms of sophistication and complexity of SQL-based operations, as well as an easy to use environment for speedy development of applications. However, there are disadvantages associated with RDBMSs in relation to constraints regarding available memory on a single server, resulting in DBMS solutions restricted to relatively small sizes.

Some organisations run a Hadoop environment and a traditional data warehouse alongside each other. This has prompted leading vendors, such as Oracle, SAS and IBM to start working towards integrating their offerings with Hadoop functionality, which involves using MapReduce to accept the data, generate reports and eventually, provide a formatted dataset which is loaded into a data warehouse (Villars et al., 2011). Table 2.3 is an overview of the different data storage technologies that are associated with big data.

Table 2.3 – Overview of different data storage techologies (Pugh, 2013)

	RDBMS	Analytic Data Stores	NoSql	Hadoop
Description	Traditional row-column databases used for transactional systems, reporting, and archiving	Optimized for data-access (as opposed to writes) and leverage columnar or in-memory technology to provide fast data access at the expense of write-performance limitations.	Designed for rapid access to “key-value” pair combinations. Useful for products like facebook/twitter where most information revolves around one “key” piece of data	An open-source approach to storing data in a file system across a range of commodity hardware and processing it utilizing parallelism (multiple systems at once)
Examples	Sql Server, MySql, Oracle, etc	Vertica, Kognitio, ParAccel, Netezza, InfoBright, Amazon RedShift	MongoDB, Cassandra	Hadoop implementations by CloudEra, Intel, Amazon,
Good for…	Reads & Writes, “reasonable” datasets (< 1B rows)	Storing lots of information, great query/retrieval speeds	Storing information of a certain type, great retrieval speed based on a key, write	Inexpensive storage of lots of data, structured & semi-structured
Not good for…	Massive data volumes, unstructured & semi-structured data	Unstructured & semi-structured data, writes (one at a time)	Not used for grouping information across keys (such as, for reporting)	Complex, code-based, incompatible approaches in market, writes (one at a time)
Notes	Challenging to “scale-out”	Often viewed as an alternative to traditional RDBMS when read performance is important	Enables faster productivity when creating data-driven applications as there is less up-front design work needed	Strong bias to the open-source community & Java

2.4.4 Big Data in the Cloud

According to Villars et al. (2011, p. 12),

> Traditional business intelligence systems have historically been centrally managed in an enterprise data centre with a scalable server and high-performance storage infrastructure built around a relational database.

The conventional IT budget planning in organisations is fast becoming obsolete as they now have the option to either buy hardware or software or to rent infrastructure or software-as-a-service with the different cloud offerings available today. These offerings come at a time where the rapid growth in data volume is much greater in magnitude than an organisation's traditional operational data (Chandran, 2013).

> Big data itself includes aspects of cloud computing and mobile computing—all of it seems to be feeding on each other and accelerating the market to a pace that we've not seen before (TDWI, 2013a, p. 10).

Ahuja and Moore (2013) are of the same view that big data is so intertwined with cloud computing. As a result of this, the emergence of cloud computing has reduced the processing costs of big data. According to Tole (2013), companies have several storage choices, such as on cloud systems, "in-house" systems or adopting a hybrid solution. In the big data era, the cloud provides a swift and responsive data platform that supports the business at a lower cost than traditional platforms (ESG, 2011).

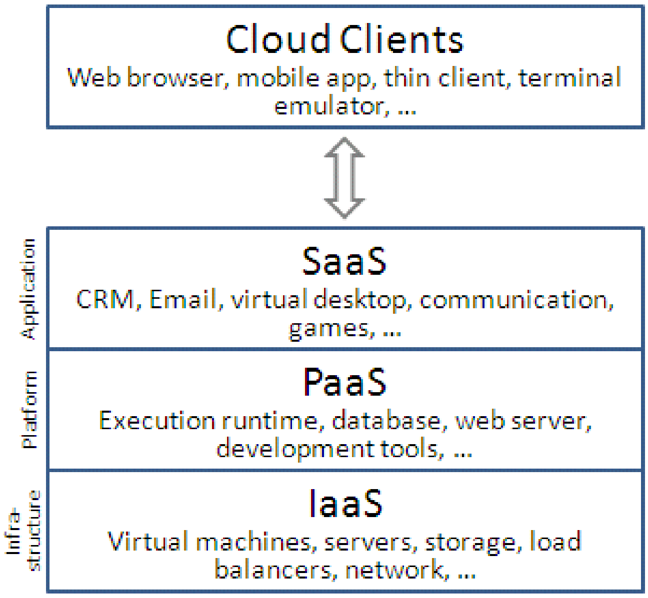


Figure 2.5 – Cloud Computing Layers (Tole, 2013)

As the rate of cloud computing adoption increases in organisations, the role of cloud computing in big data analytics will also increase (Armbrust et al., 2010). According to Armbrust et al. (2010) there are three aspects of cloud computing that bolster big data analytics. The first is the availability of infinite resources on demand, which make planning ahead for provisioning not necessary. Secondly, cloud users do not have to make an upfront commitment on the resources they require as they can start small and gradually increase resources as required. The third aspect is the fact that payment of resources can be made on a short term basis as needed and can be released when no longer needed. All these advantages aid big data implementation as storage can increase when required, allowing for flexibility in the demand for resources. These principles make the dynamics of the cloud irresistible to organisations that implement big data initiatives, and who are conscious of investment costs as the scalability of resources can be easily made.

A study by International Data Corporation (IDC) sponsored by EMC corporation, shares a similar view that cloud computing solutions provide organisations with agility and flexibility as opposed to traditional systems. The study predicts cloud computing to be the key for handling future complexity that arises from devices that generate big data.

They suggested in their paper that IT is viewed as a service and is enabled by the cloud. They go further to predict organisations would continue to view IT as an external service as they reap the benefits from cloud computing and conquer the infrastructure challenges they face due to the dawn of the “big data phenomenon”.

2.5 Big Data Management Challenges

The pursuit of conquering the challenges associated with big data management has led to a plethora of systems (Agrawal et al., 2011). A typical IT department consists of different teams, such as network, server and security teams (Fung, 2013). DBAs usually fall under the support team.

As of mid-2009, data management language and landscape was as simple as Online Transaction Processing (OLTP) supporting the enterprise’s business processes. This involved operational data stores (ODSs) storing the business tranactions in order to support operational reporting, and enterprise data warehouses (EDWs) storing and transforming business transactions in order to support both operational and strategic decision making (Ularu et al., 2012).

Big Data has now become a key area in the IT industry and managing its large amounts

of data including its storage and tools have become challenging (Kim et al., 2013).
Bertino (2013) explains that the real challenges exists not only in the volume of data generated but also in the large volumes of unstructured and structured data from numerous sources.

2.5.1 Privacy and Security

According to International Data Corporation (2011), as people continue to use digital devices and engage in social media, online data collection will become increasingly intrusive because big data has the ability to profile, as well as identify individual consumers. This has led to advocates and regulators calling for amendments to current privacy and data protection laws as acquisition and sharing of information has raised privacy concerns regarding big data. The revelations made by whistle blower Edward Snowden has seen an increase in the discussion about privacy and big data. Organisations have to weigh the subtle/delicate balance of benefits against ethical and privacy risks.

2.5.2 Technology and Infrastructure

Data volumes have grown faster than IT budgets can ever increase therefore CIOs are challenged to do more in the tight financial circumstance (Chandran, 2013). High scalability and flexibility are two major factors that have led to big data transforming the way data is stored. However, to take advantage of cloud computing technologies, data must first be in the cloud. This raises issues of data transfer as basic transfer protocols are inadequate for the sheer volume of data to be transferred (Ahuja and Moore, 2013). In recent years, the tools needed to deal with big data have become open sourced and inexpensive as Hadoop (the most popular framework), integrates with open source software, commodity hardware and stores data on cheap distributed servers. However, the skills needed to manage these new technologies are new and scarce (Harvard Business Review, 2012).

2.5.3 Talent Management

A significant factor that impacts organisations and their staff today is technological advancement and change. This can lead to a shift in the nature of work as the impact of new technology has the ability to change the role of an organisation’s employees(Gardner et al., 2003). According to Gartner, a global estimate of 4.4 million jobs will be directly created by big data by 2015 with 1.9million of them in the United States alone (Gartner, 2012b).

Hagström and Manocha (2013), proposes shortage of talent can be the greatest impediment to obtaining value from big data. According to e-skills UK (2013), the demand for big data skills is rising exponentially with an estimate of 31,000 people working in the specialist big data roles in the UK. Chandran (2013) posits that SQL and NoSQL environments are very different in terms of skills, such that vendors have tried to come up with solutions to circumvent this apparent skills shortage. MapReduce was first integrated with Hadoop but there was still a problem in terms of skills shortage for MapReduce programming compared to SQL. Pig and Hive, a “lite-SQL API”, was then applied on top of the MapReduce framework to help bridge the skills gap but disappointingly had some performance issues.

In a survey conducted by ESG, 28% of the organisations that participated cited a lack of skilled resources as a challenge in managing their current database and supporting infrastructure in relation to the specific skilled resources they lacked, almost half identified DBAs and data architects (ESG, 2011). ESG further stated that data architects are key to successful data analytic projects as they bridged the gap between the physical storage of data and the business problem. Interestingly, the report also highlighted the importance of data-based skills in the big data analytics business world in which the data scientist role did not exist a few years ago. In a survey conducted by TDWI (2011), 46% of organisations indicated inadequate staffing and skill as a leading barrier. e-skills UK (2013), identified training as the primary means of dealing with the apparent skills shortage. Results from their survey showed that 45% of organisations interviewed indicated that data/analytic skills development or training are major considerations for adoption.

2.6 Core Competencies of a DBA

According to Mullins (2002), the heart of an organisation is its data that resides in databases, which therefore places the DBA at the center of the business. Mullins defines a DBA as “the information technician responsible for ensuring the ongoing operational functionality and efficiency of an organisation's databases and the applications that access those databases” (Mullins, 2002).

Competence is defined as “a demonstrated ability to apply knowledge, skills and attitudes to achieving observable results” (European e-Competence Framework 2.0, 2010,p 6).
With the emergence of new technology, such as virtualisation, cloud computing and Big data, DBAs must have essential, specialised skills. A DBA requires a number of skills that can be broadly categorised into Technological and Interpersonal skills. Polakowski (2009)

pointed out that DBAs must not just be technically skilled, but must also have exceptional interpersonal skills as written and oral communication is a basic skill and a critical element. Mullins (2002) agrees that DBAs must possess excellent communication skills as they frequently have to interact with different teams, end users and executives in an organisation.

According to the National Careers Service (2012), a DBA should have an excellent knowledge of database management systems and computer systems in general, a high degree of accuracy, attention to detail, organisational and communication skills, as well as problem solving skills. These skills are crucial because a DBA's main responsibility is to establish, manage and support data access for a particular database technology (IBM, 2014).

In the article, techdirections (2010,p. 24) stipulates the job of a DBA can be complex, time consuming and requires significant training. It sets out the personal characteristics and skill required as:

- Strong organisational skills.
- Strong logical and analytical thinker.
- Ability to concentrate and pay close attention to detail.
- Strong written and verbal communication skills.
- Willing to pursue education throughout the career as a DBA.

Defining a DBA's job is difficult as technical responsibilities become more challenging with the emergence of new technologies.. NoSQL database systems implementation is added to the growing number of required skills DBAs must have (Mullins, 2013). It is undeniable that a major challenge DBAs encounter is the rapid and unrelenting growth of data (McKendrick, 2011). Simmonds (2013), suggests that in the big data, the DBA, System Administrator and Developer roles are blurred as a combination of skills from these roles is needed to work with a Hadoop system. There is no set path to becoming a DBA but a typical learning path begins with an undergraguate degree. Many organisations look for a degree in Information Science, Computer Science or a related Information Technology degree and require hands on experience for the position. There are several certification programs by database management system vendors, such as Oracle and Microsoft, that are highly regarded by empolyers and seen as industry standard (techdirections, 2010).

Polakowski (2009) agrees with this view stating, there are quite a number of certifications that lead to the path of database administration usually based on the background of a degree in Computer Science or Information Technology which means very little without real life experience. Polakowski argues that a DBA’s job is 24/7 considering an organisation’s data integrity must be maintained continously and the various other tasks and responsilities that come with the role. DBAs have had to acquire new skills due to the emergence of new innovative systems (Simmonds, 2013).

2.7 Skills Framework

The Skills Framework for the Information Age (SFIA) is a globally recognised framework that provides an understanding of what an ICT role entails and clearly identifies the skills required. It consist of seven generic business skills that form SFIA’s seven levels of responsibility (see Table 2.5, p. 24) ranging from a starter to senior IT manager and world-leading technologist. The content categories are: Strategy and Planning, Business Change, Solutions Development and Implementation, Service Management, Procurement and Management, and Client Interface. Each of these content categories is split into subcategories.The framework matches skills with a responsibility level and a content category. It diagnostically defines the professional skills required in jobs and roles providing a clear differentiation between the requirements at different levels of responsibility (SFIA, 2011).

Table 2.4 – SFIA Levels of Responsibility and Generic Definition (SFIA, 2011)

Level of Responsibility	Summary of Generic Definition
1	Follow
2	Assist
3	Apply
4	Enable
5	Ensure/Advise
6	Initiate/Infuence
7	Set strategy, inspire and mobilse

SFIA provides guidance on the best progression routes for career development. IT professional capability is a blend of professional skills, behavioural skills and knowledge, which in its total capacity is validated by experience and qualifications. According to SFIA, “while qualifications certify elements of skill or knowledge, experience provides a practical demonstration of capability.” (SFIA, 2011, p. 8).

Database administration and Information analysis are among the professional IT skills defined by SFIA. Database administration as broadly defined by SFIA is “The installation, configuration, upgrade, administration, monitoring and maintenance of physical databases.” (SFIA, 2011, p. 38). Information analysis is defined as “The validation and analysis of information, including the ability to discover and quantify patterns in data of any kind, including numbers, symbols, text, sound and image.” (SFIA, 2011, p. 16). According to SFIA, techniques pertinent to Information analysis include statistical and data mining or machine learning methods, for example, rule induction, artificial neural networks, genetic algorithms and automated indexing systems. Based on their definitions, these skills can be tagged to particular roles; database administration is the skill for a DBA role while Information analysis is the skill for data scientist or data analyst role. According to Accenture,

> Data scientist is the most common term for the often PhD-level experts who operate at the frontier of analytics, where data sets are so large and the data so messy that less-skilled analysts using traditional tools cannot make sense of them (Accenture, 2013, pg 3).

The SFIA framework assigns different levels of responsibility and accountability to both skills expressed in terms of Autonomy, Complexity, Influence and Business skills. Database administration has a lower start and end level than Information analysis, which clearly indicates the disparity in core competencies required for both skills. According to the framework, database administration belongs to the service operation subcategory under service management catergory while Information analysis belongs to the Information strategy subcategory under strategy and architecture category (Table 2.5)

Table 2.5 – Comparison of database administration and information analysis skills (SFIA, 2011)

Skill	Subcategory	Category	Level of Responsibility (Low)	Level of Responsibility (High)
Database administration	Service operation	Service management	2	5
Information analysis	Information strategy	Strategy and architecture	3	7

It can be argued that both categories (service management, and strategy and architecture) belong to different functional areas within a functional organisational structure in which employees with similar knowledge and skills are grouped toghether.

Research carried out by George and Glady (2013) identified two important sets of success factors for creating a sustainable Business analytics practice within an organisation. These are:

1. How to develop the data scientists' skill (4 core skills).
2. How to organise the data scientists' management dynamics within the company (4 organisation dynamics).

George and Gladdy explained that despite the importance of skills and personality development, consideration should be given to the data scientists' management dynamics within the organisation and its impact on the business analytics practice. In their attempt to identify the organisational conditions, skills and capabilities required for developing a sucessful sustainable business analytics practice, the authors proposed a competence framework (Figure 2.6) comprised of the following four core skills:

1. Analytical - quantitative skills, modeling capability, methodology, performance management.
2. Technical - data management, database design, IT programming, business intelligence tools.
3. Business - understanding of the business as such, business culture, how to make the recommendations actionable, relevancy, flexibility, urgency.
4. Communication - being able to express complex ideas in simple terms, good interaction with decision makers, right format to the right audience, emotional intelligence, etc.

George and Glady (2013) suggests that in time, data scientists will need to develop their leadership and project management skills to complement these four core skills.

Figure 2.6 – 4 by 4 Analytics Framework (George and Glady, 2013)

The authors highlight the challenges in managing data scientists' talents and developing their skills. They explained that the right balance between education,experience and exposure must be achieved to maintain motivation. Most data scientists have advanced master or PhD levels (Accenture, 2013, George and Glady, 2013), and their training are in areas of specific interests and skills. George and Glady explained that their proposed skills framework will help overcome the challenges associated with data scientists career transitions from specialised roles to more generalist jobs by developing them on the four core skills. They also stressed the importance of nuturing the data scientists' business, communication and interpersonal skills and that successful business analytics teams usually consist of a mix of data scientists from different backgrounds.

In a study carried out by Accenture (2013), it was revealed that one reason for the shortage in data scientists is that data scientists require a rare blend of skills. Accenture explained that data scientists must have the requisite technical skills, such as advanced statistical and quantitative methods and tools, as well as new computing environments, languages and techniques for managing and integrating large datasets (Accenture, 2013). They are required to have "industry knowledge and business acumen to create models and solve real-world problems, and they need excellent communication and data visualization abilities in order to explain their models and findings to others." (Accenture, 2013). Accenture identified the follwing eight skills required for data scientists:

1. Advanced analytics.
2. Business acumen.
3. Communication & collaboration.

4. Creativity.
5. Data integration.
6. Data visualization.
7. Software development.
8. Systems administration.

According to Accenture, it takes years of training to become a data scientist and many organisations are unable to find people who possess all the eight skills and abilities of data scientists. As an alternative, Accenture proposed a framework for developing a data science team (Figure 2.7), which consist of individuals who as a group possess all the skills of a data scientist but lack them individually; more like dividing the labour of a data scientist. The proposed data scientist team consist of the following roles:

- Systems architect.
- Quantitative analyst.
- Business analyst.
- Visualization designer.
- Software engineer.

In some companies, these roles may already exist under different job titles.

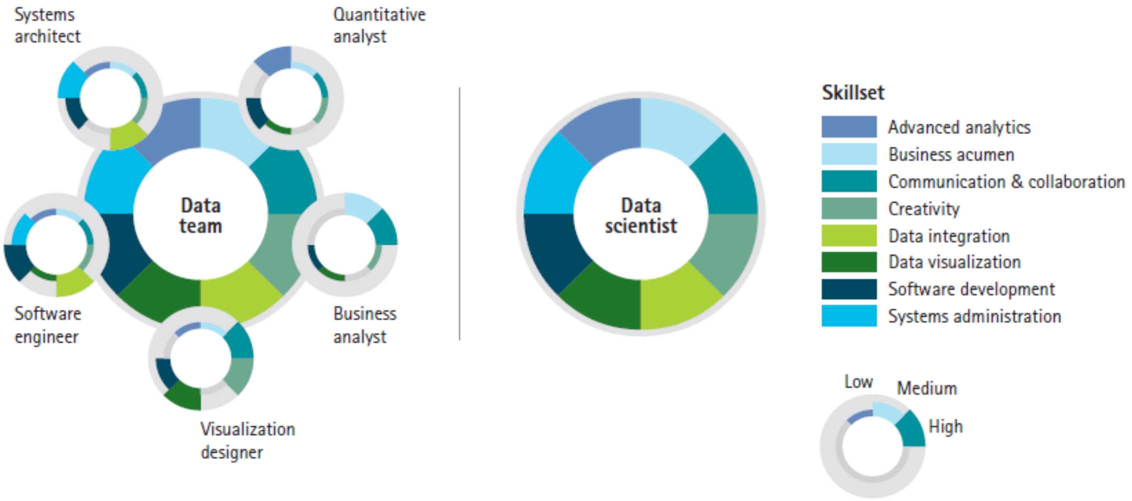


Figure 2.7 – The 8 Skills of Data Scientists (Accenture, 2013)

2.8 Conclusion

To conclude, big data has become popular in the IT world with different opinions and formal definitions of the term.

As the world becomes more information driven, challenges arise on how to harness and extract value from the growth in data as traditional systems have proved inadequate to deal with the exponential growth in data. New technologies have emerged and are rapidly changing the landscape. Big data has become synonymous with these new technologies, such as Hadoop Software and NoSql Databases. These new technologies require new skills in traditional DBAs, developers and other IT Professionals to manage the sophisticated systems as there are huge differences between the systems.

It is undeniable that a company’s data is its most strategic asset, quite expensive to manage in terms of the cost of technology and talent. However, the overall value, such as the competitive edge big data brings to an organisation, makes it is a very worthy investment. Organisations that fail to build a competency around it are more than likely to be at a disadvantage (Economist Intelligence Unit, 2011).

3 Methodology and Fieldwork

3.1 Introduction

Saunders et al. (2009) defines research as a systematic, methodological and ethical process of enquiry and investigation undertaken to find out things in order to increase knowledge, and refers to methodology as the theory of how research should be carried out. O'Leary (2004), describes methodology as the framework of paradigmatic assumptions by which research is carried out.

This chapter presents the research methodology adopted for this research. It outlines the background of different methodological approaches considered and the rationale for the approach taken for this dissertation. It describes the implementation of the research strategy, which entails the method of data collection. The Chapter discusses the skills framework proposed in this study and presents an overview of ethical considerations. The chapter concludes with a summary of the main points of the chapter.

3.2 Research Philosophy

According to Saunders et al. (2009), research philosophy contains assumptions that are directly linked to the way a researcher views the world. These assumptions dictate the research strategy and methods used, which form part of the strategy. In essence,research is carried out based on underlying philosophical assumptions; that is why it is necessary for the researcher to understand the different combinations of assumptions of research methods, in order to adequately conduct and evaluate research.

According to Flowers (2009), management research has three key paradigms namely, positivist, interpretivist and realist. These paradigms constitute the basis on which other research epistemological philosophies are formed. Saunders et al. (2009) identified four research philosophies: positivism, realism, interpretivism and pragmatism. This section briefly discusses three research philosophies considered for this study: positivism, interpretivism and pragmatism.

3.2.1 Positivism

Positivists consider reality to be stable, observable and objectively described without meddling with the current occurrences being studied. This philosophical position is characterised by hypothesis testing derived from present theory. It postulates that knowledge is only valid based on observations of external reality.This position is based on values of reason and is focused mainly on facts obtained from experience or direct observation, which are empirically measured by means of quantitative methods, such as surveys and experiments, as well as the belief that the development of theorectical models can be generalised (Flowers, 2009). This approach is generally linked with natural science research and endorses the idea of empirical testing in order to verify or invalidate hypotheses, as well as create new theory to be further tested (Greener, 2008).

Positivist researchers, according to Edirisingha (2012), proceed in a structured and controlled manner by first ascertaining a research topic, coming up with an appropriate research question and hypotheses and then implementing a research strategy, whereby statistical and mathematical techniques are an integral aspect of the research method, whilst seeking objectivity by being detached from the participants of the research in order to remain emotionally neutral. This is the chosen philosophy adopted for this research as it endeavours to test hypotheses generated from theory using purely quantitative methods.

3.2.2 Interpretivism

Interpretivism in sharp contrast, views knowledge differently from Positivism. Walsham (2009) briefly describes interpretivism as:

> interpretive methods of research start from the position that our knowledge of reality, including the domain of human action, is a social construction by human actors and that this applies equally to researchers. Thus there is no objective reality that can be discovered by researchers and replicated by others, in contrast to the assumptions of positivist science. Intepretivism is thus an epistemological position, concerned with approaches to the understanding of reality and asseting that all such knowledge is necessarily a social construction and thus subjective (Walsham, 2009,p. 5).

The Interpretivist research approach tends to focus on principal naturalistic methods, such as case studies, that involve human interaction between researchers and their subjects and this becomes the building blocks for the construction of meaningful reality (McBride, 2005). They insist reality can only be fully understood through subjective interpretation and, intervention and reality, and the researcher cannot be separated. In a nutshell, interpretivisits carry out research on people not objects and use the qualitative approach to conduct their research.

3.2.3 Pragmatism

Saunders et al. (2009) suggest that the most important determinant of a research approach is the research question. This philosophical position applies a mixed methods approach to answering the research question. Pragmatists believe that it is possible to work with both research paradigms(qualitative and quantitative research) in a single study. This leads to the idea of a particular philosophy being thought of as a continum of another instead of a direct opposite. Pragmatism is therefore not committed to any one philosophy thereby giving researchers the freedom to use any technique or method connected with quantitative or qualitative research or both if the method best suits the research question, while also recognising that different approaches can be complementary (Europe, 2009).

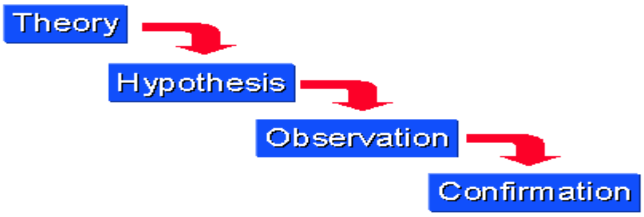
3.3 Research Approach

The approach and design of a research depends on the extent of clearity of theory within the background of the research and it is perfectly possible and advantageous to combine both approaches in a particular study (Saunders et al., 2009). Their difference is also highlighted in the way data collected is used either to test theories or to build theories (Bryman and Bell, 2007). This study will use the deductive research approach, which is one of the two main research approaches, the other being, the Inductive approach.

The deductive approach is affilated with the philosophy of positivism. It entails the testing of theory in research where data collected is used to prove or disprove hypotheses, in order to backup or refute theory from which the hypotheses is generated. This is informally referred to as the “top-down” approach, whereby research starts from the general to the more specific.

On the contrary, the inductive approach is associated with interpretivism and starts the inductive process from precise observations, detecting patterns, formulating tentative hypotheses, which will be explored and used to formulate theories. This is also informally referred to as the “bottom up” approach (Trochim, 2006). Figure 3.1 illustrates the two approaches that constitute most social and business research process.

Deductive Reasoning



Inductive Reasoning

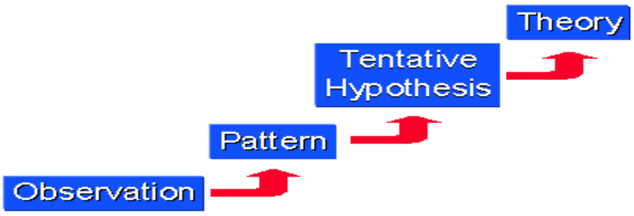


Figure 3.1 – Deductive And Inductive Research Approaches (Trochim, 2006)

3.4 Research Strategy

According to Saunders et al. (2009), the choice of a research strategy depends on how the strategy best answers the research question and set objectives. This choice is also guided by personal philosophical underpinnings, as well as practical considerations, such as available time and resources. Saunders et al. (2009) listed experiment, survey, case study, action research, grounded theory, ethnography and archival research as research strategies that could belong to the deductive or inductive approaches or classified into exploratory, descriptive and explanatory research. Saunders et al. (2009), explained that this classification is simplistic and stressed that no research strategy is more superior than the other and they are not mutually exclusive, as two strategies can be combined in a research based on how best they help answer the research question.

3.5 Research Methods

Saunders et al. (2009) refers to the way a researcher chooses to combine quantitative and qualitative procedures and techniques as a “research choice” that could be either mono-method or multiple methods. Research choice is also referred to as research design. A mono-method is described as the “use of a single data collection technique and corresponding analysis procedures to answer your research question” (Saunders et al., 2009, p. 151) and multiple methods is described as the use of more than one data collection method.

Quantitative Methods: The main aim of the use of quantitative methods is the development of generalisations that contribute to theory by enabling the researcher's understanding or explanation of a phenomenon. Quantitative research methods is described by O'Leary (2004) as the production of quantifiable data, presented numerically and analysed statistically that can be used to explain phenomena. In other words, this method involves the collection and analysis of data in numerical format through experiments and surveys.

Qualitative Methods: This method involves discovering and unconvering theories or patterns to explain a phenomenon by use of non-numeric data that can be gathered by several means, such as case studies, in-depth interviews and focus groups. Many researchers argue that the distinctions between the methods is not a hard and fast one and agree both can be used in a single or particular research project (Bryman and Bell, 2007).

3.6 Research Methodology used for this Research

There is growing consensus amongst the research community that to improve the quality of research, a combination of methods is the best approach with the acknowledgment that no single methodology is better than the other (Kaplan and Duchon, 1988). Gable (1994) acknowledges that the value of combining different methods has received attention in information systems research. However, a mono method was chosen for this study and the rationale behind the choice is explained in subsequent sections.

3.6.1 Philosopy

A positivist position was adopted in this research. This research is based on the technical knowledge of IT professionals in particular, DBAs who are involved in the storage, maintainance and administration of “big data”, as well as other business related functions that benefit from the use of big data technologies. The data required in this research can only be obtained from a wide range of job functions therefore, it is only feasible to gather this data quantitatively by means of an online survey. In this study, data will be obtained objectively rather than subjectively, therefore interpretivism will not be used because it studies participants through reflection or sensation making it inadequate to address the research question due to the large sample population involved. Moreover, critical to the interpretist view is the belief that insights into the complexity of business and management research is lost if such complexity is reduced to generalisations (Saunders et al., 2009, p. 115).

This dissertation aims to determine the impact of big data on the skills of DBAs based on unbiased responses from participants in different organisations. The positivist philosophy is therefore warranted as the details regarding the impact is measured objectively using quantitative means.

3.6.2 Approach

This research adopted a deductive approach due to the relationship between theory and collected data in this study. A big data skills framework was generated from literature (section 2.7,Skills Framework, p 23 -28) and hypotheses will be tested in accordance with the tenets of the deductive approach. This approach is indicative of the philosophy chosen for this study.

3.6.3 Strategy

Several strategies were considered and eliminated due to different reasons. Semi structured interviews with stakeholders were considered. However,

> semi-structured or in-depth interviews cannot be used to make statistical generalisations about the entire population where this is based on a small and unrepresentative number of cases (Saunders et al., 2009, p. 327).

A case study was considered for this research, however, it has disadvantages in relation to this study, which include: it is time consuming, subjective and narrow focused thereby rendering generalisation impossible. The option to use secondary data was sought but later abandoned due to the lack of secondary data, such as journal articles, related to the research question.

Survey research strategy was adopted for this study. This research seeks to garner the opinion of IT professionals, in particular, Database Adminstrators and other big data stakeholers, in order to test hypotheses. Primary data collection by means of a survey is a cost effective method of collecting large amounts of data from a geographically dispersed population, thereby making it possible for the findings to be generalised. The data gathered can also be standardised, making comparison easy; it can be quanitatively or qualitatively analysed using descriptive and inferential statistics (Saunders et al., 2009).
However, the drawbacks of this strategy include the limited number of questions that can be asked, as too many questions can result in the survey being abandoned, filled in randomly, thus increasing the possibility of a low response rate (O'Leary, 2004). To prevent these from occurring in this study, the survey was designed in such a way that the

questions were clear and concise, and to the minimum but adequate enough to tackle the research question. In spite of these limitations, a survey was deemed as the appropriate strategy, as its strenghts far outweigh the drawbacks pertinent to this research. The survey link was posted to several big data and DBA online groups on Linkedin to counteract the possibility of a low response rate.

3.6.4 Research Method

In the absence of secondary data relevant to the research topic, the research strategy involved the collection of only primary data from DBAs and other big data stakeholders through an online survey. A mono-method was therefore adopted by applying only quantitative research to this study. An online survey was adopted for this research for its versatility, efficiency and ability to form generalisations..

Another reason is quantitative data collection methods by means of an online survey is beneficial as it is an efficient and convientent means of data collection in terms of cost and time as the sample population is geographically widespread and presumably have access to computers, as well as the survey can be taken conveniently in their own time.

3.6.5 Survey Population

> The target population is the group or the individuals to whom the survey applies. In other words, you seek those groups or individuals who are in a position to answer the questions and to whom the results of the survey apply (Kitchenham and Pfleeger, 2002, p. 17)

The target population consisted of ICT professionals particularly those who may have knowledge and experience with big data technologies and are conversant with the difference between both traditional RDBMSs and big data technologies. They understand the intricate details between both technologies in terms of skills and are in the best position to predict valuable skills for the future and manage their own careers (Joseph et al., 2010)

However, the survey was open to other big data stakeholders, such as IT Managers and other business functions, who have interactions with the team administering and managing big data environment. Their perspective was also taken into account as they are likely to be involved in the decision-making process to acquire big data technologies and manpower involved in administering big data technologies.

3.6.6 Survey Design and Content

The main mechanism used in this study was a survey administered and hosted online by “SurveyMonkey”, a web-based application that allows for advanced survey design and analysis. This online method was chosen to administer the survey due to the ease of distribution it provides, flexibility it gives the respondents and the simplicity of data analysis. During the process of survey design, the research question and target population were the primary considerations for decisions on what questions to include in the questionnarie with each question serving a clear purpose.

Many authors agree that it is quite difficult to design a good survey and that the survey should be designed in such a way that it would collect accurate data around the research question and objectives (Bell, 2005, Saunders et al 2009). The survey was designed based on the following seven recommendations by Leary (1995):

1. The use of appropriate terminology in phrasing the questions.
2. Questions should be as simple as possible, avoiding jargon and difficult words.
3. Assumptions about the respondents should be avoided.
4. Conditional information should precede the key idea of the question.
5. Double-barreled questions avoided or not included all together.
6. An appropriate response format should be choosen.
7. A pilot of the questionnaire should be carried out.

The survey was designed meticulously with the notion that the questions were to be put to busy people, therefore they were short and very precise without compromising the confidence that they would answer the research question. Participants also answered the same questions in the same order. Hussey J and Hussey R (1997) pointed out the importance of keeping the potential survey respondents in mind as this guides the level of complexity of questions. Closed-ended questions were deemed appropriate as they were used to retrieve the prescise information in the mininum amount of time. Open-ended questions were avoided as they can lead to analysis and coding difficulties (O'Leary, 2004).

The survey took approximately 10mins to complete, and consisted of 20 closed questions in the form of multiple choice, ranking and matrix type questions organised in four different sections:

Section1 provided participants information regarding the survey. A consent form and a definition of big data by the McKinsey Institute was provided for a uniform understanding of the term.

Section 2 evaluated the respondents knowledge of big data and its significance to the business or the organisation where it is implemented.

Section 3 focused on technical skills. Questions enquired about the structure of big data teams, which roles constitute a big data team in their organisations or if there was a data scientist role in the organisation. Next were technology based questions, such as what big data technologies they had experience with and what barriers they thought would impede its successful implementation, ranking technical skills needed, as well as the roles needed to ensure the successful implementation of big data. The next set of questions sought to gather information in the form of a Likert scale type question where respondents where told to agree or disagree with the role a DBA plays or can play in a big data team. This was used to the extent to which respondents agreed or disagreed with the Likert statements.

Section 4 covered participants demographics and company background details. Overall, questions 15 - 20 related to participants profile e.g educational,professional experience and participants' organisational profile, such as the size and industry of the organisation.

The information gathered will be assigned appropriate weightages in order to answer the research question and test the following hypotheses:

Hypothesis 1:	The DBA role is important for the successful implementation of big data initiatives.
Hypothesis 2:	DBAs need to upskill in order to support big data initiatives.

3.6.7 Pilot Study

A pilot study can be seen as a test run of the primary study. Testing the survey is essential because it refines and filters out poorly worded or misleading questions thereby increasing the effectiveness of the survey, data collected, and its validity and reliablity (Gillman B, 2000). Welman et al. (2005) and Cohen et al. (2000) explain the aim of testing the survey by means of a pilot study as:

1. The means of identifying ambiguous and unclear questions.
2. To spot flaws or errors in the questionnarie.
3. To monitor the time taken by respondents to complete the survey.

A pilot test was conducted using the non-probability purposive sampling technique. Ten individuals with knowledge of big data technologies were chosen to take part in the pilot test. This number of respondents according to Saunders et al. (2009) is the minimum number of persons to take part in a pilot study and should have similar background and ability as the target population. Following the pilot, the questionnaire was revised following the pilot test. Double-barreled questions were split into two differnet questions. The questionnaire was revised following the pilot test. Double-barrelled questions were split into two different questions.The timeframe was deemed appropriate as the respondents filled in the survey in less than 10 minutes, which was the time allocated to complete the survey.

3.7 Big Data Skills Framework

According to Finn et al. (2000, p. 14), “Research needs theory as a framework for analysis and interpretation, and theory needs research to constantly review/modify/challenge theoretical details.”. To fully understand the impact of big data Technologies on the skills of DBAs, a framework was sought to guide the research in order to ensure data analysis conducted addressed the goal of the research. It has been noted that the requirement for soft skiils, such as communucation and business skills, in combination with technical skills are on the rise for IT professionals (Hoboken, 2007).

Luftman cited by Pratt (2012) acknowlegded that it is a challenge getting the right people to analyse big data information. He pointed out that the best candidates would be those professionals who possess an “uncommon mix of skills”, which he stated is a mix of technical know-how and business knowledge with a blend of strong mathematical and statistical backgrounds. A big data skills framework adapted from George and Glady

(2013) and SFIA (2011) was therefore developed. This framework was used to design the survey and the results will be used in testing the hypotheses stated in section 3.6.6, Survey Design and Content, p. 38. The skills framework is divided into the following four categories:

1) Analytical Skills

Analytical skills can be described as the ability to articulate and solve problems, whilst making sensible and proportionate decisions that are based on available information. Such skills include problem solving and the ablility to think logically when gathering and analysing data, interpreting data that is easy to understand, as well as discovering productive ways to successfully accomplish tasks. In the big data environment this can be broken down into the following sub-skiils:

a) Mathematics

The basic understanding of numerical analysis,computational linear algebra and the ability to dig into matrix computation and analysis is essential as most big data mining applications use matrix computations as their central algorithms. It is therefore crucial that big data professionals are competent in mathematics to be able to create new algorithms and not just build on existing algorithms.

b) Programming/Scripting Languages

A clear understanding or better still deep technical expertise in one or more programming languages, such as Python,Java,Perl,Pig and C/C++, in order to write codes and test different hypotheses.

c) Statistical Analysis

A big data professional must be competent in administering popular statistical tools, such as R, SAS, SciPy and SPSS. This understanding in correlation with looking at data from different angles is a must have skill in the big data team.

d) Data Mining

Data mininig is multidisciplinary in the sence that it ties statistics,machine learning and database systems together. Digging deep into big datasets to find new and interesting patterns is definitely a principal feature in big data management.

e) Data Modelling

Data modelling involves creating, parsing and understanding data models with tools, such as ERWin, ORM Diagrams, Agile Data Modeling and CRC. Proficiency in one or more of these tools is valuable and definitely a sought after skill in a big data team.

2) Technical Skills

Technical skills can be described as the knowledge and ability to accomplish mathematical, engineering, scientific or computer-related duties, as well as other specific tasks (Investopedia, 2014). the context of big data, technical skills can be refered to as knowledge of core big data technologies,level of education and qualification attained.

a) Programming/Scripting Languages

A clear understanding or better still deep technical expertise in one or more of these programming languages, such as Python, Java, Perl, Pig, C/C++, in order to write code and test different hypotheses.

b) Relational Databases

Fundamentals in database design and knowledge of indexing, normalisation and other basic features surrounding SQL-based systems is a must have skill as being proficient in manipulating and accessing data stored in an RDBMS is beneficial and a good foundation. Knowledge of Cloudera, VoltDB (highly scalable horizontally distributed systems) is also an added plus.

c) Distributed Computing Systems and Tools

Solid experience in NoSQL platforms and software, such as Apache Hadoop, MongoDB, Cassandra, MapReduce, Hive, as well as any new packages that emerge are benefical to have and are predominant in the big data management field.

d) Education/Qualification

Today, qualified big data professionals especially data scientists are expected to have advanced degrees, such as an M.S or Ph.D, in a reckonable field.

3) Business Skills

In this instance, business skills refer to an in depth understanding of how big data technologies can benefit or impact the organisation positively.

a) Business Acumen

Having a deep understanding of the organisation and how it operates and understanding the organisation's data is very beneficial.

b) Visualisation

The ability to work with visualisartion tools, such as Flare, HighCharts, Google Visualisation API, Tableau, after it has been scrubbed and mined is a leading skill. As a story has to be told from the bulk of data to the business in a way and manner that it can be understood and utilised by the organisation.

4) Communication Skills

Communication as described by Luftman and Kempaiah (2007, p. 166) is:

> the exchange of ideas, knowledge, and information between IT and business organisations, enabling both to clearly understand the company's strategies, plans, business and IT environments, risks, priorities, and how to achieve them.

In this case, communication skills refers to the ability to effectively and efficiently relay information or communicate with other teams in the IT department and other business stakeholders.

Communication

These are crucial skills to have as the ability to spend hours buried in data analysis and mining and then attempting to sell ideas to business users in the enterprise is very important and cannot be over emphasied.

3.8 Research Ethics

Ethics is described by Blumberg B et al. (2005) as the proper behaviour accorded to those who take part in a research study and are recommendations for responsible conduct during research. Gray (2009) pointed out that the following ethical considerations should be followed in any research process:

- The informed consent of all participants is acquired.
- Perception and harm to participants in the study should be avoided.
- The privacy of all participants should be respected.

In line with the Research Ethics Committee guidelines, the purpose and aim of the study was explained to the respondents before they gave their consent to participate. Participation was anonymous and voluntary .Participants were advised they could redraw at any point in time.

Ethics approval was sought and obtained in April 2014 from the School of Computer Science and Statistics Research Ethics Committee prior to distributing the survey. The following documents were submitted to the Ethics Committee:

- Informed consent form for organisations.
- Survey invitation.
- Completed ethics forms.
- Research proposal.

3.9 Conclusion

This chapter described a detailed account of the methodology used in this research and its limitations. The research was of a positivist nature with the use of a survey as the research strategy.

Despite the fact that there are few materials concerning the impact of big data technologies on the skills of DBAs, a big data skills framework was adapted from George and Glady (2013) and SFIA (2011). The chapter concludes with the ethical considerations considered and methodology limitations in this research. Chapter 4 will present the analysis and interpretation of the data collected from the survey.

4 Findings and Analysis

4.1 Introduction

This chapter presents the findings and detailed analysis of the empirical data collected from the online survey in order to answer the research question, sub-questions and test the hypotheses stated in the previous chapter. To effectively shed light on the topic, the survey was open to participants from IT and business backgrounds who from the IT perspective: are knowledgeable and familiar with big data technologies, and from the business perspective: interact with big data teams.

The chapter begins with a description of the response rate and data analysis. This is followed by a detailed description of the survey findings with each sub section detailing the method of analysis and subsequent findings.

4.2 Response Rate and Data Analysis

“Response representativeness is more important than response rate in survey research. However, response rate is important if it bears on representativeness.” (Cook et al., 2000, p. 821).

The link to the survey was distributed to colleagues and posted to online groups on LinkedIn made up of DBAs, IT professionals and other big data stakeholders, such as managers in the IT and business departments, who may be familiar with big data technologies. A target sample population size of 100 respondents was envisaged. However, despite the large membership of the groups and the extensive usage of personal contacts, there were 69 responses to the survey; the response rate to the survey was therefore 69%. Also, due to the poor representaton of some job roles and industries in the survey, the decision was made to combine job titles into two broad categories: technical and managerial roles, and industries combined into IT and non-IT industries.

The survey was open for four weeks from the 11th of April 2014 to the 12th of May 2014 and received a total of 69 responses. Of the 69 responses,16 were deemed invalid for the following reasons:

- Did not agree to partake in the survey.
- A majority of the questions were skipped.

As participants had the option to skip questions as part of the requirements of the ethics committee, not all questions were answered. The data was first analysed using SurveyMonkey's in-built statistical tool. Following the initial analysis, the 53 valid responses were exported to an excel spreadsheet for further analysis. A statistical calculator from StatPac Inc was used to test the results for statistical significance. The analysis and interpretation of the data gathered will ultimately provides answers to the following research question and sub-questions, and will prove or disprovbe the hypotheses tested in this study.

Research Question: What is the impact of big data technologies on the skills of DBAs?

Sub-questions:

Is the DBA role crucial to the successful implementation of big data initiatives?

Will the DBA role need to evolve into a hybrid role in order to support big data initiatives?

Will DBAs need new skills in order to fit in a big data team?

Hypotheses:

H1: The DBA role is important for the successful implementation of big data initiatives.

H2: DBAs need to upskill in order to support big data initiatives.

4.3 Survey Findings

The survey was divided into the following four sections for analytical purposes:

1. Participants Profile.
    a) Prior knowledge of participants.
    b) Experience.
    c) Education.
    d) Occupation.
    e) Organisation.
    f) Location.

2. Importance of big data to organisations and the barriers to its successful implementation.

3. Significance of the DBA role.

4. Skills and Roles required to build a big data team.

4.3.1 Participants' Profile

The profile of participants was broken down into the following sub categories:

- Prior knowledge of emerging technology.
- Experience.
- Education.
- Occupation.
- Organisation.
- Location.

➢ Prior Knowledge of Emerging Technology

Participants were asked the following questions:

How familiar are you with Big Data Technologies?

Respondents were asked to indicate their familarity with big data technologies. Of the 53 respondents eligible for further analysis, 90% indicated familarity ranging from "Slightly familiar" to "Extremely familiar". Interestingly, most DBAs indicated that they were "Moderately familiar" with big data technologies. A minority of respondents indicated that they are "Not at all familiar" with big data technologies.

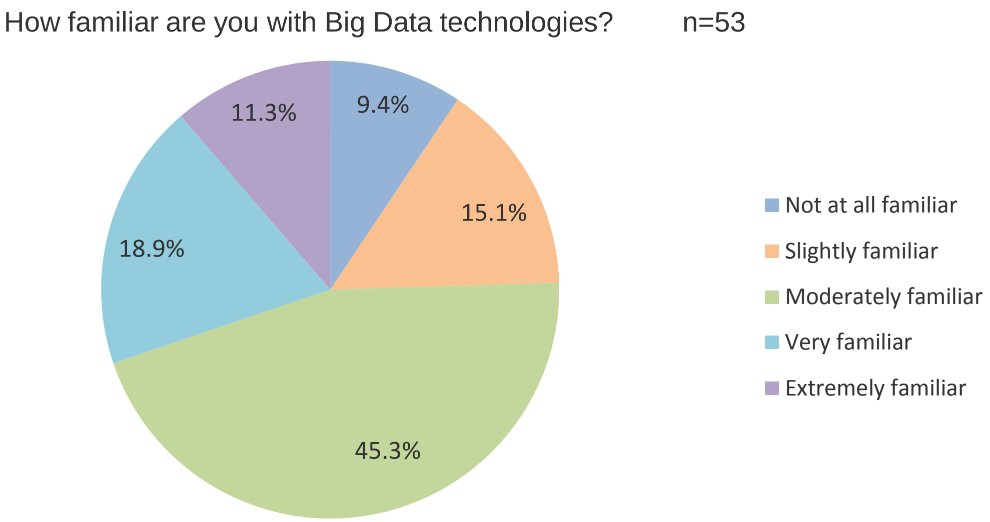


Figure 4.1 – Participants familiarity with big data technology

Responses to the question: “*How familiar are you with big data technologies*” was broken down into Technical and Managerial categories. From the data in table 4.1, a majority of respondents indicated familiarity with big data technologies ranging from “Slightly familiar” to “Extremely familiar”.

Table 4.1 **–** Familiarity with big data technology

	Not at all familiar	Slightly familiar	Moderately familiar	Very familiar	Extremely familiar	Response Count
Technical role	3	6	17	6	5	**37**
Managerial role	2	2	7	4	1	**16**
Total no. of respondents	**5**	**8**	**24**	**10**	**6**	**53**

A 5-point rating scale was applied to this question with weights ranging from 1 to 5 in the direction “Not at all familiar” to “Extremely familiar”. The average weighting was 3.08, thus indicating that on average, respondents are “Moderately familiar” with big data technologies.

How would you rate your interest in big data Technologies?

Participants were asked this question in order to find out their overall level of interest in big data technology. There were 53 responses to this question. Of the 53, 52 indicated interest ranging from “Slightly interested” to “Extremely interested”. Respondents with job titles “Database Administrator” and “Developer” indicated the highest level of interest.

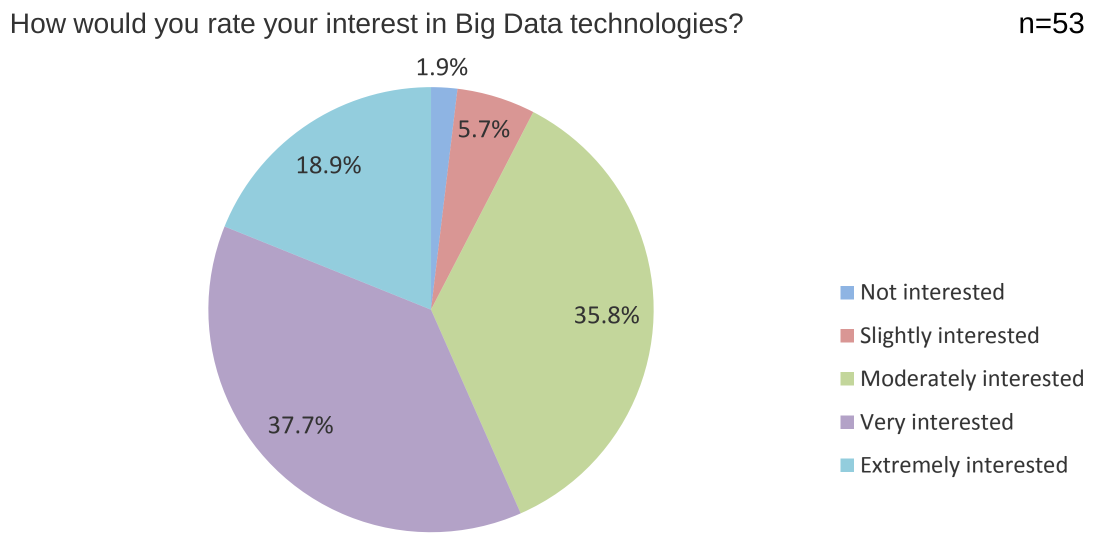


Figure 4.2 – Participants interest in technology

Like the preceding question, a 5-point rating scale was applied to this question with weights ranging from 1 to 5 in the direction “Not at all familiar” to “Extremely familiar”. The average rating was 3.66, thus indicating that, on average, respondents are “Moderately familiar” with big data technologies.

- Experience

How long have you been working in your current profession?

There were 53 responses to this questions. Of the 53, 37.7% (20 respondents) had over 10 years experience in their current profession, 17% (9 respondents) had between 1 – 3 years experience, 41.5% (22 respondents) had between 3 – 10 years experience, while 3.8% (2 respondents) represents had less than one year experience. These figures indicate that a majority of the survey population is made up mostly of participants with substantial years of experience in their current profession.

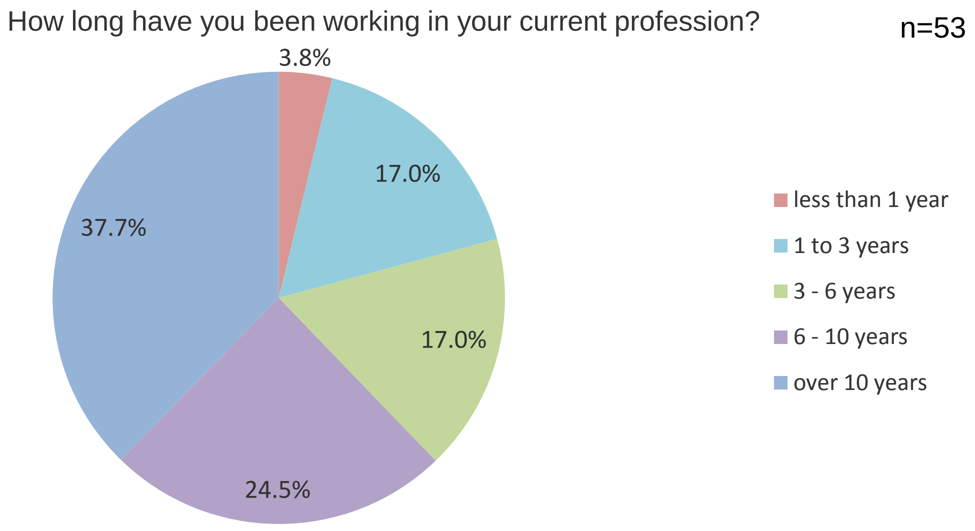


Figure 4.3 – Participants experience in current profession

- Education

What is the highest level of education you have completed?

This question sought to capture the highest level of education attained by each respondent based on their profession. As Figure 4.4 shows, of the 52 responses to this question, 50% (26 respondents) had a Master's degree, 3.8% (2 respondents) had a Doctorate degree, while 46.1% (24 respondents) a Bachelor's Degree, Professional Certificate or an Associate Degree. Further analysis using cross-tabulation to cross-reference "educational level" by "job title" revealed that a majority of the Masters Degree holders were DBAs.

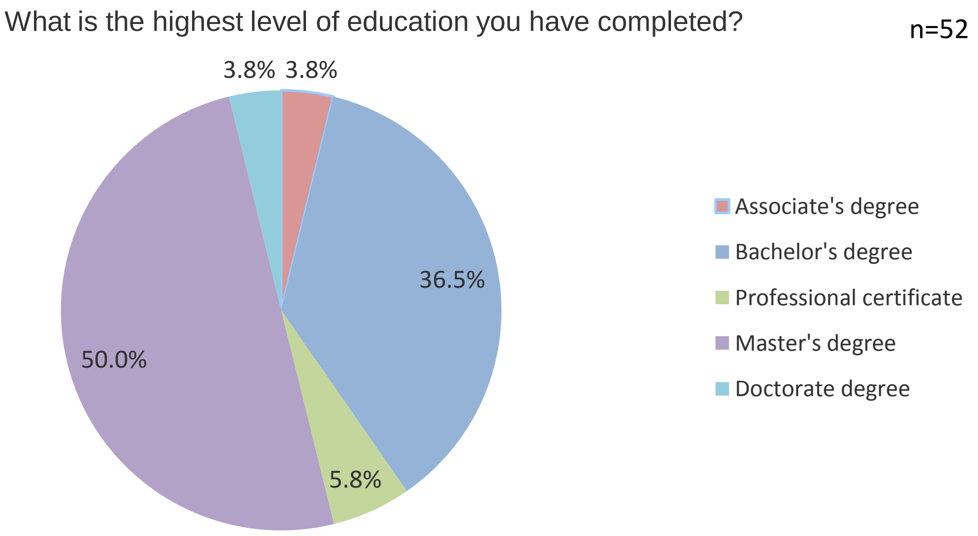


Figure 4.4 – Educational level of participants

➢ Occupation

What is your job title?

The survey was targeted at DBAs and other big data stakeholders. As Figure 4.5 shows, a majority of participants (69.81%, 37 respondents) were in technical roles, most of whom are DBAs and Developers, while 30.19% (16 respondents) were in Non-technical roles.

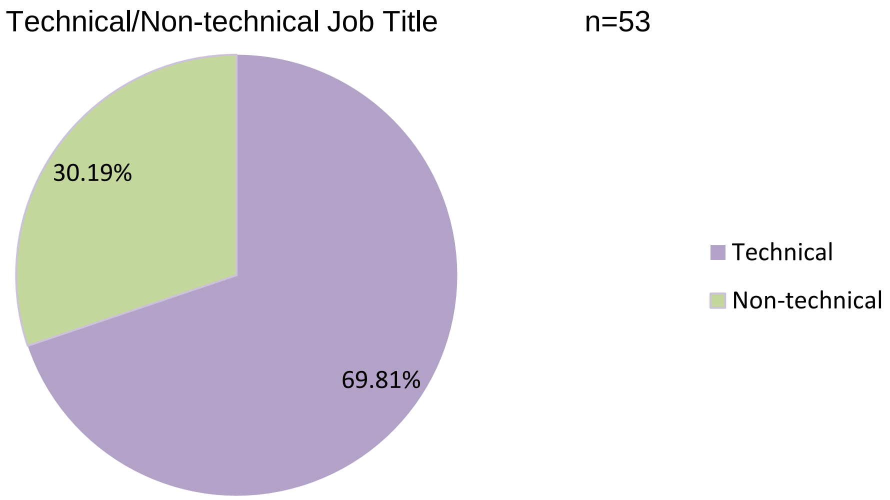


Figure 4.5 – Job titles of participants

The non-technical roles consisted of participants on the managerial level as shown in Table 4.2.

Table 4.2 **–** Participants Managerial job title

Managerial Job title	Response rate
Chief Technical Officer	5.67%
Chief Information Officer	3.77%
IT/IS Manager	9.43%
Project Manager	9.43%
Director	1.89%

➢ Organisation

How many people are employed where you work?

This question was deemed important as participants' organisation in terms of employee numbers could impact the size of a big data team; the bigger the organisation the more complex in diversity the big data team is likely to be. As Figure 4.6 shows, a mjority of the 53 respondents indicated that they work in organisations with 1 to 500 employees.

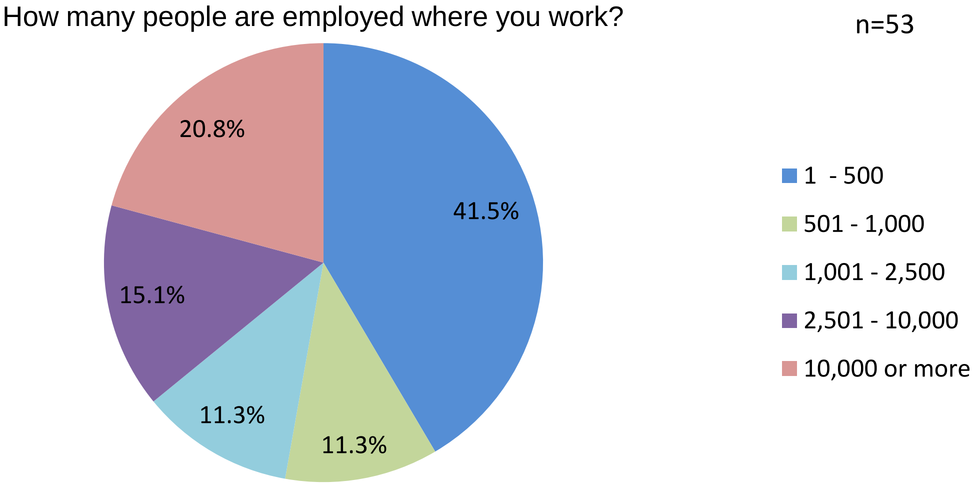


Figure 4.6 – Size of employees in organisation

Which industry do you work in?

A majority of respondents (58.50%) indicated they work in the IT sector; the IT services/Consulting industry accounted for a majority of the respondents who worked in the IT sector. There were 41.50% of respondents who indicated that they worked in the Non-IT sector, in industries, such as the manufacturing, energy and financial services sectors. The "Other" option was chosen by 4 respondents who entered "public sector", "call center industry", "higher education" and "market research" in the text entry box provided.

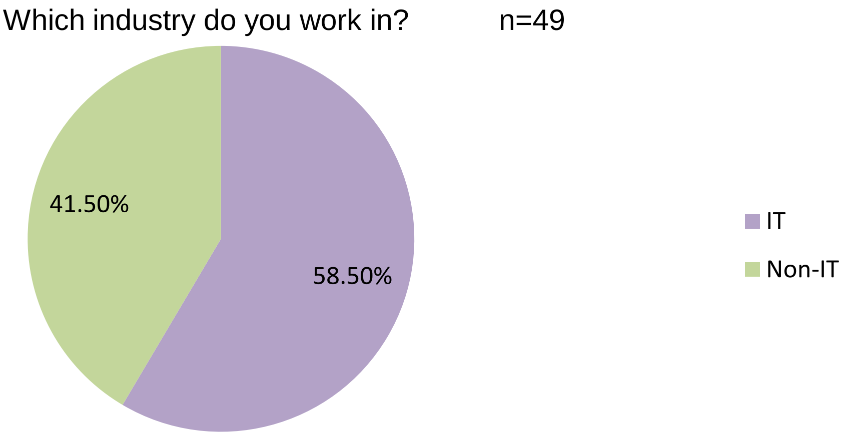


Figure 4.7 – Participants Industry

- Location

Which country do you work in?

There were 43 responses to this question. Of the 43, Ireland (56%) had the highest number of participants, while 44% of respondents were from other countries in Europe, Asia, and America (Figure 4.8).

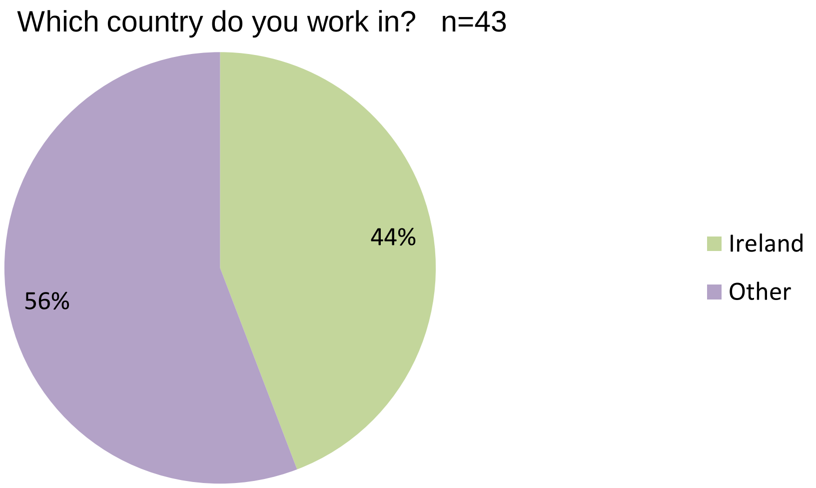


Figure 4.8 – Participants geographical location

4.3.2 Importance of Big Data to Organisations and the Barriers to its Successful Implementation

How important is big data to your business?

There were 52 responses to this question. Of the 52, 75% (39 respondents) indicated importance ranging from “Somewhat important” to “Extremely important”, while 25% (13 respondents) either did not know or indicated that it was not important to their organisation. When cross referenced against “participants industry” the IT Services/Consulting and telecomunication industries had the highest number of participants who indicated that big data was “Extremely important” their organisation.

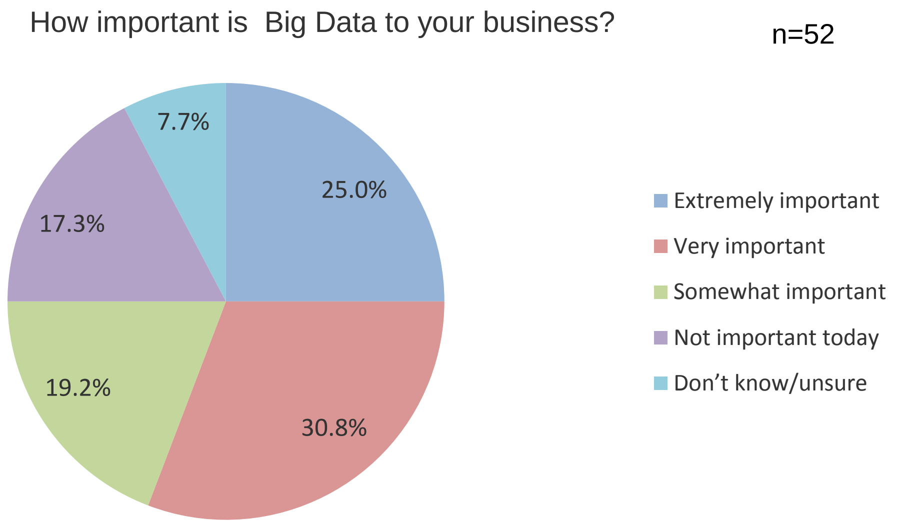


Figure 4.9 **–** Importance of big data to participants business

Is there currently a big data initiative where you work?

Participants were asked if there was a big data initiative in their organisation. From the data in the chart (Figure 4.10), it can be seen that of the 52 respondents who answered this question, 53.8% (28 respondents) indicated that their organisation currently has a big data initiative, 34.6% (18 respondents) indicated that their organisation did not have one, while 11.5% (6 respondents) didn’t know.

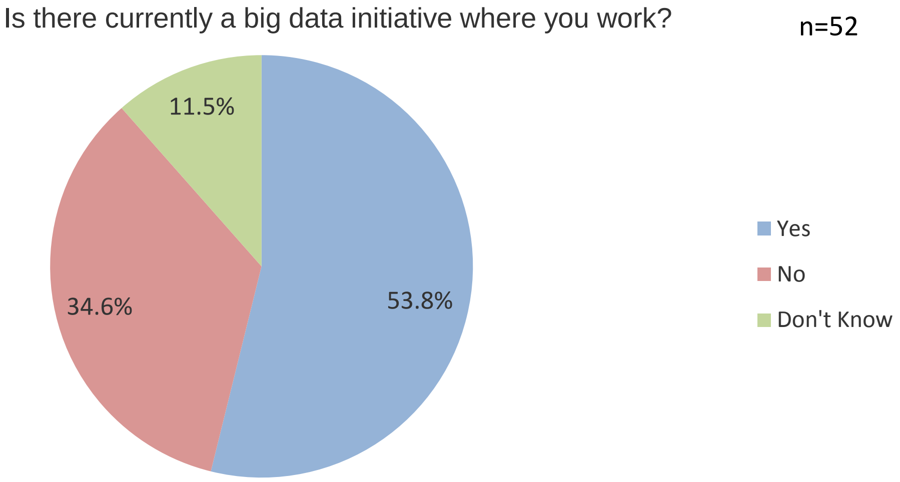


Figure 4.10 – Current big data initiative in participants organisation

When cross referenced against industry as shown in Table 4.3 , the IT industry comprising of the IT services/consulting and telecommunications industries were the most popular industries with big data initiatives; akin to results published by e-skills UK on behalf of

SAS UK and Ireland that identified IT firms as having the highest growth in the implementation of big data technologies in Ireland (e-skills UK, 2013).

Table 4.3 – Current big data initiatives by industry

Industry	Yes	No	Don't Know	Response Count
IT	17	12	1	**30**
Non IT	10	3	5	**18**
Total no. of respondents	**27**	**15**	**6**	**48**

Who owns or drives the big data initiative?

There were 29 respondents to this question due to the skip logic applied to the previous question, which automatically skipped this question for the 24 participants who indicated that their organisation did not have a big data initiative or that they did not know if big data was implemented in their organisation. As Table 4.4 shows, of the 29 respondents who answered this question, an overwhelming 21 indicated that the big data intiative in their organisation is driven by the collaboration of Business and IT, 5 respondents indicated that it is mostly IT driven with minimal business involvement, while 3 respondents indicated that it is mostly business-driven and with minimal IT support.

Table 4.4 – Big data initiative drive

Initiative drive	Response Count
Mostly business-driven with minimal IT support	3
Business/IT collaboration	21
Mostly IT driven with minimal business support	5
Total no. of respondents	**29**

On a scale of 1 to 6, with 1 being the highest and 6 the lowest, participants were asked to rate a list of barriers to the successful implementation of big data initiatives. Figure 4.11 offers a breakdown of the significant barriers identified in this research.

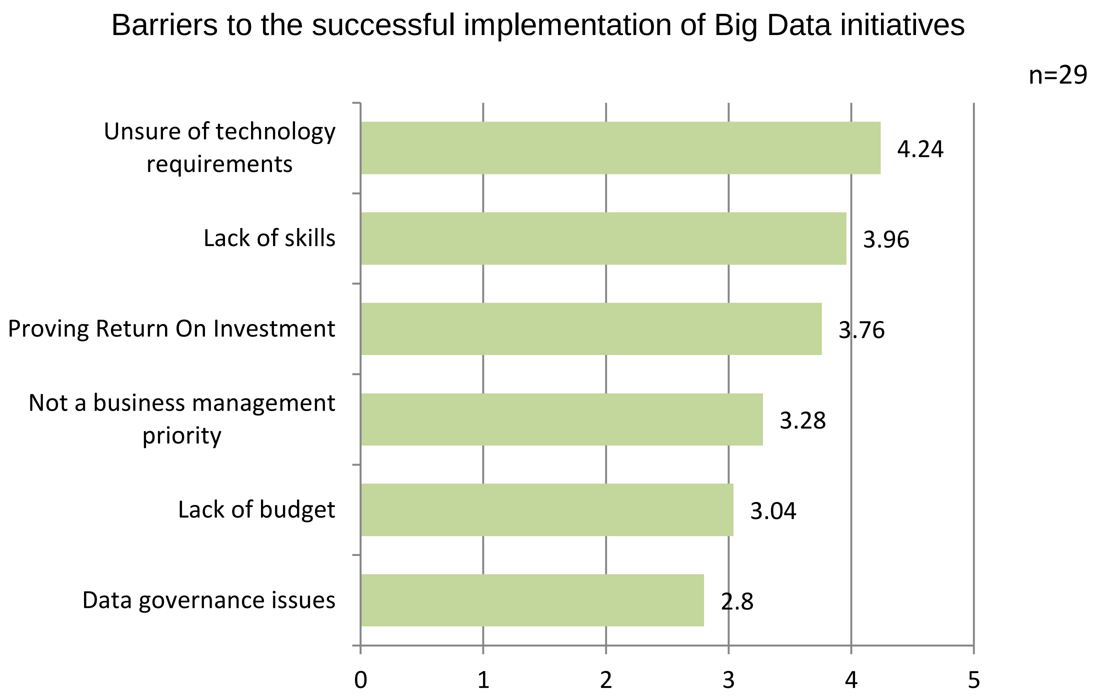


Figure 4.11 – Barriers to successful implementation of big data initiatives

These results not surprising, but interesting to see none the less, indicate that a majority of respondents chose “Unsure of technology requirements”, as the highest barrier to successful implementation of big data initiatives. Understandably, big data technologies have a wide array of vendor offerings that may make it difficult for organisations to determine which big data technology is most suitable for its business needs. Lack of skills was the second hightest barrier to the successful implementation of big data with an average ranking of 3.96. This supports the notion that there is a general lack of skills to properly administer and manage big data technology. Other barriers cited by respondents are: proving return on investment, not a business priority, lack of budget and data governance issues.

To gain in-depth insights into the use of big data technologies, participants were also asked to select which data processing/management technologies were in use or planned for adoption. Relational databases was the most selected technology, closely followed by Hadoop. The least selected were Column based DB and Dynamo as shown in figure 4.12.

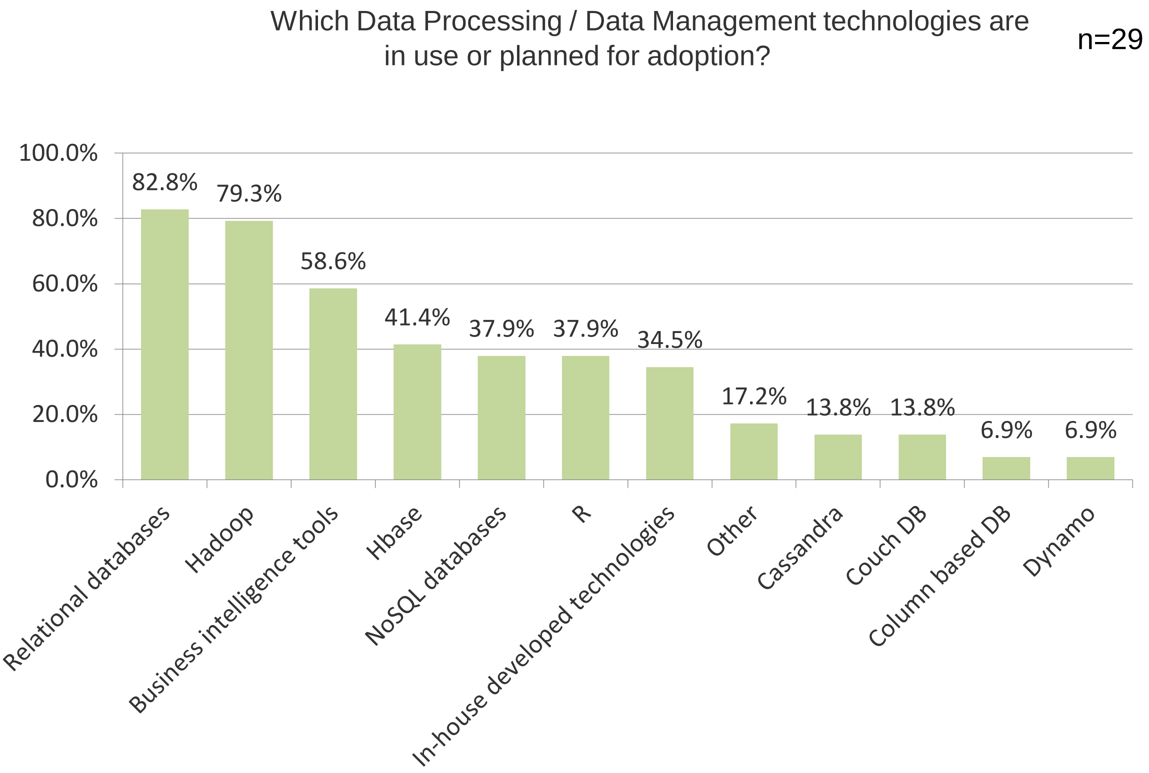


Figure 4.12 – Data Management Technologies Planned for adoption

4.3.3 Significance of the DBA Role

To adequately determine the significance of the DBA role in a big data team, participants were asked to indicate the extent to which they agree or disagree with each of the Likert statements in table 4.5. These statements constitute the hypotheses that will be tested in this study.

Hypothesis 1: The DBA role is important for the successful implementation of big data initiatives.

Respondents were asked to rate their level of agreement or disagreement with the following statements in Table 4.5.

Table 4.5 – First Hypothesis statements

	Question: To what extent do you agree or disagree with the following statements?
S1	The DBA role is gaining importance with the emergence of big data
S2	The DBA plays a critical role in big data

Of the 53 responsdents analysed, 19 indicated disagreement with statement S1 ranging from “Strongly disagree” to “Disagree”, while 34 expressed agreement ranging from “Agree” to “Strongly agree”. Similar proportions were also seen in S2 where 17 and 36 respondents indicated disagreement and agreement respectively.

The 5 response categories of the Likert scale shown in Table 4.6 were collapsed to three for further analysis. The “Agree” and “Strongly agree” response categories were combined into a single category and the “Disagree” and “Strongly disagree” into another. The “Neither agree” nor “Disagree” category was excluded from further analysis as it had no response resulting in two response categories. This allowed for further analysis with a t-test to determine the statistical significance of the result.

Table 4.6 – Results of first hypothesis statements

Statements	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total
S1	6	13	0	28	6	**53**
S2	5	12	0	26	10	**53**

A one-sample t-test between the percentages for statement S1 showed that the t-statistic was significant at the 0.05 critical alpha level, t(52)=2.148, p=0.0364.

A one-sample t-test between the percentages for statement S2 showed that the t-statistic was significant at the 0.05 critical alpha level, t(52)=2.796, p=0.0072.

The results were deemed statistically significant as the p-values are less than $\alpha = 0.05$. Thus the null (H_0) hypothesis that the DBA role is not important for the successful implementation of big data initiatives is rejected thereby supporting the alternate hypothesis that the DBA role is important for the successful implementation of big data initiatives.

Hypothesis 2: DBAs need to upskill in order to support big data initiatives.

Respondents were asked to rate their level of agreement or disagreement with the following statements in Table 4.7:

Table 4.7 – Second Hypotheis statements

	Question: To what extent do you agree or disagree with the following statements?
S1	The role of DBAs needs to evolve to handle big data
S2	The DBA role fits into the big data team
S3	DBAs need to upskill to support big data initiatives

Of the 53 responsdents analysed for each statement, 14 respondents indicated disagreement with statement S1 ranging from “Strongly disagree” to “Disagree”, while 39 respondents expressed agreement ranging from “Agree” to “Strongly agree”. For statement S2, 18 and 35 respondents indicated disagreement and agreement respectively, while for statement S3, 16 respondents expressed disagreement, 36 respondents indicated agreement and 1 respondent neither agreed nor disagreed.

The five response categories of the Likert scale shown in Table 4.8 were collapsed to three for further analysis. The “Agree” and “Strongly agree’” response categories were combined into a single category and the “Disagree” and “Strongly disagree” into another. The “Neither agree” nor “Disagree” category was excluded from further analysis as it had only one response resulting in two response categories. This allowed for further analysis with a t-test to determine the statistical significance of the result.

Table 4.8 – Results of second hypothesis statements

Statements	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Total
S1	3	11	0	31	8	53
S2	5	13	0	28	7	53
S3	8	8	1	27	9	53

A one-sample t-test between the percentages for statement S1 showed that the t-statistic was significant at the 0.05 critical alpha level, t(52)=3.898, p=0.0003.

A one-sample t-test between the percentages for statement S2 showed that the t-statistic was significant at the 0.05 critical alpha level, t(52)=2.459, p=0.0173.

A one-sample t-test between the percentages for statement S3 showed that the t-statistic was significant at the 0.05 critical alpha level, t(51)=2.968, p=0.0046.

The results were deemed statistically significant as the p-values are less than $\alpha = 0.05$. Thus the null (H_0) hypothesis that DBAs do not need to upskill in order to support big data initiatives is rejected thereby supporting the alternate hypothesis that DBAs need to upskill in order to support big data initiatives.

4.3.4 Skills and Roles required to Build a Big Data Team

To determine the skills and roles required to build a big data team, participants were asked specific questions. There were 29 responses to the following question due to the skip logic applied to the previous question:.

Is there a big data team where you work?

From the data in Table 4.9, it can be seen that 25 respondents indicated that they have a big data team established in their organisation, while 4 respondents either don't have or are planning to set up one.

Table 4.9 – Presence of big data teams in participants place of work

Is there a big data team where you work?	Response Count
Yes	25
No yet, but planning to set up one	1
No	3
Total no. of respondents	**29**

A clearer picture emerged when the respondents were categorised into IT and Non-IT sub sections as shown in Table 4.10 . A majority of the respondents who indicated that their organisation had a big data team worked for IT companies, such as IT services/consulting service and telecommunications.

Table 4.10 – Presence of big data teams in participants place of work by industry

	Yes	Not yet, but planning to set up one	No	Response Count
IT	16	1	1	**18**
Non IT	10	0	1	**11**
Total no. of respondents	**26**	**1**	**2**	**29**

Which of the following roles are in your big data team?

As a follow up question participants were asked to select the roles that constitute the big data team in their organisation. As Figure 4.13 shows, Developers and Data Scientists accounted for the most selected roles while enterprise data architects, DBAs and business analysts were the least selected roles.

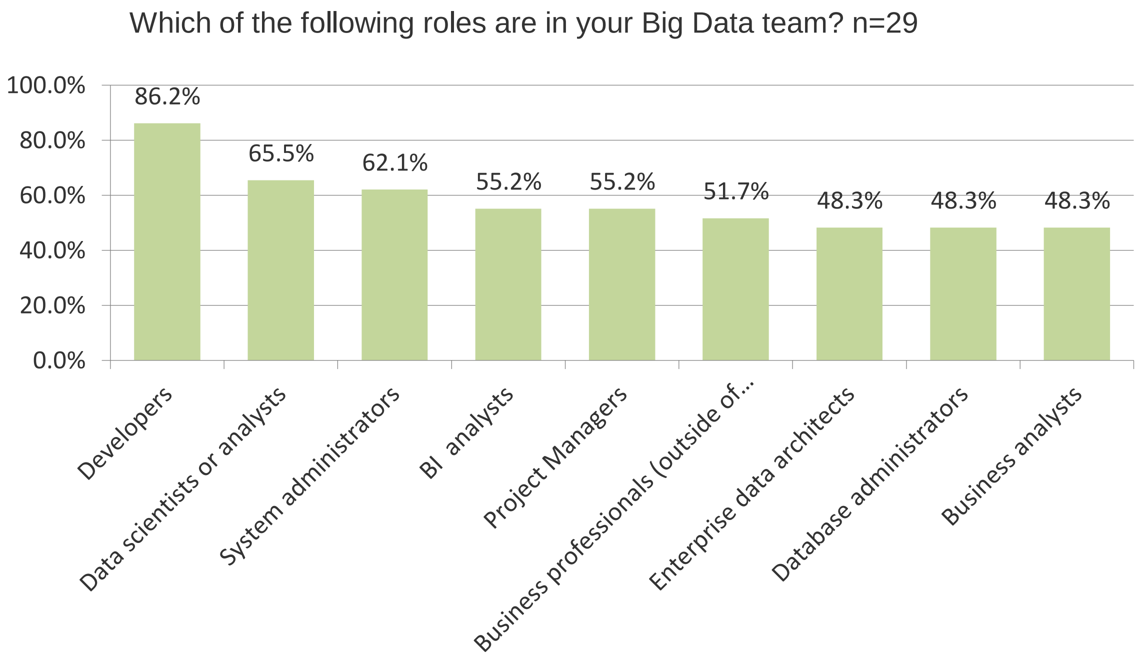


Figure 4.13 – Breakdown of big data roles in teams

Do you belong in the big data team?

This question sought to further capture respondents' level of exposure and knowledge of big data technologies. From the data in Table 4.11, 14 respondents indicated that they belong in a big data team, 10 indicated that they do not belong but interact with the team, while 4 indicated that they do not belong or even interact with the team. As Table 4.12 shows, with the exception of an IT/IS manager and CTO, all other job titles appear to belong in the big data team.

Table 4.11 **–** Percentage of participants who belong to a big data team

Do you belong in the big data team?	Response Count
Yes	14
No, but in my current role, I interact with the big data team	10
No	4
Total no. of respondents	**28**

Table 4.12 – Participants exposure to big data teams by job title

	Yes	No, but in my current role, I interact with the big data team	No	Response Count
IT/IS Manager	0	1	0	**1**
CIO	1	1	0	**2**
IT Consultant	2	1	0	**3**
DBAs	3	2	1	**6**
CTO	0	3	1	**3**
Developer	3	2	0	**5**
Project Manager	1	0	1	**2**
Systems Administrator	1	0	1	**2**
Analyst	3	0	1	**4**
Total no. of respondents	**14**	**10**	**4**	**28**

In your opinion, how important are the following skills for ensuring the success of a big data Intiative?

There were 52 responses to this question in which respondents were asked to rank in the order of importance the skills required for ensuring the success of a big data initative. From the data in the chart (Figure 4.14), it is apparent that respondents indicated that critical and analytical thinking is the most important skill for ensuring the successful implementation of big data initiatives, database administration was ranked 6th, while decision making was the least important.

.

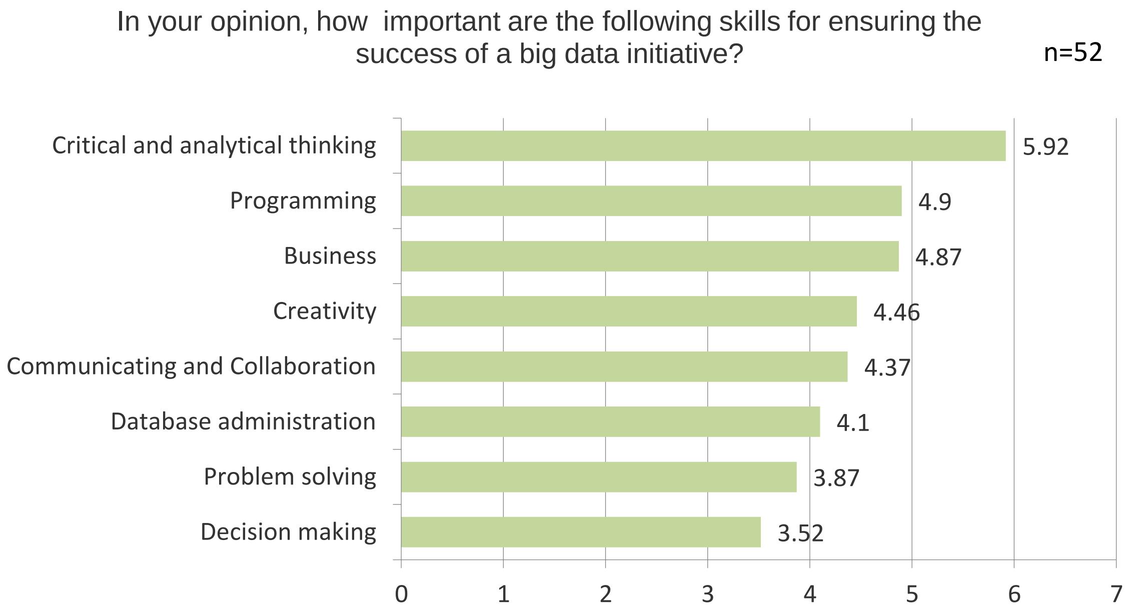


Figure 4.14 – Important skills for successful big data implementation

Table 4.13 – Skills ranked by number of respondents

Skill	Number of Respondents that ranked skill the highest	Number of respondents that ranked skill the lowest
Critical and analytical thinking	18	1
Programming	8	7
Business	9	6
Creativity	3	5
Communication and Collaboration	3	4
Database Administration	6	6
Problem Solving	2	5
Decision Making	3	18
Total no. of respondents	**52**	**52**

The skills in table 4.13 constitute the proposed skills framework for this study, which will be used to assess the impact of big data technologies on DBAs from a skills perspective. It is further broken down by job title as follows:

- Critical and analytical thinking: This was the most important skill identified, which had a weighted average ranking of 5.92. Eighteen respondents ranked this skill as the most important skill for ensuring the success of a big data initiative. The 18 respondents comprised of DBAs, developers, CTOs and IT/IS managers with only one respondent (project manager) ranking it as the least important skill to have.

- Programming: Programming skills with 4.90 average ranking was the second most important skill. Analysts and developers made up the majority of the 8 respondents that ranked programming as the most important skill while an IT/IS manager, CIO,3 DBAs, a CTO and a developer ranked programming as the least important skill.

- Business: Business skills was third with an average weighting of 4.87. Nine respondents ranked this as most important skill to have. Other stakeholders who have interaction with a big data team, such as the IT /IS manager, CTO and IT consultant, indicated that it was the most important skill to have. Four DBAs and 1 analyst ranked business skills as the least important.

- Creativity: Creativity had a weighting of 4.46 with one CIO and 2 DBAs ranking creativity as the most important. Three IT/IS managers and 2 systems administrators felt creativity was the least important skill.

- Communication and collaboration skills: Communication and collaboration with an average ranking of 4.37 had 3 respondents (1 analyst, 1 IT Consultant and 1 DBA) who indicated that it was the most important skill. Four respondents (1 IT consultant, 1 developer and 2 project managers) thought it was the least skill to have.

- Database administration: Database Administration had an average weighting of 4.10 with 6 respondents who ranked it as the most important skill to have. The 6 respondents included 3 DBAs, 1 IT consultant, 1 Developer and 1 Systems Administrator. Six respondents also ranked database administration as the least skill to have and they are: 1 CIO, 1 CTO, 1 IT Consultant and 3 Developers.

- Problem solving: Problem solving had a weighting of 3.87 and 2 respondents (Project manager and Developer) thought it was the most important skill. One IT consultant, 1 developer, 1 project manager, 1 systems administrator and 1 analyst made up the 5 respondents who indicated that problem solving was the least important skill to have.

- Decision making: Decision making had an average weighting of 3.52 with 3 respondents (1 Project manager, 1 analyst and 1 developer) who ranked it the most important skill. The 18 respondents that indicated that decision making was the least important skill to have were 4 DBAs, 3 IT Consultants, 5 developers, 3 analysts, 1 IT/IS manager,1 CTO and 1 project manager.

How important are the following roles to ensure the adoption and use of big data?

There were 51 responses to this question. As Figure 4.15 shows, data scientists or analysts were ranked the most important roles required to ensure the adoption and use of big data with an average rating of 6.18. This was followed by developer role with an average rating of 5.98. The DBA role comes in at the penultimate rank of 8 with a low average rating of 4.14.

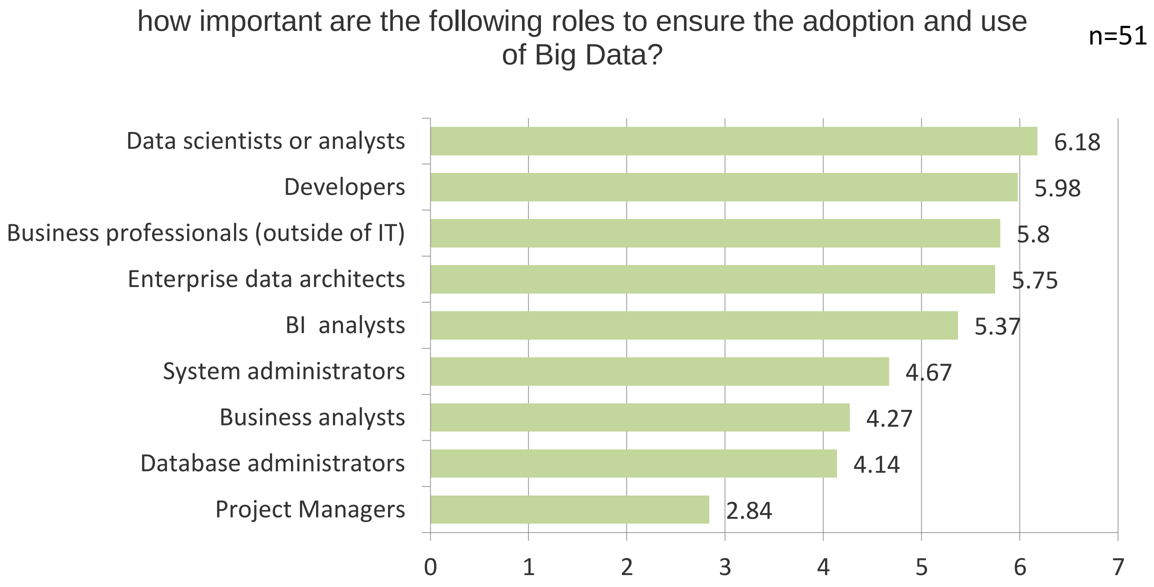


Figure 4.15 – Importance of roles to ensure the adoption and use of big data

4.4 Summary of Findings

This chapter presented the detailed analysis of the results of the data gathered by the research. As part of the overall survey findings, a detailed breakdown of the profile of participants was presented indicating a diverse mix of experienced and knowledgeable participants in technical and managerial roles; of these participants, a majority indicated that they interact with a big data team or were in a big data team. Futhermore, most of these participants were located in Ireland, with more than 50% employed in mid to large size organisations ranging from 501 to 10,000 employees.

A big data skills framework comprising of four parts was formulated and tested to determine the skills needed by DBAs to fit into a big data team. The findings revealved that critical and anlytical thinking topped the list of skills needed to successfully implement a big data initiative. As expected, the data scientist role topped the list of important roles needed for the successful implementation of big data initiatives.

The research question and hypotheses were used as a framework in order to guide data analysis. The findings of the study corroborate findings from the reviewed literature and support the hypotheses tested in this research. Chapter 5 highlighs the key conclusions of the research, discusses its limitations and recommends areas for future research.

5 Conclusions and Future Work

5.1 Introduction

This chapter discusses how the findings of the study answered the research question and sub-questions. It highlights the key findings and interesting new finding, and briefly discusses the contribution of this research to the existing body of knowledge. Subsequent sections discuss the limitations of the research and outlines recommendations for future research in this area.

5.2 Key Findings from the Research

From the findings of the data gathered in this research and the reviewed literature, the interpretation and understanding of the term big data is unambigous especially among data management professionals. DBAs were the most familiar and interested in big data technology, which is spread across different sectors of the economy and varying enterprises irrespective of size, particularly, the IT services/Consulting industries.

The findings of the research suggest that big data technologies are implemented in various industries as over 50% of participants indicated that their organisations currently have a big data initiative. Most of the participants that indicated that a big data team currently exist in their organisation work in the IT services/consulting services and telecommunications industries. They also indicated that big data is extremely important to their business.

The findings from this research, similar to findings of a study conducted by TDWI (2011), shows that lack of skills is one of the top barriers to the successful implemention of big data initiatives. It was chosen by participants as the second highest barrier to the successful implementation of big data initiatives after “unsure of technological requiremennts” (see Figure 4.11, section 4.3.2, Importance of Big Data to Organisations and Barriers to its Successful Implementation, p. 54).

The findings from the research show that relational databases topped the list of data management technologies planned for adoption, closely followed by Hadoop (see Figure 4.12, section 4.3.2, Importance of Big Data to Organisations and Barriers to its Successful Implementation, p. 55). These findings corroborate the findings of Villars et al. (2011), which suggest that a Hadoop environment and a traditional data warehouse environment run alongside each other in some organisations.

5.3 Answering the Research Question

The main aim of this research is to determine the impact of big data technologies on Database Administrators from a skills perspective. As part of this study, an online survey was conducted, the conclusions from the findings are summarised and disscussed in sub sections in order to adequately address the research question and sub-questions. The answers to the sub-questions will further guide the conclusions of the research and together answer the main reasearch question, which is centered on the importance of the role of a DBA in supporting big data initiatives. Following a comprehensive review of the literature, a skills framework was formulated; this skills framework can be used as a guide to ascertain the skills needed by DBAs in order to fit in a big data team (see section 3.7, Big Data Skills Framework, pp. 38-41). This framework, adapted from George and Glady (2013) and SFIA (2011) was divided into four sections namely Analytical, Technical, Business and Communication Skill, and used to develop the following hypotheses tested in this study:

H1: The DBA role is important for the successful implementation of big data initiatives.

H2: DBAs need to upskill in order to support big data initiatives.

The results presented in section 4.3.3 (Significance of the DBA role, pp. 55-58) suggest that the DBA role is important for the successful implementation of big data initiatives. This corroborates the report by Enterprise Strategy Group (2011) that suggest that Data Architects or DBAs are key to the successful implementation of data analytics projects as they bridged the gap between the physical storage of data and the business problem. The findings suggest that the DBA role as a data professional will evolve as DBAs become more involved in the implementation of big data initiaitives and develop their skills to become key players in big data teams.

The results of the second hypothesis support the findings of earlier research carried out on the evolutionary effect of big data technologies in the area of database administration by Simmonds (2013), which suggest that a new range of skills is needed by DBAs in addition to traditional database administration skills, so much so that the role of the DBA would have to evolve to handle big data technologies. The data in Figure 4.14 (see section 4.3.4, Skills and Roles required to Build a Big Data Team, p. 61) shows that critical and analytical thinking topped the list of important skills for ensuring the successful implementation of big data initiatives while decision making was the least important.

The findings from this study provides the following answers to the sub-questions, which ultimately answer the research question (see section 1.2, Research Question, pp. 2-3):

- The DBA role is crucial to the successful implementation of big data initiatives.
- The DBA role needs to evolve into a hybrid role in order to support big data initiaties.
- DBAs need new skills in order to fit in a big data team.

5.4 Interesting New Finding from the Research

Findings from this study and the reviewed literature (Accenture, 2013) indicate that the unique blend of skills and years of training required to undertake the role of a data scientist accounts for the shortage of data scientists Professionals . Although this is common with emerging technologies, the vast amount of skills that cut across different roles makes this quite exceptional. Hence organisations have resorted to building big data teams, which comprise of individual roles that have the skills outlined by Accenture (2013) (see Figure 2.7, section 2.7, Skills Framework, p. 27). Accenture (2013) proposed a data science team, which consist of the following roles:

- Systems architect.
- Quantitative analyst.
- Business analyst.
- Visualization designer.
- Software engineer.

This study revealed that although data scientists, developers and system administrators were the most prominent roles in big data teams, the DBA role was also surprisingly found to be part of big data teams in some organisations (see Figure 4.13, section 4.3.4, Skills and Roles required to Build a Big Data Team, p. 59). This new finding buttresses the need for this research and furture indepth research on the effect of big data technologies on the skills and tasks of data management professionals, particularly, DBAs.

5.5 Contribution to the Body of Knowledge

This reserch provides an insight into the impact of big data technologies on the skills of DBAs. As new forms of data become readily available and more questions are asked of data, it has become relevant for DBAs to develop big data skills in order to harness the oppourtunities that have arisen from the deluge of data in the data management industry .

The study provides a conceptual skills framework (see section 3.7, Big Data Skills Framework, pp. 38-41) that could be used by organisations who plan on implementing big data initiatives. It pinpoints the skills and roles needed to establish a big data team and the specific skills a DBA would need to fit into the team. The skills framework provides a starting point, whereby DBAs can leverage and expand their current database administration skills (see section 2.6, Core competencies of a DBA, pp. 21-23) in order to adequately acquire big data administration skills.

This study also provides a basis for further research in this area and for other emerging technologies that may become a future phenomenon in the data management Industry, as there are ever increasing volumes and variety of available data.

5.6 Limitations of the Research

This study has a number of limitations due to the limited time frame and exploratory nature of this research. The sample size was small and is therefore not representative of the target population. A larger sample size with an increased number of participants representing other big data stakeholders from the business aspect of big data may have been more beneficial and representative of the target population; this would have increased the validity of the findings and conclusions of this research.

It is worth noting that a key limitation in this study is the purely quantiative nature of the study. A mixed method approach of both quantitative and qualitative would have been advantageous as their combination would accommodate objective data mixed with the responses from the participants, which would reveal in depth human aspect of the research, thereby making it a more comprehensive study.

5.7 Recommendations for Future Work

This section proposes interesting directions for a number of related areas for future research in respect of big data technologies.

This study proposes further research on big data initiatives in organisations focusing on the specific activities performed by established big data teams, the different team structues and setup, as well as the skills these teams may posssess.

Further research in specific sectors could be beneficial as big data technologies become more widespread. Future research could identify adoption challenges in respect of databases as there seems to be a gap between the uses of different platforms. It would be interesting to see how newly implemented big data solutions can be integrated from a different perspective and what problems can arise.

Future research could focus on emerging database technologies and how their adoption can affect existing skills of DBAS; an interesting area of study. This should be continually looked into as one thing is certain: the volume, variety and velocity of data is unprecedented and continuously on the rise.

5.8 Summary

“Big data has had an evolutionary effect in the area of database administration, requiring in addition to traditional DBA expertise, a new range of skill sets” (Simmonds, 2013,p. 48)

DBAs have important data management skills that can be used in the new world of databases that have arisen with the advent of big data technologies, but current skills are nevertheless inadequate. One of the major findings of this research is that DBAs need to upskill to fit in the big data world and this was proven in this study.

The aim of this research is to determine the effect of big data technologies on the skills of DBAs. These skills, a majority of which DBAs possess, are magnified when seen through the eyes of big data. This research shows that most organisations today construct a team of different professionals, of which DBAs are a crucial part, to make up for the scarcity of the wealth of skills the data scientists role possesses, which is a combination of analytical, technical, business and communication skills. These roles or skills already exist in most organisations or are evolving internally, with organisations offering on the job big data training and implementing changes to enhanced data-driven practises.

This research provides a big data skills framework, which will aid organisations in selecting the right roles and skills when acquiring big data talent. It will also aid DBAs in developing the skills they need to fit into a big data team and play a crucial role in implementing big data intiatives. Successful DBAs will be those who are ready to see the oppourtunities available with the advent of big data complexity and also embrace the inevitable changes that come with big data.

References

ABI Research. 2013. *More Than 30 Billion Devices Will Wirelessly Connect to the Internet of Everything in 2020* [Online]. Available: https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne [Accessed 11 February 2014].

ACCA. 2013. *Big data: its power and perils* [Online]. Available: http://www.accaglobal.com/bigdata [Accessed 11 February 2014].

Accenture. 2013. *The Team Solution to the Data Scientist Shortage* [Online]. Available: http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Team-Solution-Data-Scientist-Shortage.pdf [Accessed 6 March 2014].

Agrawal, D., Das, S. & Abbadi, A. E. Big Data and Cloud Computing: Current State and Future Opportunities. Proceedings of the 14th International Conference on Extending Database Technology, March 22–24 2011. ACM, pp. 530-533.

Ahuja, S. P. & Moore, B. 2013. State of Big Data Analysis in the Cloud. *Network and Communication Technologies,* Vol. 2.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. & Zaharia, M. 2010. A View of Cloud Computing. *Communications of the ACM,* vol. 53.

Bean, R., Davenport, T. & Barth, P. 2012. How Big Data is Different. *MIT Sloan Mangement Review,* VOL.54.

Bell, J. 2005. *Doing your Research Project* [Online]. Open University Press. Available: http://moodle.up.pt/pluginfile.php/60167/mod_forum/attachment/9449/LIVRO_Judith%20Bell%20-%20Doing_Your_Research_Project.pdf [Accessed 06 April 2014].

Bertino, E. 2013. Big Data - Opportunities and Challenges. *IEEE 37th Annual Computer Software and Applications Conference.*

Blumberg B, Cooper D.R & Schindler P.S 2005. *Business Research Methods* McGraw-Hill Higher Education

Borthakur, D. 2008. *HDFS Architecture Guide* [Online]. Hadoop Apache Project. Available: http://archive.cloudera.com/cdh/3/hadoop-0.20.2-cdh3u6/hdfs_design.pdf.

Bryman & Bell 2007. *Business Research Methods*, Oxford University Press.

Cattell, R. 2011. Scalable SQL and NoSQL Data Stores. *ACM SIGMOD Record,* Vol 39**,** pp 12-27.

Centre for Economics and Business Research. 2013. *Data equity – Ireland Unlocking the value of big data* [Online]. Available: http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf [Accessed 24 January 2014].

Chandran, R. 2013. Big Data Management Platforms: Architecting Heterogeneous Solutions. *Business Intelligence Journal,* vol.18**,** pp 32-38.

Codd, E. F. 1970. A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM,* Volume 13**,** pp 377 -387.

Cohen, L., Manion, L. & Morrison, K. 2000. *Research Methods in Education*, RoutledgeFalmer.

Cook, C., HEATH, F. & THOMPSON, R. L. 2000. A META-ANALYSIS OF RESPONSE RATES IN WEB- OR INTERNET-BASED SURVEYS. *Educational and Psychological Measurement,* Vol 60**,** pp 821-836.

Daas, P. J. H., Puts, M. J., Buelens, B. & van den Hurk, P. A. M. 2012. Big Data and Official Statistics. *Sharing Advisory Board, Software Sharing Newsletter 7 pg 2-3*.

e-skills UK. 2013. *Big Data Analytics Adoption and Employment Trends, 2012–2017* [Online]. Available: http://www.sas.com/offices/europe/uk/downloads/bigdata/eskills/eskills.pdf [Accessed 8 February 2014].

Economist Intelligence Unit. 2011. *Big data Harnessing a game-changing asset* [Online]. Available: http://www.sas.com/resources/asset/SAS_BigData_final.pdf [Accessed 9 February 2014].

Edirisingha, P. 2012. What is Ontology and What is Epistemology? . *prabash78* [Online]. Available from: http://prabash78.wordpress.com/2012/03/14/interpretivism-and-postivism-ontological-and-epistemological-perspectives/ 17 March 2014].

EMC Corporation 2011. Big Data: Big Opportunities to Create Business Value. Report and recommendations based on discussions with the Leadership Council for Information Advantage [Online]. Available: http://www.emc.com/microsites/cio/articles/big-data-big-opportunities/LCIA-BigData-Opportunities-Value.pdf [Accessed 30th April 2014].

Enterprise Strategy Group 2011. ESG white paper: Big Data Advisory Service. [Online]. Available: [Accessed January 2014].

Europe, A. 2009. *The four main approaches* [Online]. Available: http://www.alzheimer-europe.org/Research/Understanding-dementia-research/Types-of-research/The-four-main-approaches [Accessed 18 March 2014].

European e-Competence Framework 2.0. 2010. *User Guidelines 2.0 for the application of the European e-competence framework* [Online]. Available: http://www.ecompetences.eu/site/objects/download/5999_EUeCF2.0userguide.pdf [Accessed 02 March 2014].

Fingar, P. 2011. *Consider Big Data as the Most Important Thing for Business since the Internet* [Online]. www.bptrends.com. Available: http://www.bptrends.com/publicationfiles/12-06-2011-COL-Extreme%20Competition-Big%20Data-Fingar.pdf [Accessed 12 February 2014].

Flowers, P. 2009. *Research Philosophies – Importance and Relevance* [Online]. Available: https://www.networkedcranfield.com/cell/Assigment%20Submissions/research%20philosophy%20-%20issue%201%20-%20final.pdf [Accessed 15 March 2014].

Fung, H. P. 2013. Using Big Data Analytics in Information Technology (IT) Service Delivery. *Internet Technologies and Applications Research,* 1**,** 6-10.

Gable, G. 1994. Integrating Case Study and Survey Research Methods: An Example in Information Systems. *European Journal of Information Systems,* Vol 3**,** pp.112-126.

Gardner, S. D., Lepak, D. P. & Bartolc, K. M. 2003. Virtual HR: The impact of information technology on the human resource professional. *Journal of Vocational Behavior,* 63**,** pp 159-179.

Gartner. 2011. *Extreme Data: Challenges and Opportunities for Large-Scale Data Warehousing, BI and Analytics* [Online]. Available:

http://gartnerinfo.com/futureofit2011/MEX38L_C3%20mex38l_c3.pdf [Accessed 8 May 2014].

Gartner. 2012a. *Big Data* [Online]. Gartner IT Glossary. Available: http://www.gartner.com/it-glossary/big-data/ [Accessed 20 January 2014].

Gartner. 2013. *Driving Value from Big Data* [Online]. Available: http://www.gartner.com/technology/research/big-data/ [Accessed 02 May 2014].

Gartner, T. P. 2012b. Gartner Upbeat on Big Data Jobs. *Computerworld* 46**,** pp 6-6.

George, M. & Glady, N. 2013. *4 by 4 analytics framework: developing organizations and their data scientists in business analytics* [Online]. Available: http://www.baqmar.eu/2013/4-by-4-analytics-framework-developing-organizations-and-their-data-scientists-in-business-analytics/ [Accessed 06 March 2014].

Gillman B 2000. *Developing a questionnaire*, Continuum.

Gray 2009. *Doing Research in the Real World*, SAGE.

Greener, S. 2008. *Business Research Methods* [Online]. Available: http://dl.is.vnu.edu.vn/bitstream/123456789/269/1/introduction-to-research-methods.pdf [Accessed 14 March 2014].

Gualtieri, M. 2012. *The Pragmatic Definition Of Big Data* [Online]. Forrester. Available: http://blogs.forrester.com/mike_gualtieri/12-12-05-the_pragmatic_definition_of_big_data [Accessed 9 February 2014].

Hagström, M. & Manocha, I. 2013. The Big Opportunity for Inclusive Growth. The Global Information Technology Report [Online]. Available: http://www3.weforum.org/docs/GITR/2013/GITR_Chapter1.9_2013.pdf [Accessed.

Harvard Business Review. 2012. *Big Data: The Management Revolution* [Online]. Harvard Business School Publishing Corporation. Available: http://automotivedigest.com/wp-content/uploads/2013/01/BigDataR1210Cf2.pdf [Accessed 05 February 2014].

Hoboken, N. J. 2007. SIM & Professor Luftman Release 2007 CIO Survey Data. *Developer's Power Builder Journal*.

Hussey J & Hussey R 1997. *Business Research: A Practical Guide for Undergraduate and Postgraduate Students,* Basingstoke, Macmillan Press Limited.

IBM. 2013a. *What is big data?* [Online]. Available: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html [Accessed 25 January 2014].

IBM. 2013b. *What is Hadoop?* [Online]. Available: http://www-01.ibm.com/software/data/infosphere/hadoop/ [Accessed 12 February 2014].

IBM. 2014. *Manager's Guide: Database Administrator* [Online]. Available: http://skills.salary.com/Job/Database-Administrator [Accessed 27 February 2014].

International Data Corporation. 2011. *Extracting Value from Chaos* [Online]. IDC iview: IDC. Available: http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf [Accessed 12 February 2014].

International Data Corporation. 2012. *THE DIGITAL UNIVERSE IN 2020: Big Data,Bigger Digi tal Shadows, and Biggest Growth in the Far East* [Online]. IDC iview: IDC. Available: http://idcdocserv.com/1414 [Accessed 26 January 2014].

Investopedia. 2014. *Technical Skills* [Online]. Available: http://www.investopedia.com/terms/t/technical-skills.asp [Accessed 10 March 2014].

ISACA. 2013. *Big Data Impacts and Benefits An ISACA White Paper* [Online]. Available: http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Big-Data-Impacts-and-Benefits.aspx.

James, J. 2012. *How Much Data is Created Every Minute* [Online]. Available: http://www.domo.com/blog/2012/06/how-much-data-is-created-every-minute/?dkw=socf3 [Accessed 25 January 2014].

Joseph, Ang, S., Chang, R. H. L. & Slaughter, S. A. 2010. Practical Intelligence in IT : Assessing Soft Skills of IT Professionals. *Communications of the ACM,* Vol 53**,** pp 149-154.

Kaplan & Duchon. 1988. *Combining Qualitative and Quantitative Methods In Information Systems Research: A Case Study* [Online]. MIS Quarterly. Available: http://staff.blog.ui.ac.id/r-suti/files/2010/03/qualquant_in_is.pdf [Accessed 05 April 2014].

Kim, H. J., Pelaez, A. & Winston, E. R. 2013. Experiencing Big Data Analytics: Analyzing Social Media Data In Financial Sector as a Case Study. *Northeast Decision Sciences Institute Annual Meeting Proceedings*, pp.62-69.

Kitchenham, B. & Pfleeger, S. L. 2002. Principles of Survey Research Part 5: Populations and Samples. *ACM SIGSOFT,* vol 27 pp 17-20.

Laney, D. 2001. 3D data management: Controlling data volume, velocity. and variety. [Online]. Available: [Accessed.

Leary 1995. Introduction To Behavioural Research Methods. 2nd ed.

Letouzé, E. 2012. Big Data for Development: Challenges & Opportunities. [Online]. Available: [Accessed.

Lo, F. 2013. *Big Data Technology* [Online]. DataJobs.com. Available: https://datajobs.com/what-is-hadoop-and-nosql [Accessed 17 February 2014].

McBride, N. 2005. Chaos theory as a model for interpreting information systems in organizations. *Information Systems Journal* 15**,** pp 233-254.

McKendrick, J. 2011. Managing The Rapid Rise In Database Growth: 2011 IOUG Survey On Database Manageability. [Online]. Available: [Accessed.

McKinsey Global Institute 2011. Big Data: The Next Frontier For Innovation, Competition & Productivity,. *Business Source Complete*, pp.1-143.

Mithas, S., Lee, M. R., Earley, S., Murugesan, S. & Djavanshir, R. 2013. Leveraging Big Data and Business Analytics. *IT Professional,* 15**,** pp 18-20.

Moniruzzaman, A. B. M. & Hossain, S. A. 2013. NoSQL Database: New Era of Databases for Big data Analytics-Classification, Characteristics and Comparison. *International Journal of Database Theory & Application,* 6.

Mullins, C. 2013. The Ever-Changing Role of the DBA. Available: http://www.b-eye-network.com/view/17002.

Mullins, C. S. 2002. *Database Administration: The Complete Guide to Practices and Procedures* [Online]. Addison Wesley. Available:

https://jogins.com/media/Science%20&%20Engineering/Database_ebook_collection/Addison%20Wesley%20-%20Database%20Administration,%20The%20Complete%20Guide%20to%20Practices%20and%20Procedures%20(2002).pdf [Accessed 01 March 2014].

National Careers Service. 2012. *Database Administrator* [Online]. Available: https://nationalcareersservice.direct.gov.uk/advice/planning/jobprofiles/Pages/databaseadministrator.aspx [Accessed 01 March 2014].

Navint. 2012. *Why is BIG Data Important?* [Online]. Available: http://www.navint.com/images/Big.Data.pdf [Accessed 07 February 2014].

O'Leary. 2004. *The essential guide for doing research* [Online]. Sage. Available: http://eunacal.org/metodakerkimi/wp-content/uploads/spss/The_essential_guide_to_doing_research.pdf [Accessed 06 April 2014].

O'leary, D. E. 2013. 'Big Data', The 'Internet of Things' and the 'Internet of Signs' *Intelligent Systems In Accounting, Finance and Management,* 20**,** pp.53-65.

Peppard, J. 2013. The big data oppourtunity. *Management Focus***,** pg 20.

Polakowski, M. 2009. Database Administrators: It’s All in a Day’s Work. *Certification Magazine.*

Pratt, M. K. 2012. 10 hot IT skills for 2013. *Computerworld*.

Pugh, M. 2013. *Big Data 101 – NoSql, Hadoop, Analytic DS, RDBMS Differences for Business People* [Online]. Available: http://www.logianalytics.com/blog/logixml-big-data-101-hadoop-Nosql-AnalyticDS-RDBMS-differences-for-business-people [Accessed 17 February 2014].

Saunders, Lewis & Thornhill. 2009. *Research methods for business students* [Online]. Available: http://doha.ac.mu/ebooks/Research%20Methods/ResearchMethodsForBusinessStudents_Saunders.pdf [Accessed 01 March 2014].

Sheppard, B. 2011. *Putting Big Data to Work: Opportunities for Enterprises* [Online]. GigaOM Pro. Available:

http://gigaomevents.files.wordpress.com/2011/03/puttingbigdatatowork_032011.pdf [Accessed 28 January 2014].

Simmonds, T. 2013. Teaching Database Administration in the World of Big Data and Small Budgets. University of Sunderland.

Skills Framework for the Information Age. 2011. SFIA 5 framework reference.

Snijders, C., Matzat, U. & Reips, U.-D. 2012. “Big Data”: Big Gaps of Knowledge in the Field of Internet Science *International Journal of Internet Science,* 7**,** 1-5.

TechAmerica Foundation’s Federal Big Data Commission 2004. Demystifying Big Data: A Practical Guide To Transforming The Business of Government. [Online]. Available: http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf [Accessed.

techdirections. 2010. *Database Administrator* [Online]. www.techdirections.com: Prakken Publications. [Accessed 01 March 2014].

The Data Warehousing Institute 2011. *Big Data Analytics* [Online]. TDWI Best Practices Report. Available: ftp://129.35.224.12/software/tw/Defining_Big_Data_through_3V_v.pdf [Accessed 26 January 2014].

The Data Warehousing Institute 2013a. *ENHANCING YOUR DATA WAREHOUSE* [Online]. TDWI. Available: http://tdwi.org/~/media/TDWI/TDWI/EBooks/2013/TDWI_E-Book_Enhancing-Data-Warehouse_IBM_web [Accessed 12 February 2014].

The Data Warehousing Institute 2013b. Where Hadoop Fits in your Data Warehouse Architecture. TDWI Checklist Report [Online]. Available: [Accessed.

TMC News. 2012. *Top Ten IT Trends for 2012 (Malaysian Business (Malaysia)* [Online]. Available: http://www.tmcnet.com/usubmit/2012/01/31/6086248.htm [Accessed 21st May 2014].

Tole, A. A. 2013. Big Data Challenges. *Database Systems Journal,* vol. IV.

Trochim, W. M. K. 2006. *Deduction & Induction* [Online]. Available: http://www.socialresearchmethods.net/kb/dedind.php [Accessed 19 March 2014].

Ularu, E. G., Puican, F. C., Apostu, A. & Velicanu, M. 2012. Perspectives on Big Data and Big Data Analytics. *Database Systems Journal,* vol.III.

Villars, R. L., Olofson, C. W. & Eastwood, M. 2011. *Big Data: What It Is and Why You Should Care* [Online]. International Data Corporation (IDC). Available: http://sites.amd.com/us/Documents/IDC_AMD_Big_Data_Whitepaper.pdf [Accessed 7 February 2014].

Walsham, G. 2009. *Interpreting Information Systems in Organizations* [Online]. Available: http://pustakalaya.org/eserv.php?pid=Pustakalaya:1663&dsID=GeoffWalsham2009_InterpretingInformationSystems.pdf [Accessed 17 March 2014].

Welman, C., Kruger, F. & Mitchell, B. 2005. *Research Methodology*, Oxford University Press.

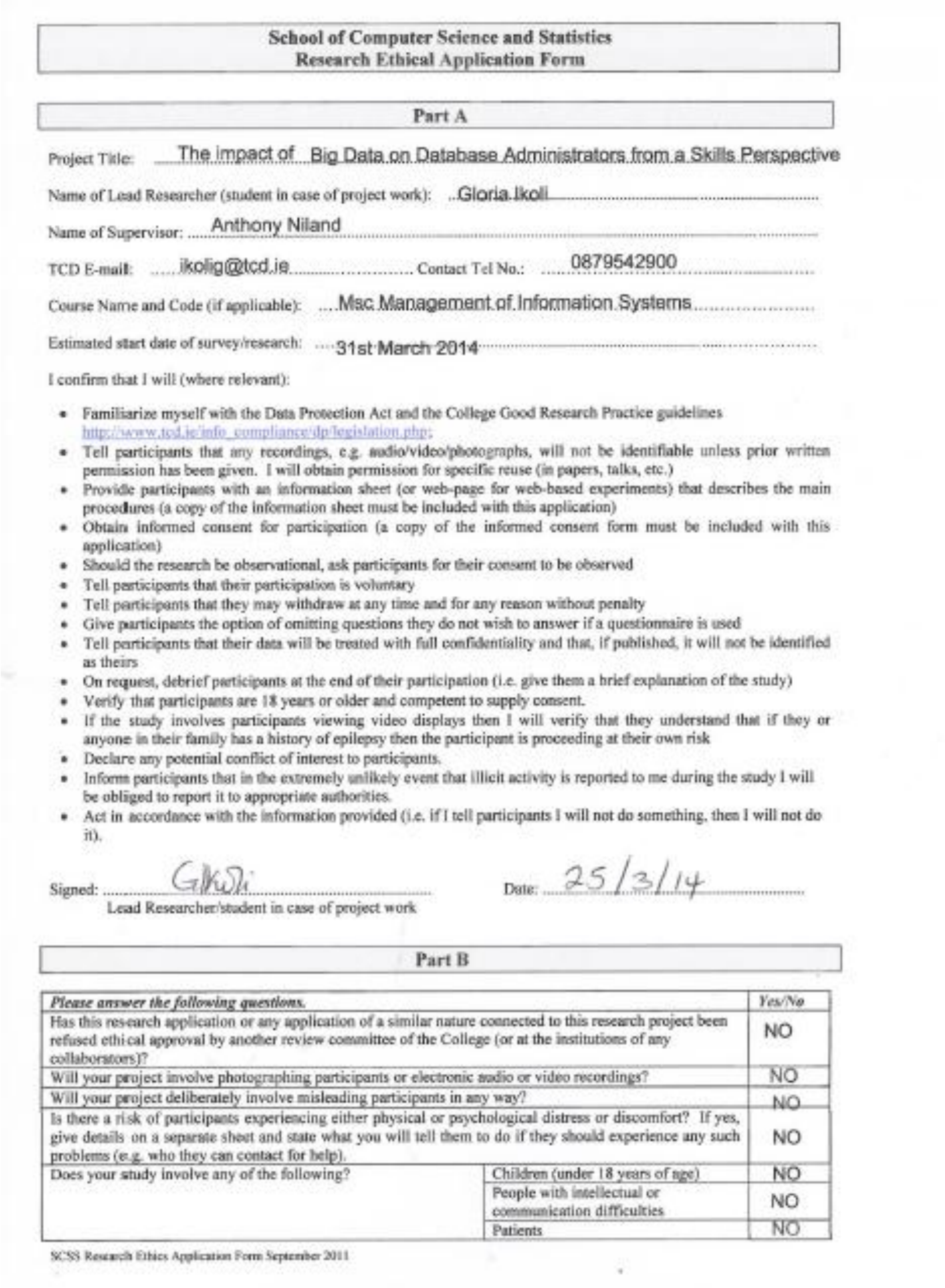
Wu, X., Zhu, X., Wu, G.-Q. & Ding, W. 2014. Data Mining with Big Data. *Knowledge and Data Engineering, IEEE Transactions,* Volume 26**,** pp 97-107.

Zikopoulos, deRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D. & Giles, J. 2013. Harness the Power of Big Data. [Online]. Available: [Accessed.

Zikopoulos, Eaton, C., DeRoos, D., Deutsch, T. & Lapis, G. 2012. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media.

Appendices

Appendix 1 – Ethics Application



School of Computer Science and Statistics
Research Ethical Application Form

Part A

Project Title: The impact of Big Data on Database Administrators from a Skills Perspective

Name of Lead Researcher (student in case of project work): Gloria Ikoli

Name of Supervisor: Anthony Niland

TCD E-mail: ikolig@tcd.ie Contact Tel No.: 0879542900

Course Name and Code (if applicable): Msc Management of Information Systems

Estimated start date of survey/research: 31st March 2014

I confirm that I will (where relevant):

- Familiarize myself with the Data Protection Act and the College Good Research Practice guidelines http://www.tcd.ie/info_compliance/dp/legislation.php;
- Tell participants that any recordings, e.g. audio/video/photographs, will not be identifiable unless prior written permission has been given. I will obtain permission for specific reuse (in papers, talks, etc.)
- Provide participants with an information sheet (or web-page for web-based experiments) that describes the main procedures (a copy of the information sheet must be included with this application)
- Obtain informed consent for participation (a copy of the informed consent form must be included with this application)
- Should the research be observational, ask participants for their consent to be observed
- Tell participants that their participation is voluntary
- Tell participants that they may withdraw at any time and for any reason without penalty
- Give participants the option of omitting questions they do not wish to answer if a questionnaire is used
- Tell participants that their data will be treated with full confidentiality and that, if published, it will not be identified as theirs
- On request, debrief participants at the end of their participation (i.e. give them a brief explanation of the study)
- Verify that participants are 18 years or older and competent to supply consent.
- If the study involves participants viewing video displays then I will verify that they understand that if they or anyone in their family has a history of epilepsy then the participant is proceeding at their own risk
- Declare any potential conflict of interest to participants.
- Inform participants that in the extremely unlikely event that illicit activity is reported to me during the study I will be obliged to report it to appropriate authorities.
- Act in accordance with the information provided (i.e. if I tell participants I will not do something, then I will not do it).

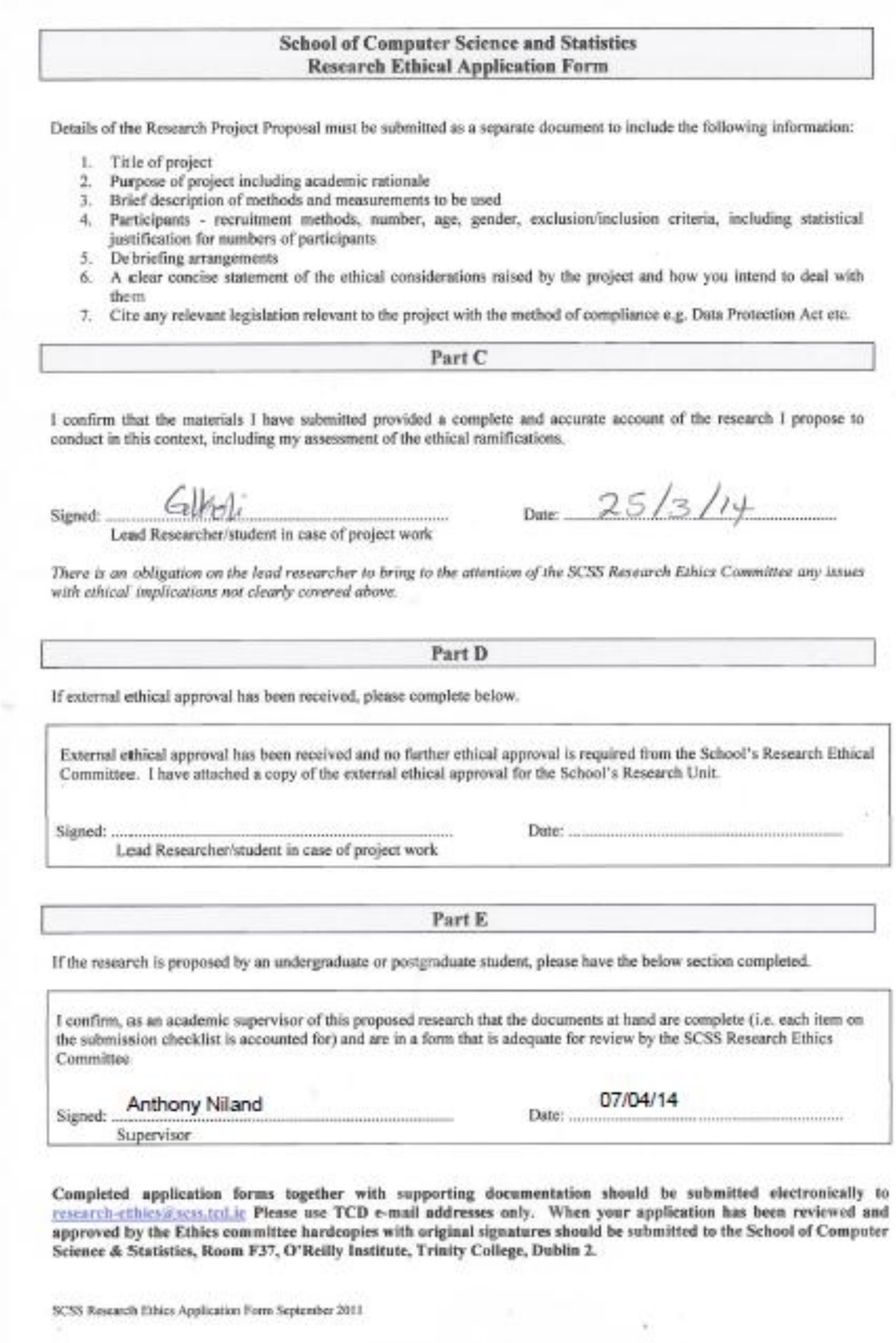
Signed: GIkoli
Lead Researcher/student in case of project work

Date: 25/3/14

Part B

Please answer the following questions.		*Yes/No*
Has this research application or any application of a similar nature connected to this research project been refused ethical approval by another review committee of the College (or at the institutions of any collaborators)?		NO
Will your project involve photographing participants or electronic audio or video recordings?		NO
Will your project deliberately involve misleading participants in any way?		NO
Is there a risk of participants experiencing either physical or psychological distress or discomfort? If yes, give details on a separate sheet and state what you will tell them to do if they should experience any such problems (e.g. who they can contact for help).		NO
Does your study involve any of the following?	Children (under 18 years of age)	NO
	People with intellectual or communication difficulties	NO
	Patients	NO

SCSS Research Ethics Application Form September 2011



School of Computer Science and Statistics
Research Ethical Application Form

Details of the Research Project Proposal must be submitted as a separate document to include the following information:

1. Title of project
2. Purpose of project including academic rationale
3. Brief description of methods and measurements to be used
4. Participants - recruitment methods, number, age, gender, exclusion/inclusion criteria, including statistical justification for numbers of participants
5. Debriefing arrangements
6. A clear concise statement of the ethical considerations raised by the project and how you intend to deal with them
7. Cite any relevant legislation relevant to the project with the method of compliance e.g. Data Protection Act etc.

Part C

I confirm that the materials I have submitted provided a complete and accurate account of the research I propose to conduct in this context, including my assessment of the ethical ramifications.

Signed: .... Date: 25/3/14
Lead Researcher/student in case of project work

There is an obligation on the lead researcher to bring to the attention of the SCSS Research Ethics Committee any issues with ethical implications not clearly covered above.

Part D

If external ethical approval has been received, please complete below.

External ethical approval has been received and no further ethical approval is required from the School's Research Ethical Committee. I have attached a copy of the external ethical approval for the School's Research Unit.

Signed: .... Date: ....
Lead Researcher/student in case of project work

Part E

If the research is proposed by an undergraduate or postgraduate student, please have the below section completed.

I confirm, as an academic supervisor of this proposed research that the documents at hand are complete (i.e. each item on the submission checklist is accounted for) and are in a form that is adequate for review by the SCSS Research Ethics Committee

Signed: Anthony Niland Date: 07/04/14
Supervisor

Completed application forms together with supporting documentation should be submitted electronically to research-ethics@scss.tcd.ie Please use TCD e-mail addresses only. When your application has been reviewed and approved by the Ethics committee hardcopies with original signatures should be submitted to the School of Computer Science & Statistics, Room F37, O'Reilly Institute, Trinity College, Dublin 2.

SCSS Research Ethics Application Form September 2011

Appendix 2 – Information Page for Participants

BACKGROUND OF RESEARCH:

This research attempts to investigate the impact of the emergence of Big Data on the skills of Database Administrators. In order to carry out this research a web based survey will be conducted on all stakeholders to determine the impact of this technology from a skills perspective.

THE SURVEY PROCESS:

The following points should be noted about the survey:

- Your participation is voluntary and anonymous;
- The survey will take you approximately 10 minutes to complete and each question is optional;
- You have the right to withdraw from the survey at any time during the process without penalty;
- You may refuse to answer a question without penalty.

OTHER INFORMATION:

- This information is being gathered for the completion of a dissertation as part of the M.Sc.in Management of Information Systems.
- This dissertation along with the gathered anonymous data may be published in Trinity College Dublin Library along with all other theses and dissertations.
- I have no conflict of interest with regard to the research topic and with any of the participants either individually or at an organisational level.
- I am required by TCD to inform you that, in the extremely unlikely event that illicit activity is reported I will be obliged to report it to appropriate authorities.
- Please do not name third parties in any open text field of the questionnaire. Any such replies will be anonymised.

Appendix 3 – Informed Consent Form

RESEARCHER: Gloria Ikoli

CONTACT DETAILS: ikolig@tcd.ie

BACKGROUND OF RESEARCH

Big data is comprised of datasets too large to be handled by traditional database systems. The advent of big data has brought with it new ways to save, retrieve and process data, which has given rise to technical challenges. This new technology poses opportunities and challenges for businesses as changes to hardware, software, and data processing techniques are necessary.

This research attempts to investigate the impact of the emergence of Big Data on the skills of Database Administrators. In order to carry out this research a web based survey will be conducted on all stakeholders to determine the impact of this technology from a skills perspective.

PROCEDURES OF THIS STUDY

Participants must be 18 years of age or older. Participants in this survey will remain anonymous. Participation is voluntary and should take no longer than 10 minutes. All survey questions are optional and can be skipped nevertheless the researcher would appreciate it if all questions are answered. Participants have the option to “Exit” any time before completing the questionnaire by clicking the “Exit This Survey” button, in so doing; answers will NOT be recorded.

PUBLICATION
At the end of the survey, individual results will be aggregated anonymously and research reported on aggregate results. The results will be used solely for a dissertation as part of the completion of a M.Sc. in Management of Information Systems course at Trinity College Dublin (TCD).

CONFIDENTIALITY
Your responses will be kept completely confidential.

If you have any questions or concerns or if you have any difficulties accessing this survey,

you may contact me at: ikolig@tcd.ie.
DECLARATION:
- I am 18 years or older and am competent to provide consent.

- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.

- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.

- I understand that if I make illicit activities known, these will be reported to appropriate authorities.

- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.

- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.

- I understand that my participation is fully anonymous and that no personal details about me will be recorded.

- Since this research involves viewing materials via a computer monitor I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.

By submitting this form you are indicating that you have read the description of the study, are over the age of 18, and that you agree to the terms as described.

Thank you in advance for your participation!

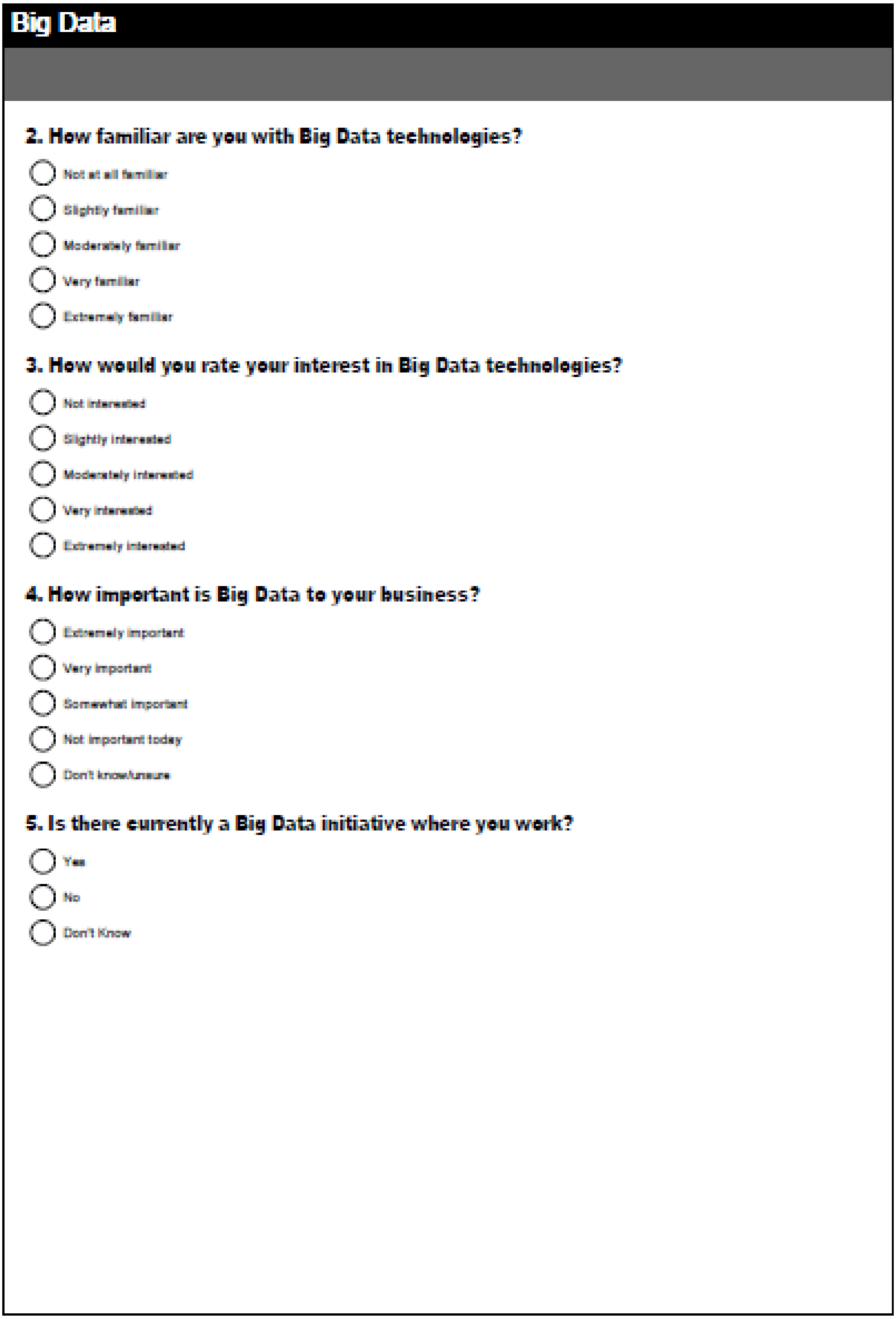
Gloria Ikoli

1. Do you agree to the consent information listed on this form?

○ Yes

○ No

Appendix 4 – Questionnaire



Big Data

2. How familiar are you with Big Data technologies?

○ Not at all familiar
○ Slightly familiar
○ Moderately familiar
○ Very familiar
○ Extremely familiar

3. How would you rate your interest in Big Data technologies?

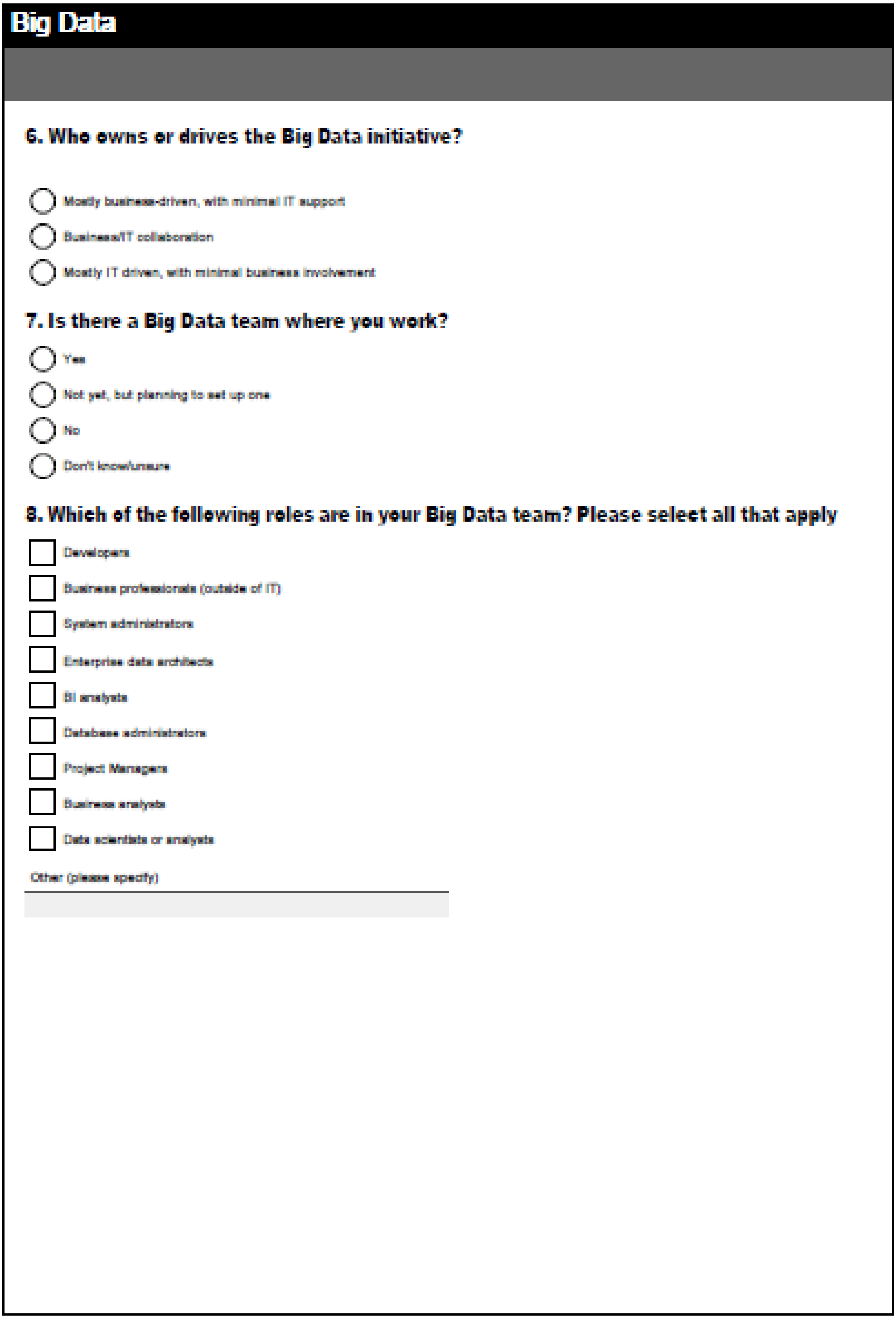
○ Not interested
○ Slightly interested
○ Moderately interested
○ Very interested
○ Extremely interested

4. How important is Big Data to your business?

○ Extremely important
○ Very important
○ Somewhat important
○ Not important today
○ Don't know/unsure

5. Is there currently a Big Data initiative where you work?

○ Yes
○ No
○ Don't Know



Big Data

6. Who owns or drives the Big Data initiative?

○ Mostly business-driven, with minimal IT support
○ Business/IT collaboration
○ Mostly IT driven, with minimal business involvement

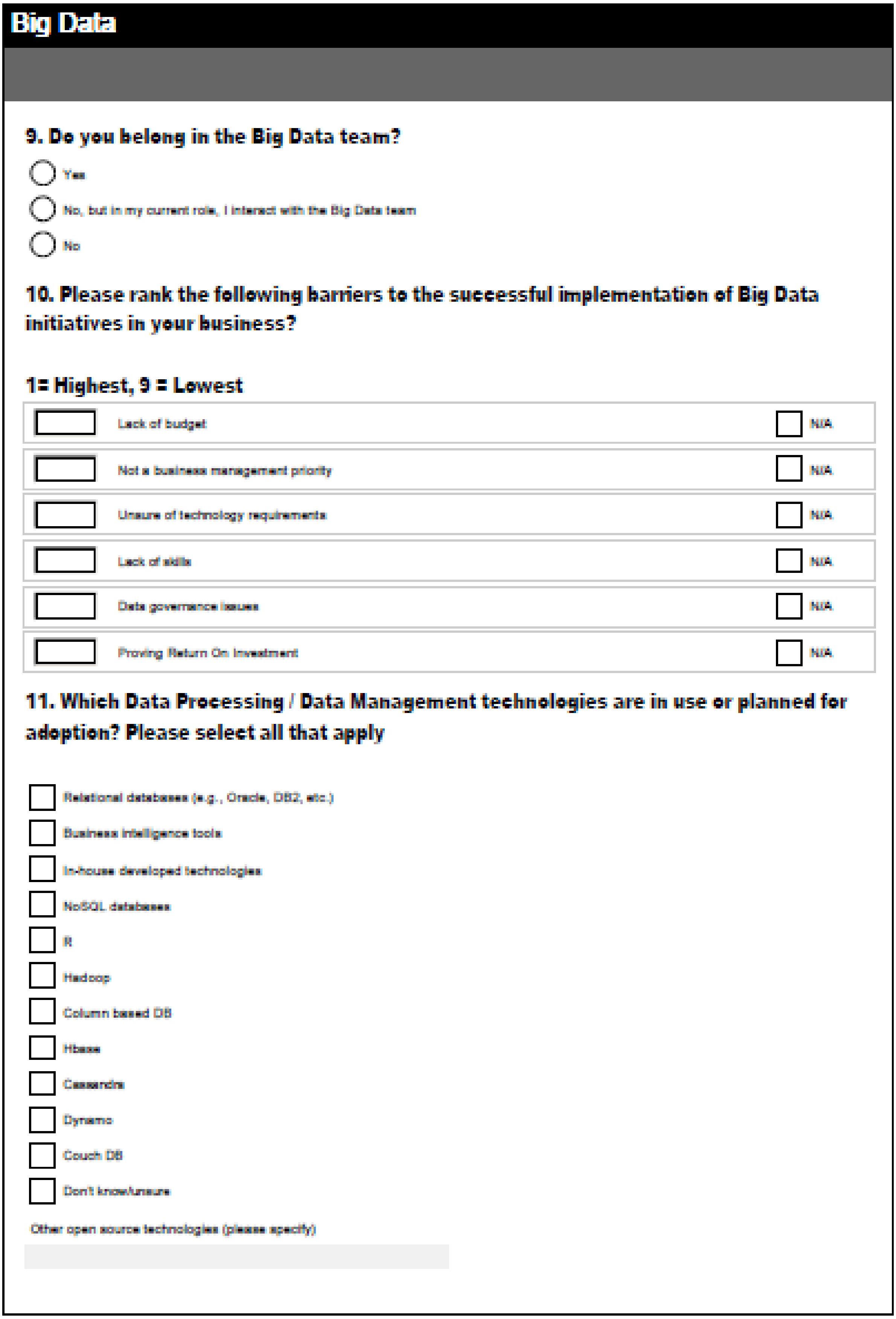
7. Is there a Big Data team where you work?

○ Yes
○ Not yet, but planning to set up one
○ No
○ Don't know/unsure

8. Which of the following roles are in your Big Data team? Please select all that apply

☐ Developers
☐ Business professionals (outside of IT)
☐ System administrators
☐ Enterprise data architects
☐ BI analysts
☐ Database administrators
☐ Project Managers
☐ Business analysts
☐ Data scientists or analysts

Other (please specify)



Big Data

9. Do you belong in the Big Data team?

○ Yes
○ No, but in my current role, I interact with the Big Data team
○ No

10. Please rank the following barriers to the successful implementation of Big Data initiatives in your business?

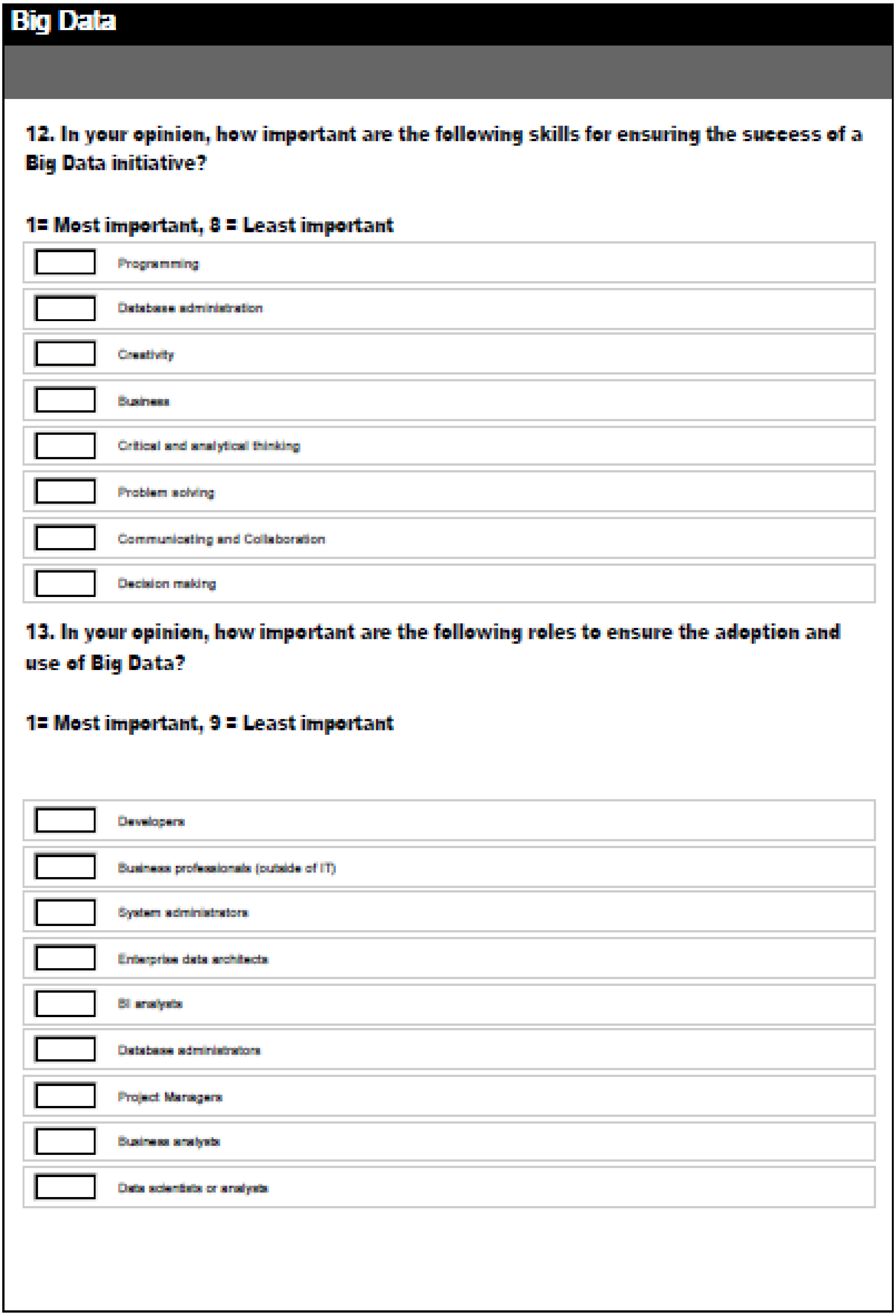
1= Highest, 9 = Lowest

☐	Lack of budget	☐ N/A
☐	Not a business management priority	☐ N/A
☐	Unsure of technology requirements	☐ N/A
☐	Lack of skills	☐ N/A
☐	Data governance issues	☐ N/A
☐	Proving Return On Investment	☐ N/A

11. Which Data Processing / Data Management technologies are in use or planned for adoption? Please select all that apply

☐ Relational databases (e.g., Oracle, DB2, etc.)
☐ Business intelligence tools
☐ In-house developed technologies
☐ NoSQL databases
☐ R
☐ Hadoop
☐ Column based DB
☐ Hbase
☐ Cassandra
☐ Dynamo
☐ Couch DB
☐ Don't know/unsure

Other open source technologies (please specify)



Big Data

12. In your opinion, how important are the following skills for ensuring the success of a Big Data initiative?

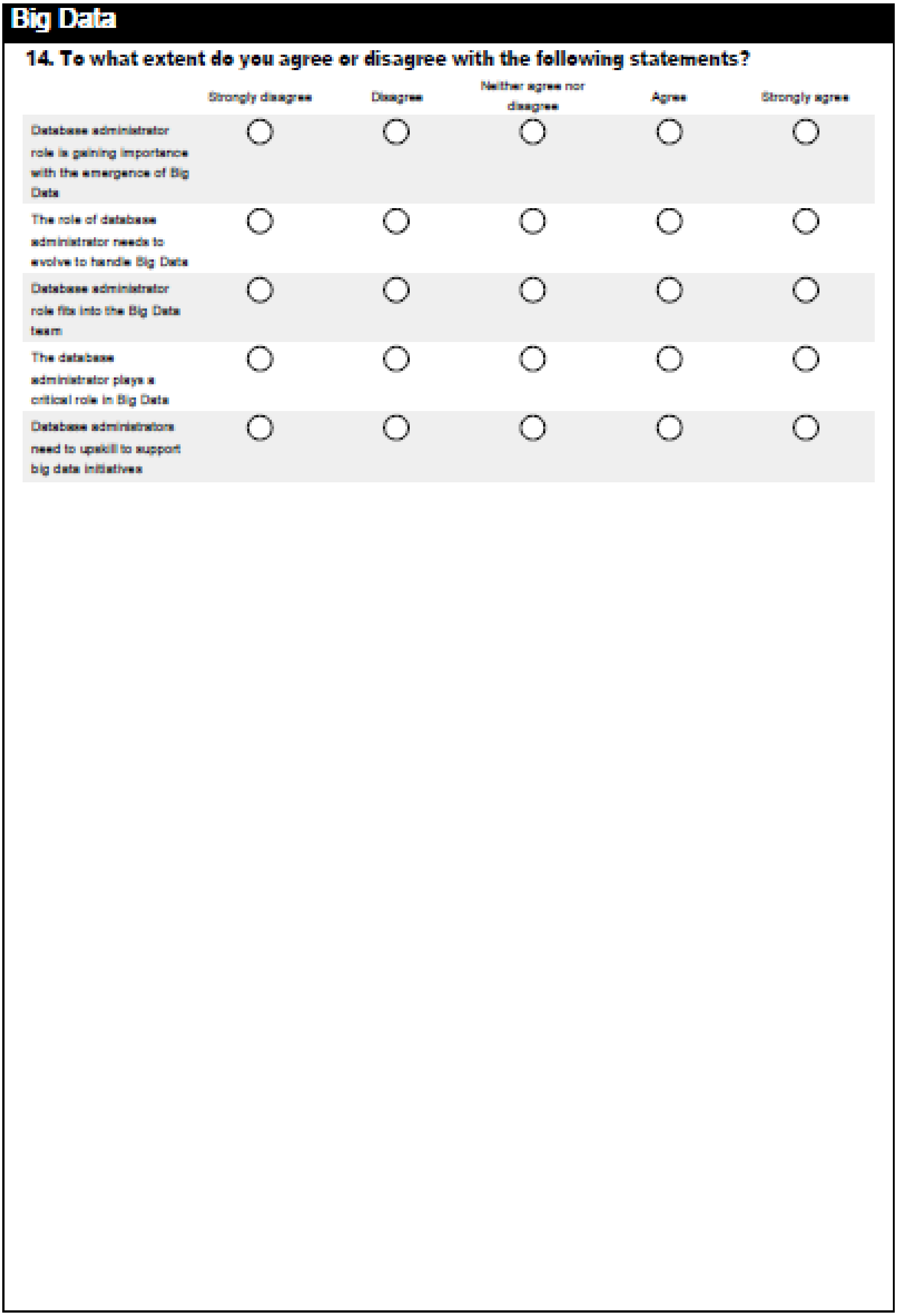
1= Most important, 8 = Least important

☐ Programming
☐ Database administration
☐ Creativity
☐ Business
☐ Critical and analytical thinking
☐ Problem solving
☐ Communicating and Collaboration
☐ Decision making

13. In your opinion, how important are the following roles to ensure the adoption and use of Big Data?

1= Most important, 9 = Least important

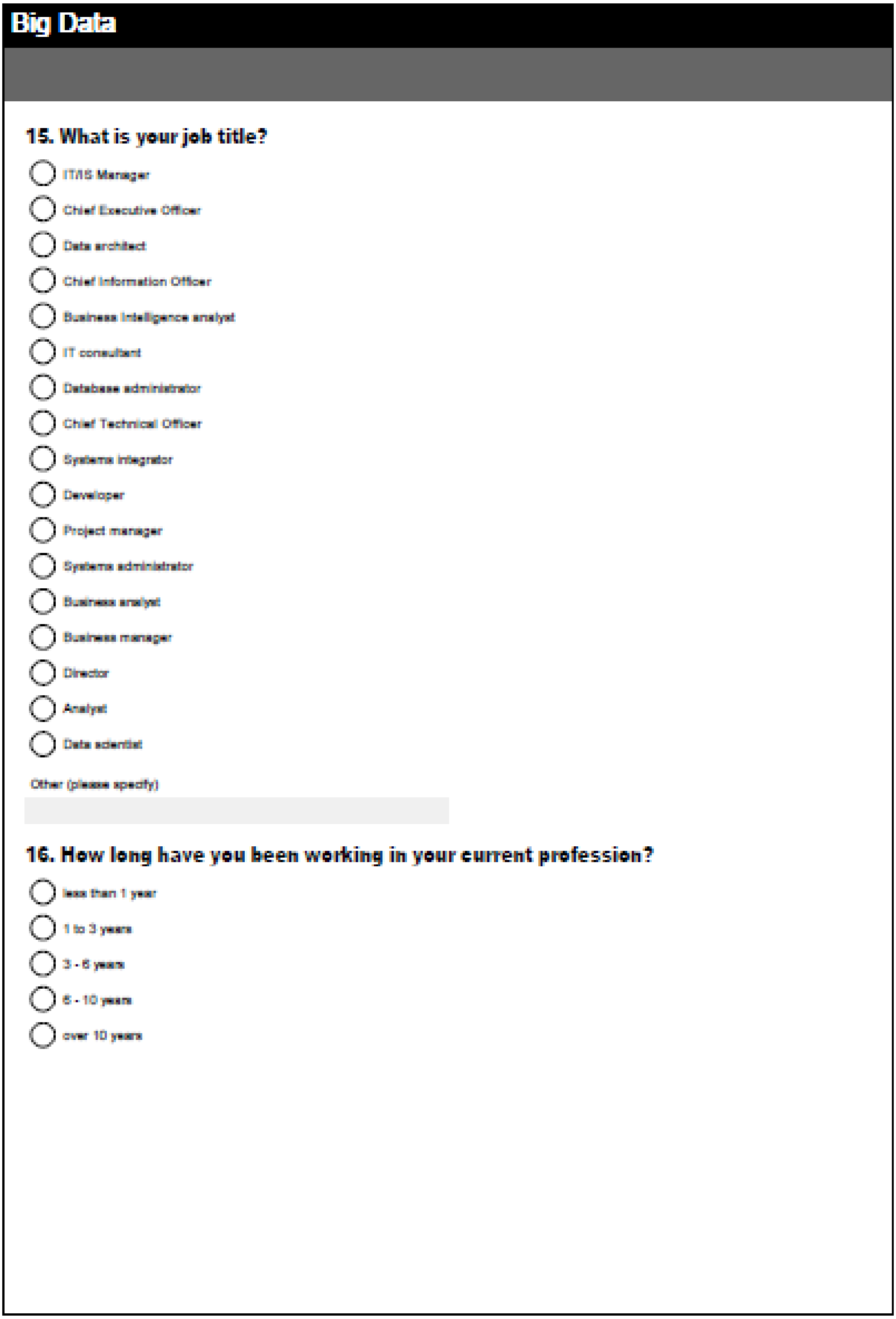
☐ Developers
☐ Business professionals (outside of IT)
☐ System administrators
☐ Enterprise data architects
☐ BI analysts
☐ Database administrators
☐ Project Managers
☐ Business analysts
☐ Data scientists or analysts



Big Data

14. To what extent do you agree or disagree with the following statements?

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Database administrator role is gaining importance with the emergence of Big Data	○	○	○	○	○
The role of database administrator needs to evolve to handle Big Data	○	○	○	○	○
Database administrator role fits into the Big Data team	○	○	○	○	○
The database administrator plays a critical role in Big Data	○	○	○	○	○
Database administrators need to upskill to support big data initiatives	○	○	○	○	○



Big Data

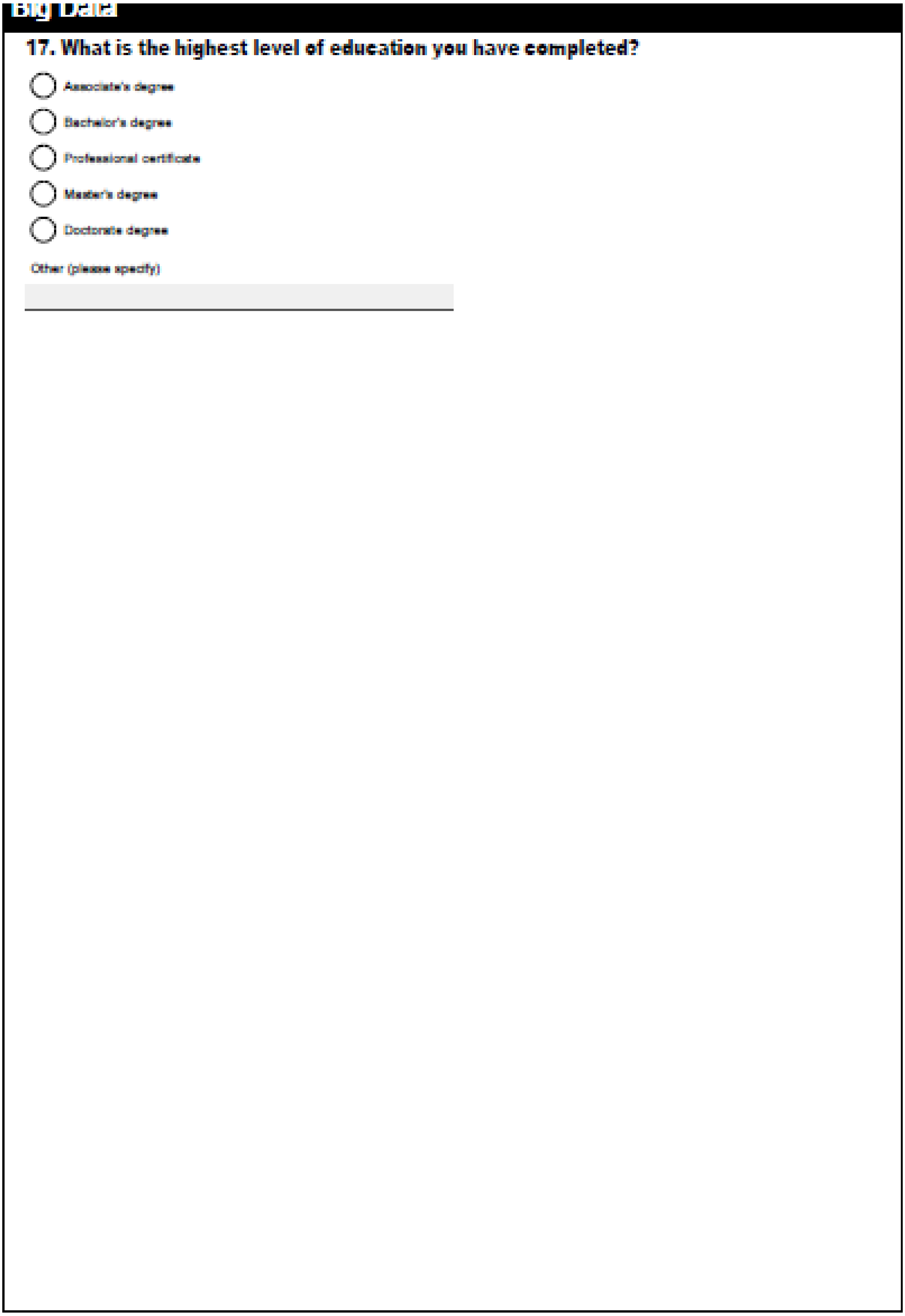
15. What is your job title?

- ○ IT/IS Manager
- ○ Chief Executive Officer
- ○ Data architect
- ○ Chief Information Officer
- ○ Business Intelligence analyst
- ○ IT consultant
- ○ Database administrator
- ○ Chief Technical Officer
- ○ Systems integrator
- ○ Developer
- ○ Project manager
- ○ Systems administrator
- ○ Business analyst
- ○ Business manager
- ○ Director
- ○ Analyst
- ○ Data scientist

Other (please specify)

16. How long have you been working in your current profession?

- ○ less than 1 year
- ○ 1 to 3 years
- ○ 3 - 6 years
- ○ 6 - 10 years
- ○ over 10 years

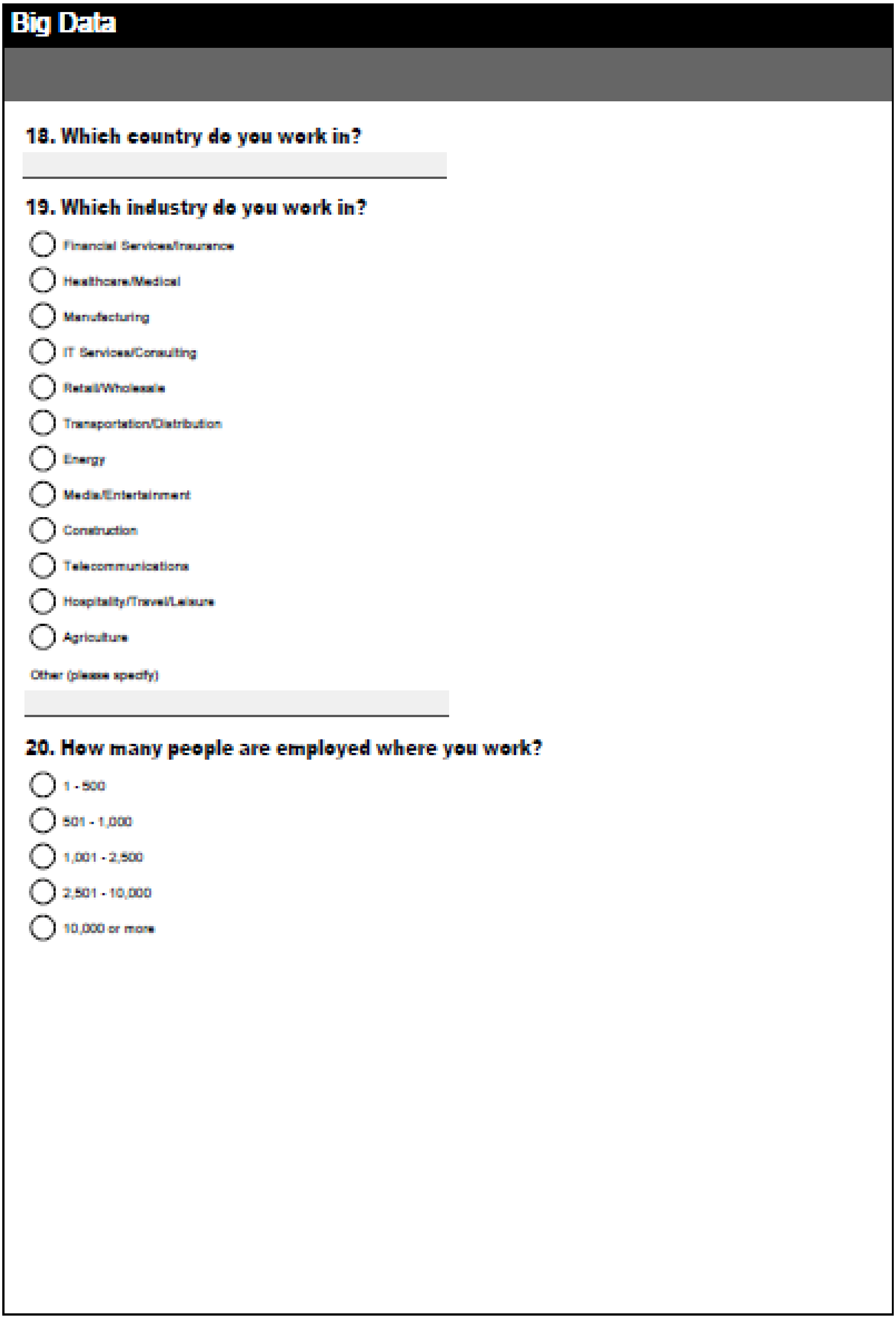


Big Data

17. What is the highest level of education you have completed?

○ Associate's degree
○ Bachelor's degree
○ Professional certificate
○ Master's degree
○ Doctorate degree

Other (please specify)



Big Data

18. Which country do you work in?

19. Which industry do you work in?

○ Financial Services/Insurance
○ Healthcare/Medical
○ Manufacturing
○ IT Services/Consulting
○ Retail/Wholesale
○ Transportation/Distribution
○ Energy
○ Media/Entertainment
○ Construction
○ Telecommunications
○ Hospitality/Travel/Leisure
○ Agriculture

Other (please specify)

20. How many people are employed where you work?

○ 1 - 500
○ 501 - 1,000
○ 1,001 - 2,500
○ 2,501 - 10,000
○ 10,000 or more